Springer Handbook of

Electronic and Photonic Materials

Kasap
Capper
Editors

2nd Edition

Springer

Springer Handbook of Electronic and Photonic Materials

Springer Handbooks provide a concise compilation of approved key information on methods of research, general principles, and functional relationships in physical and applied sciences. The world's leading experts in the fields of physics and engineering will be assigned by one or several renowned editors to write the chapters comprising each volume. The content is selected by these experts from Springer sources (books, journals, online content) and other systematic and approved recent publications of scientific and technical information.

The volumes are designed to be useful as readable desk reference book to give a fast and comprehensive overview and easy retrieval of essential reliable key information, including tables, graphs, and bibliographies. References to extensive sources are provided.

Springer Handbook

# of Electronic and Photonic Materials

Safa Kasap, Peter Capper (Eds.)

2nd Edition
With 1088 Figures and 185 Tables

Springer

Editors
Safa Kasap
University of Saskatchewan
Saskatoon, Canada
safa.kasap@usask.ca

Peter Capper
Southampton, UK
pete.capper@btinternet.com

ISBN: 978-3-319-48931-5 e-ISBN: 978-3-319-48933-9
DOI 10.1007/978-3-319-48933-9
Library of Congress Control Number: 2017944229

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Production and typesetting: le-tex publishing services GmbH, Leipzig
Typography and layout: schreiberVIS, Seeheim
Illustrations: le-tex publishing services GmbH, Leipzig
Cover design: eStudio Calamar Steinen, Barcelona
Cover production: WMXDesign GmbH, Heidelberg
Printing and binding: Printer Trento s.r.l., Trento

Printed on acid free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Prof. Arthur Willoughby
University of Southampton

The editors and authors of the *Springer Handbook of Electronic and Photonic Materials* are to be congratulated on the second edition of this distinguished volume, which will be an invaluable source of information for all students, scientists, professionals, and other interested people working in these areas. Having made contributions to earlier handbooks, I am well aware of the considerable and sustained work that is necessary to produce a volume of this kind. This particular Handbook, however, is distinguished by its breadth of coverage in the field and the way in which it discusses the very latest developments. In such a rapidly moving field, this is a considerable challenge, which has been met admirably.

Previous handbooks and encyclopedias have tended to concentrate on semiconducting materials, for the understandable reason of their dominance in the electronics field, and the wide range of semiconducting materials and phenomena that must be covered. Few have been courageous enough to predict future trends, but in 1992 *Mahajan* and *Kimerling* attempted this in the Introduction to their *Concise Encyclopaedia of Semiconducting Materials and Related Technologies* (Pergamon) and foresaw future challenges in the areas of nanoelectronics, low dislocation-density III–V substrates, semi-insulating III–V substrates, patterned epitaxy of II–Vs, alternative dielectrics and contacts for silicon technology, and developments in ion-implantation and diffusion. To a greater or lesser extent, all of these have been proved to be true, but it illustrates how difficult it is to make such a prediction.

Not many people would have thought, two decades ago, that the III–nitrides would occupy an important position in this book. As high melting point materials, with the associated growth problems, they were not high on the list of favorites for light emitters at the blue end of the spectrum! The story is a fascinating one – at least as interesting as the solution to the problem of the short working life of early solid-state lasers at the red end of the spectrum. Likewise, more recently, the advances in white-light emitting diodes (WLEDs) have impacted us all, in replacement of traditional incandescent bulbs for applications such as car headlights, torches, and other lighting, and are making a huge impact in saving energy. Optoelectronics and photonics, in general, have seen one of the most spectacular advances over the last decade, and this is fully reflected in the book, ranging from visible-light emitters and phosphors to infrared materials. Part D of this book covers a wide range of work in this area, including III–V and II–VI optoelectronic materials and bandgap engineering, as well as photonic glasses, liquid crystals, organic photoconductors, and the new area of photonic crystals. The entire part reflects materials for light generation, processing, transmission and detection – all the essential elements for using light instead of electrons.

New materials have emerged since the first edition of this book was published in 2006 and they are also fully reflected in the book. One of these is graphene, which is under active research in order to discover its most effective applications, and a whole chapter is deservedly devoted to this. Other 2D materials have followed this discovery, and offer similarly exciting possibilities.

Another strong feature of this book is its coverage of fundamental properties in Part A, including optical, magnetic, electronic and dielectric properties, which lays the groundwork for the rest of the book. Likewise, in Part B, we learn about single-crystal growth, epitaxy, semiconductor compounds' growth and related characterization, all of which will be extremely instructive to those new to the field.

In the *Materials for Electronics* section (Part C), the book charts the progress in silicon – overwhelmingly the dominant material for a whole range of electronic functions and circuitry – including new dielectrics and other issues associated with shrinking geometry of circuits and devices to produce ever higher packing densities. It also includes areas rarely covered in other books: thick films, high-temperature electronic materials, and amorphous and microcrystalline materials. The existing developments that extend the life of silicon technology, including silicon/germanium alloys, appear too, and raise the question again as to whether the predicted timetable for the demise of silicon has again been declared too early! Ferroelectrics – a class of materials used so effectively in conjunction with silicon – certainly deserve to be here.

The chapters in Part E (*Novel Materials and Selected Applications*), break new ground in a number of admirable ways. Most of us are familiar with informa-

tion recording devices such as CDs, DVDs, etc., but few are aware of the materials, or principles, involved. This book describes magnetic information storage materials, as well as phase-change optical recording, keeping us fully up to date with respect to recent developments. The chapters in this section cover a diverse range of materials with applications such as solar cells, sensors, photoconductors, carbon nanotubes, graphene, thermoelectric materials, and metamaterials on which such a huge volume of work is presently being pursued worldwide. Both ends of the spectrum, from research to applications, are represented in chapters on organic electronics and packaging materials, as well as flexible substrates, perovskites, and materials for terahertz engineering.

So, a particular strength of this book is that it ranges from the fundamental science (Part A) through growth and characterization of the materials (Part B) to applications (Parts C–E). Virtually all the materials covered here have a wide range of applications, which is one of the reasons why this book is going to be so useful. As I indicated before, few of us will be successful in predicting the future direction and trends that occupy the high-ground in this field in the coming decade, but this book teaches us the basic principles of materials and leaves it to us to adapt these to the needs of tomorrow. I commend it to you most warmly.

Arthur Willoughby

Preface

After several years in use, this second edition of the *Springer Handbook of Electronic and Photonic Materials* is brought out to update the coverage and include new topics. Like the first edition, it represents a comprehensive treatise on electronic and photonic materials, with each chapter written by experts in the field. It is aimed at senior undergraduate and graduate students, researchers, and practicing professionals working in the area of electronic, optoelectronic, and photonic materials.

The idea for the first edition of this volume grew out of the need for a reference work that covers the wide spectrum of topics in materials that today's electrical engineers, material scientists, and physicists need. Electronic and photonic materials is a truly interdisciplinary subject that encompasses a number of traditional disciplines such as materials science, electrical engineering, chemical engineering, mechanical engineering, physics, and chemistry. It is not unusual to find a mechanical engineering faculty carrying out research on electronic packaging, and electrical engineers carrying out characterization measurements on semiconductors. There are only a few established university departments specializing in electronic or photonic materials. In general, electronic materials as a *discipline* appears as a research group or as an interdisciplinary activity within a college or faculty. One could argue that, because of the very fact that it is such an interdisciplinary field, there is a greater need to have a handbook that covers not only fundamental topics but also advanced topics; hence the present Handbook.

The chapters in this second edition of the *Springer Handbook of Electronic and Photonic Materials* provide the necessary background and up-to-date knowledge in a wide range of topics. Each chapter has an introduction to the topic, many clear illustrations, and numerous references. Clear explanations and illustrations make the handbook useful for all levels of research. All chapters are as self-contained as possible. There are both fundamental and advanced chapters that will appeal to readers with different backgrounds. This is particularly important for this Handbook since the subject matter is highly interdisciplinary. For example, there will be readers with a background (first degree) in chemical engineering working on semiconductor processing and who need to learn the fundamentals of semiconductor physics. Someone with a first degree in physics would need to quickly update himself on materials science concepts such as liquid phase epitaxy and so on. Difficult mathematics has been avoided and, whenever possible, the explanations have been given semiquantitatively. There is a Glossary of Defining Terms at the end of the Handbook, which is a particularly useful feature in an interdisciplinary handbook.

As indicated before, the second edition is a major revision of the first one. Some chapters of the first edition have been left out and many chapters have gone through major revisions. Quite a few new chapters have been added: Photoconductivity in Materials Research; Temperature-Insensitive Band-Gap III-V Semiconductors: Tl-III-V and III-V-Bi; Graphene; Organic Solar Cells; Materials for Terahertz Engineering; Metamaterials; Thermoelectric Materials; Transparent Conductive Oxides; and Inorganic Perovskite Oxides.

The editors are very grateful to all the authors for their excellent contributions and for their cooperation in delivering their manuscripts in the various stages of the production of this Handbook. Sincere thanks go to Judith Hinterberg for all her support and help throughout the long period of commissioning the contributions and for getting them ready for production. Ursula Barth and Constanze Ober have very skilfully handled the myriad issues involved in getting the Handbook ready for production, copy-editing, figure redrawing, proof preparation and correction, and our sincere thanks go to them also for all their dedicated work in making the Handbook attractive to read.

It is a pleasure to thank Professor Arthur Willoughby for his many helpful suggestions that have made this a better Handbook. His wealth of experience as the Editor-in-Chief of the *Journal of Materials Science: Materials in Electronics* played an important role not only in selecting chapters but also in finding the right authors.

Finally, the Editors wish to thank all the members of their families (Nicolette (SK) and Marian, Samuel and Thomas (PC)) for their support and particularly for their endurance during the entire project.

Safa Kasap — Saskatoon, Canada
Peter Capper — Southampton, UK

About the Editors

Safa Kasap is currently a Distinguished Professor of Electronic and Optoelectronic Materials and Devices in the Electrical Engineering Department at the University of Saskatchewan, Canada. He obtained his BSc (1976), MSc (1978), and PhD (1983) degrees from the Imperial College of Science, Technology and Medicine, University of London, UK, specializing in amorphous semiconductors and chalcogenide glasses. In 1996 he was awarded the DSc (Engineering) from London University for his research contributions to materials science in electrical engineering. He is a Fellow of the American Physical Society and the Royal Society of Canada. He is the author of Optoelectronics and Photonics: Principles and Practices (2nd Edition) and Principles of Electronic Materials and Devices (4th Edition). Both textbooks are widely used by many major universities and professionals. He is the Deputy Editor-in-Chief for the Journal of Materials Science: Materials in Electronics.

Peter Capper was formerly Materials Team Leader at SELEX ES, which is now part of Leonardo MW Ltd. Dr Capper obtained his BA (1968) and PhD (1971) in Chemical Physics and Physics from Essex University and Portsmouth Polytechnic, UK, respectively. He has spent some 40 years in the infrared industry, growing and characterizing cadmium mercury telluride (CMT) and other tellurium-based materials. He holds patents in the bulk growth of CMT, has authored/co-authored over 100 papers and given several invited talks at international crystal growth/infrared (IR) conferences. He is a Fellow of the Institute of Physics and an Editor for the Journal of Materials Science: Materials in Electronics. He has edited/co-edited ten books in IR materials and devices and crystal growth.

List of Authors

Martin Abkowitz (deceased)

Sadao Adachi
Gunma University
Faculty of Science and Technology
Gunma, Japan
adachi@gunma-u.ac.jp

Guy J. Adriaenssens
University of Leuven
Lab. for Semiconductor Physics
Leuven, Belgium
guy.adri@fys.kuleuven.be

Wilfried von Ammon
Von Ammon Consulting
Roth, Germany
wilfried.ammon@aon.at

Hajime Asahi
Osaka University
Inst. of Scientific and Industrial Research
Osaka, Japan
asahi@sanken.osaka-u.ac.jp

Peter Ashburn
University of Southampton
School of Electronics and Computer Science
Southampton, UK
pa@ecs.soton.ac.uk

Mark Auslender
Ben-Gurion University of the Negev Beer Sheva
Dept. of Electrical and Computer Engineering
Beer Sheba, Israel
marka@bgu.ac.il

Darren M. Bagnall
University of New South Wales
School of Photovoltaic and Renewable Energy Engineering
Sydney, Australia
d.bagnall@unsw.edu.au

Ian M. Baker
Leonardo MW Ltd
Southampton, UK
ian.m.baker@leonardocompany.com

Sergei Baranovskii
Philipps University
Dept. of Physics
Marburg, Germany
baranovs@staff.uni-marburg.de

Mark Baxendale
Queen Mary University of London
School of Physics and Astronomy
London, UK
m.baxendale@qmul.ac.uk

Mohammed L. Benkhedir
Tebessa University
Lab. of Applied and Theoretical Physics
Tebessa, Algeria
benkhedir@gmail.com

Monica Brinza
ASML Netherlands B.V.
Veldhoven, The Netherlands
monica.brinza@asml.com

Paul D. Brown
University of Nottingham
Dept. of Mechanical, Materials and Manufacturing Engineering
Nottingham, UK
paul.brown@nottingham.ac.uk

Peter Capper
Formerly with Leonardo MW Ltd
Southampton, UK
pete.capper@btinternet.com

R. Lawrence Comstock (deceased)

Ray DeCorby
University of Alberta
Electrical and Computer Engineering
Edmonton, Canada
rdecorby@ualberta.ca

M. Jamal Deen
McMaster University
Dept. of Electrical and Computer Engineering
Hamilton, Canada
jamal@mcmaster.ca

Leonard Dissado
University of Leicester
Dept. of Engineering
Leicester, UK
lad4@leicester.ac.uk

David Dunmur
University of Southampton
Chemistry
Southampton, UK
d.dunmur@tiscali.co.uk

Lester F. Eastman (deceased)

Andy Edgar
Victoria University
School of Chemical and Physical Sciences
Wellington, New Zealand
andy.edgar@vuw.ac.nz

Vassili Fedotov
University of Southampton
Optoelectronics Research Centre
Southampton, UK
vaf@orc.soton.ac.uk

Romualdo A. Ferreyra
Kyoto Institute of Technology
Innovation Center
Kyoto, Japan
raferrey@kit.ac.jp

Paul Fons
National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan
paul-fons@aist.go.jp

Brian E. Foutz
Charlottesville, USA
foutz@cadence.com

Mark Fox
University of Sheffield
Dept. of Physics and Astronomy
Sheffield, UK
mark.fox@sheffield.ac.uk

Darrel Frear
NXP Semiconductor
Core Technology in Package Innovation
Chandler, USA
darrel.frear@nxp.com

Jacek Gieraltowski
European Institute for Marine Studies
Oceanic Domains Laboratory
Plouzané, France
jacek.gieraltowski@univ-brest.fr

Robert D. Gould (deceased)

Yaser M. Haddara
McMaster University
Dept. of Electrical and Computer Engineering
Hamilton, Canada
yaser@mcmaster.ca

Walid A. Hadi
University of Windsor
Dept. of Electrical and Computer Engineering
Windsor, Canada
abdulhw@uwindsor.ca

Shlomo Hava
Ben-Gurion University of the Negev Beer Sheva
Dept. of Electrical and Computer Engineering
Beer Sheba, Israel
hava@bgu.ac.il

Masahiro Hiramoto
Institute for Molecular Science
Dept. of Materials Molecular Science
Okazaki, Japan
hiramoto@ims.ac.jp

Hideo Hosono
Tokyo Institute of Technology
Materials Research Center of Element Strategy
Yokohama, Japan
hosono@msl.titech.ac.jp

Colin Humphreys
University of Cambridge
Dept. Materials Science and Metallurgy
Cambridge, UK
colin.humphreys@msm.cam.ac.uk

Mehrdad Irannejad
University of Waterloo
Dept. of Mechanical and Mechatronics Engineering
Waterloo, Canada
mehrdad.irannejad@uwaterloo.ca

Stuart Irvine
Swansea University
College of Engineering
Swansea, UK
s.j.c.irvine@swansea.ac.uk

Tatsumi Ishihara
Kyushu University
Dept. of Applied Chemistry
Fukuoka, Japan
ishihara@cstf.kyushu-u.ac.jp

Radu Ispasoiu
ON-Semiconductor
Image Sensor Group
San José, USA
radu.ispasoiu@onsemi.com

Minoru Isshiki
Tohoku University
Inst. of Multidisciplinary Research for Advanced Materials
Sendai, Japan
isshiki@agate.plala.or.jp

Tim Joyce
University of Liverpool
School of Engineering
Liverpool, UK
t.joyce@liverpool.ac.uk

M. Zahangir Kabir
Concordia University
Dept. of Electrical and Computer Engineering
Montreal, Canada
zahangir.kabir@concordia.ca

Safa Kasap
University of Saskatchewan
Saskatoon, Canada
safa.kasap@usask.ca

Gudrun Kissinger
Innovations for High Performance Microelectronics (IHP)
Frankfurt/Oder, Germany
gkissinger@ihp-microelectronics.com

Alexander V. Kolobov
National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan
a.kolobov@aist.go.jp

Cyril Koughia
University of Saskatchewan
Dept. of Electrical and Computer Engineering
Saskatoon, Canada
cyril.koughia@usask.ca

Roger Lewis
University of Wollongong
School of Physics
Wollongong, Australia
roger@uow.edu.au

Geoffrey Luckhurst
University of Southampton
Chemistry
Southampton, UK
gl@soton.ac.uk

Jiří Málek
University of Pardubice
Dept. of Physical Chemistry
Pardubice, Czech Republic
jiri.malek@upce.cz

Akihisa Matsuda
Osaka University
Dept. of Systems Innovation
Osaka, Japan
matsuda@ee.es.osaka-u.ac.jp

Naomi Matsuura
University of Toronto
Dept. of Materials Science & Engineering
Toronto, Canada
naomi.matsuura@utoronto.ca

Jan Mistrik
University of Pardubice
Center of Materials and Nanotechnology
Pardubice, Czech Republic
jan.mistrik@upce.cz

Winfried Mönch
University Duisburg-Essen
Faculty of Physics
Duisburg, Germany
winfried.moench@uni-due.de

Donald T. Morelli
Michigan State University
Chemical Engineering and Materials Science
East Lansing, USA
dmorelli@egr.msu.edu

Kazuo Morigaki
University of Tokyo
Tokyo, Japan
k.morigaki@yacht.ocn.ne.jp

Hadis Morkoç
Virginia Commonwealth University
Electrical and Computer Engineering
Richmond, USA
hmorkoc@vcu.edu

Jayanta Mukherjee
Compound Photonics UK Ltd
Durham, UK
jayanta.mukherjee@compoundphotonics.com

Takashi Nagase
Osaka Prefecture University
Dept. of Physics and Electronics
Sakai, Japan
nagase@pe.osakafu-u.ac.jp

Hiroyoshi Naito
Osaka Prefecture University
Dept. of Physics and Electronics
Sakai, Japan
naito@pe.osakafu-u.ac.jp

Arokia Nathan
University of Cambridge
Dept. of Electrical Engineering
Cambridge, UK
an299@eng.cam.ac.uk

Stephen K. O'Leary
University of British Columbia
School of Engineering
Kelowna, Canada
stephen.oleary@ubc.ca

Chisato Ogihara
Yamaguchi University
Graduate School of Science and Engineering
Yamaguchi, Japan
ogihara@yamaguchi-u.ac.jp

Fabien Pascal
University of Montpellier
Institute of Electronic and Systems
Montpellier, France
fabien.pascal@ies.univ-montp2.fr

Michael C. Petty
Durham University
School of Engineering and Computing Sciences
Durham, UK
m.c.petty@durham.ac.uk

Henry H. Radamson
Chinese Academy of Science
Inst. of Microelectronics
Beijing, China
and
KTH Royal Institute of Technology
School of Information and Communication Technology
Stockholm, Sweden
rad@ime.ac.cn

Asim K. Ray
Brunel University London
Inst. of Materials and Manufacturing
Middlesex, UK
asim.ray@brunel.ac.uk

Stephen Reynolds
University of Dundee
School of Science and Engineering
Dundee, UK
s.z.reynolds@dundee.ac.uk

Oleg Rubel
McMaster University
Dept. of Materials Science and Engineering
Hamilton, Canada
rubelo@mcmaster.ca

Harry E. Ruda
University of Toronto
Dept. of Materials Science and Engineering, and
Edward Rogers Dept. of Electrical and Computer Engineering
Toronto, Canada
harry.ruda@utoronto.ca

Andreas Sattler
Siltronic AG
München, Germany
andreas.sattler@siltronic.com

Peyman Servati
University of British Columbia
Electrical and Computer Engineering
Vancouver, Canada
peymans@ece.ubc.ca

Derek Shaw
Hull University
Dept. of Physics
Hull, UK
derekshaw47@gmail.com

Fumio Shimura
Shizuoka Institute of Science and Technology
Dept. of Materials and Life Science
Fukuroi, Japan
shimura@ms.sist.ac.jp

Yusuke Shinmura
Institute for Molecular Science
Dept. of Materials Molecular Science
Okazaki, Japan
shinmura@ims.ac.jp

Michael S. Shur
Rensselaer Polytechnic Institute
Dept. of Physics, Applied Physics, and Astronomy
New York, USA
shurm@rpi.edu

Poppy Siddiqua
University of British Columbia
School of Engineering
Kelowna, Canada
siddiquapoppy@gmail.com

Jai Singh
Charles Darwin University
School of Engineerng and Information Technology
Darwin, Australia
jai.singh@cdu.edu.au

Tim Smeeton
SHARP Laboratories of Europe
Oxford, UK
tim.smeeton@sharp.co.uk

Hitoshi Suzuki
Hiroshima University
Graduate School of Advanced Sciences of Matter
Higashihiroshima, Japan
hitoshi-suzuki@hiroshima-u.ac.jp

Roman Svoboda
University of Pardubice
Dept. of Physical Chemistry
Pardubice, Czech Republic
roman.svoboda@upce.cz

Stephen J. Sweeney
University of Surrey
Dept. of Physics and Advanced Technology Institute
Guildford, UK
s.sweeney@surrey.ac.uk

David Sykes
Loughborough Surface Analysis Ltd
Loughborough, UK
d.e.sykes@lsaltd.co.uk

Keiji Tanaka
Hokkaido University
Dept. Applied Physics
Sapporo, Japan
keiji@eng.hokudai.ac.jp

Charbel Tannous
University of Western Brittany
Lab. of Magnetism of Brittany
Brest, France
charbel.tannous@univ-brest.fr

Ali Teke
Balikesir University
Dept. of Physics
Balikesir, Turkey
ateke@balikesir.edu.tr

Junji Tominaga
National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan
j-tominaga@aist.go.jp

Harry Tuller
Massachusetts Institute of Technology
Dept. of Materials Science and Engineering
Cambridge, USA
tuller@mit.edu

Kazushige Ueda
Kyushu Institute of Technology
Dept. of Materials Science
Kitakyushu, Japan
kueda@che.kyutech.ac.jp

Robert M. Wallace
University of Texas at Dallas
Dept. of Materials Science and Engineering
Richardson, USA
rmwallace@utdallas.edu

Jifeng Wang
Rasa Industries Ltd.
Advanced Microelectronic Materials Development Center
Tokyo, Japan
wypwm6866@yahoo.co.jp

David S. Weiss
Molecular Glasses, Inc.
Rochester, USA
dweiss2@rochester.rr.com

Rainer Wesche
Swiss Federal Institute of Technology Lausanne
Swiss Plasma Center
Villigen, Switzerland
rainer.wesche@psi.ch

Roger Whatmore
Imperial College London
Dept. of Materials
London, UK
roger.whatmore@gmail.com

Neil White
University of Southampton
School of Electronics and Computer Science
Southampton, UK
nmw@ecs.soton.ac.uk

Congyong Zhu
Infineon Technologies
Chandler, USA
zhuc@mymail.vcu.edu

Contents

List of Abbreviations XXIX

1 **Perspectives on Electronic and Photonic Materials**
Tim Smeeton, Colin Humphreys 1
1.1 Tremendous Integration 2
1.2 The Silicon Age 3
1.3 The Compound Semiconductors 7
References 14

Part A Fundamental Properties

2 **Electrical Conduction in Metals and Semiconductors**
Safa Kasap, Cyril Koughia, Harry E. Ruda 19
2.1 Fundamentals: Drift Velocity, Mobility and Conductivity 20
2.2 Matthiessen's Rule 22
2.3 Resistivity of Metals 23
2.4 Solid Solutions and Nordheim's Rule 26
2.5 Carrier Scattering in Semiconductors 28
2.6 The Boltzmann Transport Equation 29
2.7 Resistivity of Thin Polycrystalline Films 30
2.8 Inhomogeneous Media: Effective Media Approximation 32
2.9 The Hall Effect 35
2.10 High Electric Field Transport 37
2.11 Impact Ionization 38
2.12 Two-Dimensional Electron Gas 40
2.13 One-Dimensional Conductance 42
2.14 The Quantum Hall Effect 43
References 44

3 **Optical Properties of Electronic Materials: Fundamentals and Characterization**
Jan Mistrik, Safa Kasap, Harry E. Ruda, Cyril Koughia, Jai Singh 47
3.1 Optical Constants 47
3.2 Refractive Index 50
3.3 Optical Absorption 53
3.4 Optical Characterization 69
3.5 Optical Materials 79
References 82

4 **Magnetic Properties: From Traditional to Spintronic**
Charbel Tannous, Jacek Gieraltowski 85
4.1 Traditional Magnetism 87
4.2 Nonconventional Magnetism and Progress Toward Spintronics and Quantum Devices 99
4.3 Spintronics and Quantum Information Devices 104
References 109

5 **Defects in Monocrystalline Silicon**
Wilfried von Ammon, Andreas Sattler, Gudrun Kissinger 111
5.1 Technological Impact of Intrinsic Point Defects Aggregates 112
5.2 Thermophysical Properties of Intrinsic Point Defects 113
5.3 Aggregates of Intrinsic Point Defects 115
5.4 Formation of OSF Ring 127
References 129

6 **Diffusion in Semiconductors**
Derek Shaw 133
6.1 Basic Concepts 134
6.2 Diffusion Mechanisms 134
6.3 Diffusion Regimes 136
6.4 Internal Electric Fields 138
6.5 Measurement of Diffusion Coefficients 139
6.6 Hydrogen in Semiconductors 140
6.7 Diffusion in Group IV Semiconductors 141
6.8 Diffusion in III–V Compounds 143
6.9 Diffusion in II–VI Compounds 144
6.10 Nano Volume Diffusion 146
6.11 Diffusion in Molten Semiconductors 146
6.12 The Meyer–Neldel Rule 146
6.13 Conclusions 147
6.14 General Reading and References 147
References 147

7 **Photoconductivity in Materials Research**
Stephen Reynolds, Monica Brinza, Mohammed L. Benkhedir, Guy J. Adriaenssens 151
7.1 Steady-State Photoconductivity (SSPC) 153
7.2 Constant Photocurrent Method (CPM) and Related Techniques 157
7.3 Steady-State Photocarrier Grating Method (SSPG) 160
7.4 Modulated Photocurrent Spectroscopy (MPC) 161
7.5 Switch-on and Switch-off Transients 164
7.6 Transient Photocurrent Spectroscopy (TPC) 166
7.7 Time-of-Flight (TOF) and Related Techniques 168
7.8 Other Photoconductivity-Related Techniques 171
References 172

8 **Electronic Properties of Semiconductor Interfaces**
Winfried Mönch 175
8.1 Experimental Database 177
8.2 IFIGS-and-Electronegativity Theory 181
8.3 Comparison of Experiment and Theory 183
8.4 Modifications of Schottky Contacts 187
8.5 Graphene Schottky Contacts 190
8.6 Final Remarks 190
References 191

9 **Charge Transport in Disordered Materials**
Sergei Baranovskii, Oleg Rubel 193
9.1 General Remarks on Charge Transport in Disordered Materials 195
9.2 Charge Transport in Disordered Materials via Extended States 198
9.3 Hopping Charge Transport in Disordered Materials via Localized States 200
9.4 Concluding Remarks 215
References 216

10 **Dielectric Response**
Leonard Dissado 219
10.1 Definition of Dielectric Response 220
10.2 Frequency-Dependent Linear Responses 222
10.3 Information Contained in the Relaxation Response 228
10.4 Charge Transport 240
10.5 Data Presentation 243
10.6 A Few Final Comments 243
References 244

11 **Ionic Conduction and Applications**
Harry Tuller 247
11.1 Conduction in Ionic Solids 248
11.2 Fast Ion Conduction 251
11.3 Mixed Ionic–Electronic Conduction 256
11.4 Applications 258
11.5 Future Trends 262
References 263

Part B Growth and Characterization

12 **Bulk Crystal Growth: Methods and Materials**
Peter Capper 269
12.1 Background 269
12.2 History 270
12.3 Techniques 271
12.4 Materials Grown 278
12.5 Conclusions 289
References 289

13 **Single-Crystal Silicon: Growth and Properties**
Fumio Shimura 293
13.1 Overview 294
13.2 Starting Materials 295
13.3 Single-Crystal Growth 296
13.4 New Crystal Growth Methods 304
References 306

14 **Epitaxial Crystal Growth: Methods and Materials**
Peter Capper, Stuart Irvine, Tim Joyce 309
14.1 Liquid-Phase Epitaxy (LPE) 309
14.2 Metal Organic Chemical Vapor Deposition 319

14.3 Molecular Beam Epitaxy (MBE) 329
References 337

15 **Narrow Bandgap II–VI Semiconductors: Growth**
Peter Capper 343
15.1 Bulk Growth Techniques 344
15.2 Liquid Phase Epitaxy (LPE) 349
15.3 Metal–Organic Vapor Phase Epitaxy (MOVPE) 353
15.4 Molecular Beam Epitaxy (MBE) 358
15.5 Alternatives to MCT 361
References 362

16 **Wide-Bandgap II–VI Semiconductors: Growth and Properties**
Minoru Isshiki, Jifeng Wang 365
16.1 Crystal Properties 367
16.2 Epitaxial Growth 371
16.3 Bulk Crystal Growth 375
16.4 Conclusions 381
References 381

17 **Structural Characterization**
Paul D. Brown 385
17.1 Radiation–Material Interactions 386
17.2 Particle–Material Interactions 387
17.3 X-ray Diffraction 390
17.4 Optics, Imaging and Electron Diffraction 393
17.5 Characterizing Functional Activity 403
17.6 Sample Preparation 404
17.7 Case Studies – Complementary Characterization of Electronic and Optoelectronic Materials 406
17.8 Concluding Remarks 411
References 412

18 **Surface Chemical Analysis**
David Sykes 413
18.1 Electron Spectroscopy 414
18.2 Glow-Discharge Spectroscopies (GDOES and GDMS) 416
18.3 Secondary Ion Mass Spectrometry (SIMS) 417
18.4 Conclusion 423

19 **Thermal Properties and Thermal Analysis: Fundamentals, Experimental Techniques and Applications**
Safa Kasap, Jiří Málek, Roman Svoboda 425
19.1 Heat Capacity 426
19.2 Thermal Conductivity 431
19.3 Thermal Expansion 436
19.4 Enthalpic Thermal Properties 438
19.5 Temperature-Modulated DSC (TMDSC) 446
References 449

20 **Electrical Characterization of Semiconductor Materials and Devices**
M. Jamal Deen, Fabien Pascal 453
20.1 Resistivity 454
20.2 Hall Effect 462
20.3 Capacitance–Voltage Measurements 464
20.4 Current–Voltage Measurements 470
20.5 Charge Pumping 472
20.6 Low-Frequency Noise 473
20.7 Deep-Level Transient Spectroscopy 477
References 479

Part C Materials for Electronics

21 **Single-Crystal Silicon: Electrical and Optical Properties**
Mark Auslender, Shlomo Hava 485
21.1 Silicon Basics 485
21.2 Electrical Properties 495
21.3 Optical Properties 515
References 520

22 **Silicon-Germanium: Properties, Growth and Applications**
Yaser M. Haddara, Peter Ashburn, Darren M. Bagnall 523
22.1 Physical Properties of Silicon-Germanium 524
22.2 Optical Properties of SiGe 530
22.3 Growth of Silicon-Germanium 534
22.4 Polycrystalline Silicon-Germanium 537
References 539

23 **Temperature-Insensitive Band-Gap III-V Semiconductors: TI-III-V and III-V-Bi**
Hajime Asahi 543
23.1 TI-III-V Alloy Semiconductors 544
23.2 III-V-Bi Alloy Semiconductors 550
23.3 Summary 554
References 554

24 **Amorphous Semiconductors: Structure, Optical, and Electrical Properties**
Kazuo Morigaki, Chisato Ogihara 557
24.1 Electronic States 557
24.2 Structural Properties 560
24.3 Optical Properties 561
24.4 Electrical Properties 565
24.5 Light-Induced Phenomena 567
24.6 Nanosized Amorphous Structure 569
References 570

25 **Amorphous and Microcrystalline Silicon**
Akihisa Matsuda 573
25.1 Reactions in SiH_4 and SiH_4/H_2 Plasmas 574
25.2 Film Growth on a Surface 575

25.3 Defect Density Determination for a-Si:H and μc-Si:H 580
25.4 Device Applications 582
25.5 Recent Progress in Material Issues Related to Thin-Film Silicon Solar Cells 583
25.6 Summary 586
References 586

26 **Ferroelectric Materials**
Roger Whatmore 589
26.1 Definitions and Background 589
26.2 Ferroelectric Materials 593
26.3 Ferroelectric Materials Fabrication Technology 600
26.4 Ferroelectric Applications 607
References 613

27 **Dielectric Materials for Microelectronics**
Robert M. Wallace 615
27.1 Overview 615
27.2 Gate Dielectrics 620
27.3 Isolation Dielectrics 635
27.4 Capacitor Dielectrics 636
27.5 Interconnect Dielectrics 639
27.6 Summary 641
References 641

28 **Thin Films**
Robert D. Gould†, Safa Kasap, Asim K. Ray 645
28.1 Deposition Methods 647
28.2 Structure 669
28.3 Properties 680
28.4 Concluding Remarks 699
References 702

29 **Thick Films**
Neil White 707
29.1 Thick Film Processing 708
29.2 Substrates 710
29.3 Thick Film Materials 711
29.4 Components and Assembly 714
29.5 Sensors 717
References 720

Part D Materials for Optoelectronics and Photonics

30 **III–V Ternary and Quaternary Compounds**
Sadao Adachi 725
30.1 Introduction to III–V Ternary and Quaternary Compounds 726
30.2 Interpolation Scheme 726
30.3 Structural Parameters 727
30.4 Mechanical, Elastic and Lattice Vibronic Properties 729
30.5 Thermal Properties 731

30.6 Energy Band Parameters 733
30.7 Optical Properties 737
30.8 Carrier Transport Properties 739
References 741

31 **Group III Nitrides**
Romualdo A. Ferreyra, Congyong Zhu, Ali Teke, Hadis Morkoç 743
31.1 Crystal Structures of Nitrides 745
31.2 Lattice Parameters of Nitrides 746
31.3 Mechanical Properties of Nitrides 748
31.4 Thermal Properties of Nitrides 752
31.5 Electrical Properties of Nitrides 757
31.6 Optical Properties of Nitrides 769
31.7 Properties of Nitride Alloys 782
31.8 Doped GaN 786
31.9 Defects in GaN 789
31.10 GaN-Based Nanostructures 794
31.11 Summary and Conclusions 801
References 802

32 **Electron Transport Within III-V Nitride Semiconductors**
Stephen K. O'Leary, Poppy Siddiqua, Walid A. Hadi, Brian E. Foutz, Michael S. Shur, Lester F. Eastman† 829
32.1 Electron Transport Within Semiconductors and the Monte Carlo Simulation Approach 830
32.2 Steady-State and Transient Electron Transport Within Bulk Wurtzite GaN, AlN, and InN 834
32.3 Electron Transport Within III–V Nitride Semiconductors: A Review 845
32.4 Conclusions 850
References 850

33 **II-IV Semiconductors for Optoelectronics: CdS, CdSe, CdTe**
Minoru Isshiki, Jifeng Wang 853
33.1 Background 853
33.2 Solar Cells 853
33.3 Radiation Detectors 858
33.4 Conclusions 863
References 863

34 **II-VI Narrow Bandgap Semiconductors: Optoelectronics**
Ian M. Baker 867
34.1 Overview 867
34.2 Applications and Sensor Design 870
34.3 Photoconductive Detectors in HgCdTe 871
34.4 Sprite Detectors 872
34.5 Introduction to Photovoltaic Devices in HgCdTe 873
34.6 Manufacturing Technology for Photodiode Arrays 879
34.7 Advanced HgCdTe Technologies 886
34.8 Emission Devices in II–VI Semiconductors 892
34.9 Potential for Reduced Dimensionality in HgTe–CdTe 893
References 893

35 **Optoelectronic Devices and Materials**
Stephen J. Sweeney, Jayanta Mukherjee 897
35.1 Introduction to Optoelectronic Devices 898
35.2 Light-Emitting Diodes and Semiconductor Lasers 900
35.3 Single- and Multimode Lasers 913
35.4 Optical Amplifiers 922
35.5 Modulators 923
35.6 Photodetectors 927
35.7 Conclusions 929
References 930

36 **Liquid Crystals**
Geoffrey Luckhurst, David Dunmur 933
36.1 Introduction to Liquid Crystals 933
36.2 The Basic Physics of Liquid Crystals 940
36.3 Liquid-Crystal Devices 947
36.4 Materials for Displays 956
References 964

37 **Organic Photoconductors**
David S. Weiss, Martin Abkowitz† 967
37.1 Chester Carlson and Xerography 968
37.2 Operational Considerations and Critical Materials Properties 970
37.3 OPC Characterization 978
37.4 OPC Architecture and Composition 980
37.5 Photoreceptor Fabrication 990
37.6 Summary 990
References 992

38 **Luminescent Materials**
Andy Edgar 997
38.1 Luminescent Centres 999
38.2 Interaction with the Lattice 1001
38.3 Thermally Stimulated Luminescence 1003
38.4 Optically (Photo-)Stimulated Luminescence 1004
38.5 Experimental Techniques – Photoluminescence 1005
38.6 Applications 1006
38.7 Representative Phosphors 1011
References 1011

39 **Nano-Engineered Tunable Photonic Crystals**
Harry E. Ruda, Naomi Matsuura 1013
39.1 PC Overview 1014
39.2 Traditional Fabrication Methodologies for Static PCs 1017
39.3 Tunable PCs 1027
39.4 Summary and Conclusions 1030
References 1031

40 **Quantum Wells, Superlattices, and Band-Gap Engineering**
Mark Fox, Radu Ispasoiu 1037
40.1 Principles of Bandgap Engineering and Quantum Confinement 1038
40.2 Optoelectronic Properties of Quantum-Confined Structures 1040
40.3 Emitters 1047

40.4 Detectors 1050
40.5 Modulators 1052
40.6 Quantum Dots and Nanowires 1053
40.7 Conclusions 1055
References 1055

41 **Glasses for Photonic Integration**
Ray DeCorby, Mehrdad Irannejad 1059
41.1 Main Attributes of Glasses as Photonic Materials 1061
41.2 Glasses for Integrated Optics 1068
41.3 Laser Glasses for Integrated Light Sources 1071
41.4 Summary 1077
References 1077

42 **Optical Nonlinearity in Photonic Glasses**
Keiji Tanaka 1081
42.1 Third-Order Nonlinearity in Homogeneous Glass 1083
42.2 Second-Order Nonlinearity in Poled Glass 1088
42.3 Particle-Embedded Systems 1089
42.4 Photoinduced Phenomena 1090
42.5 Summary 1091
References 1092

Part E Novel Materials and Selected Applications

43 **Solar Cells and Photovoltaics**
Stuart Irvine 1097
43.1 Figures of Merit for Solar Cells 1098
43.2 Crystalline Silicon 1100
43.3 Amorphous Silicon 1102
43.4 GaAs Solar Cells 1102
43.5 CdTe Thin-Film Solar Cells 1104
43.6 $CuInGaSe_2$ ($CIGS_2$) Thin-Film Solar Cells 1105
43.7 Excitonic PV 1107
43.8 Conclusions 1107
References 1108

44 **Disordered Semiconductors on Mechanically Flexible Substrates for Large-Area Electronics**
Peyman Servati, Arokia Nathan 1111
44.1 a-Si:H TFTs on Flexible Substrates 1112
44.2 Field-Effect Transport in Amorphous Films 1112
44.3 Electronic Transport Under Mechanical Stress 1118
References 1123

45 **Photoconductors for X-Ray Image Detectors**
M. Zahangir Kabir, Safa Kasap 1125
45.1 X-Ray Photoconductors 1127
45.2 Dark Current Limitations 1136
45.3 Metrics of Detector Performance 1139
45.4 Summary 1145
References 1145

46 **Phase-Change Memory Materials**
Alexander V. Kolobov, Junji Tominaga, Paul Fons 1149
46.1 Structure of Ge-Sb-Te Phase-Change Alloys 1150
46.2 Mechanism of the Phase-Change Process 1156
46.3 Present Applications and Future Trends 1158
References 1160

47 **Carbon Nanotubes and Bucky Materials**
Mark Baxendale 1163
47.1 Carbon Nanotubes 1163
47.2 Bucky Materials 1170
References 1170

48 **Graphene**
Henry H. Radamson 1173
48.1 Graphene Synthesis 1174
48.2 Band Structure and Electronic Applications 1176
48.3 Characterization of Graphene Material 1178
48.4 Photonic Applications 1180
References 1182

49 **Magnetic Information-Storage Materials**
Charbel Tannous, R. Lawrence Comstock† 1185
49.1 Magnetic Recording Technology 1186
49.2 Magnetic Random-Access Memory 1215
49.3 Extraordinary Magnetoresistance (EMR) 1220
49.4 Summary 1220
References 1220

50 **High-Temperature Superconductors**
Rainer Wesche 1225
50.1 The Superconducting State 1228
50.2 Cuprate High-T_c Superconductors: An Overview 1234
50.3 Physical Properties of Cuprate Superconductors 1239
50.4 Superconducting Films 1244
50.5 The Special Case of MgB_2 1245
50.6 Iron-Based Superconductors 1248
50.7 Summary 1250
References 1252

51 **Molecular Electronics**
Michael C. Petty, Takashi Nagase, Hitoshi Suzuki, Hiroyoshi Naito 1257
51.1 Electrically Conductive Organic Compounds 1258
51.2 Materials 1261
51.3 Plastic Electronics 1263
51.4 Molecular-Scale Electronics 1266
51.5 DNA Electronics 1274
51.6 Conclusions 1276
References 1276

52 **Organic Materials for Chemical Sensing**
Asim K. Ray 1281
52.1 Analyte Requirements 1282
52.2 Brief Review of Inorganic Materials 1283
52.3 Macrocylic Compounds for Sensing 1286
52.4 Sensing with Phthalocyanine and Porphyrin 1291
52.5 Polymeric Materials 1297
52.6 Cavitand Molecules 1303
52.7 Concluding Remarks 1305
References 1305

53 **Packaging Materials**
Darrel Frear 1311
53.1 Package Applications 1312
53.2 The Materials Challenge of Electronic Packaging 1313
53.3 Materials Coefficient of Thermal Expansion 1315
53.4 Wirebond Materials 1315
53.5 Solder Interconnects 1317
53.6 Substrates 1322
53.7 Underfill and Encapsulants 1323
53.8 Electrically Conductive Adhesives (ECAs) 1324
53.9 Thermal Issues 1326
53.10 Summary 1327
References 1327

54 **Organic Solar Cells**
Masahiro Hiramoto, Yusuke Shinmura 1329
54.1 History 1330
54.2 Excitons 1331
54.3 Donor–Acceptor Sensitization 1332
54.4 Exciton Diffusion 1332
54.5 Blended Junctions 1333
54.6 Route Formation 1334
54.7 π–π Stacking 1335
54.8 HOMO–LUMO Gap 1335
54.9 Tandem Cells 1336
54.10 Conclusions 1337
References 1337

55 **Materials for Terahertz Engineering**
Roger Lewis 1339
55.1 Overview 1339
55.2 Terahertz Sources 1343
55.3 Terahertz Detectors 1345
55.4 Terahertz Components 1346
55.5 Conclusion 1347
References 1347

56 **Metamaterials**
Vassili Fedotov 1351
56.1 Bulk Metamaterials 1351
56.2 Planar Metamaterials 1365
References 1375

57 **Thermoelectric Materials**
Donald T. Morelli 1379
57.1 Overview of the Field 1380
57.2 Semiconductors as Thermoelectric Materials 1381
57.3 New Concepts in Thermoelectric Materials Design 1385
57.4 Summary and Future Outlook 1389
References 1390

58 **Transparent Conductive Oxides**
Hideo Hosono, Kazushige Ueda 1391
58.1 Overview 1391
58.2 Materials Design for TCOs 1395
58.3 New Approach to Explore Candidate Materials: Materials Genome Approach 1402
References 1403

59 **Inorganic Perovskite Oxides**
Tatsumi Ishihara 1405
59.1 Typical Properties of Perovskite Oxides 1409
59.2 Photocatalytic Activity 1411
59.3 Application for Solid Oxide Fuel Cells (SOFCs) 1413
59.4 Oxygen Separating Membrane 1418
59.5 Summary 1419
References 1420

About the Authors 1421
Detailed Contents 1435
Glossary of Defining Terms 1461
Subject Index 1495

List of Abbreviations

2DEG	two-dimensional electron gas
2DHG	two-dimensional hole gas
3DAP	three-dimensional atom probe microscopy

A

ABE	acceptor-bound exciton
AC	alternating current
ACPM	absolute constant photocurrent method
ACRT	accelerated crucible rotation technique
ACT	advanced Compton telescope
ADX	angular dispersive x-ray diffraction
AECA	anisotropic electrically conductive adhesives
AEM	analytical electron microscope
AES	Auger electron spectroscopy
AFC	antiferromagnetically coupled
AFM	atomic force microscopy
AG	Adam-Gibbs
AHE	anomalous Hall effect
ALD	atomic layer deposition
ALE	atomic layer epitaxy
AMA	active matrix array
AMFPI	active matrix flat-panel imager
AMOLED	active matrix organic light-emitting diode
AMR	anisotropic magnetoresistance
APD	avalanche photodiode
ASTC	Advanced Storage Technology Consortium
AuNp	hybrid graphene-gold nanoparticles
AWG	arrayed-waveguide grating

B

BAR	bit aspect ratio
BE	bound exciton
BEEM	ballistic-electron-emission microscopy
BEOL	back end of line
BER	bit error rate
BET	Brunauer, Emmett and Teller
BIL	burst-illumination LIDAR
BJT	bipolar junction transistor
BLG	bilayer graphene
BLIP	background limited detectors
BPSG	boro-phosphosilicate glass
BSA	bovine serum albumin
BSE	backscattered electron
BSF	basal stacking fault
BZ	Brillouin zone

C

CAIBE	chemically assisted ion beam etching
CARR	compound annual reduction rate
CB	conduction band
CBD	chemical bath deposition
CBE	chemical beam epitaxy
CBED	convergent beam electron diffraction
CC	constant current
CCD	charge-coupled device
CCSS	coclose-spaced sublimation
CCZ	continuous Czochralski method
CDD	charge density difference
CFL	compact fluorescent lamps
CFLPE	container-free LPE
CGC	coupled granular continuous
CGL	charge-generation layer
CHCC	conduction–hole–conduction–conduction
CHSH	conduction–hole–spin–hole
CID	compositional interdiffusion
CIP	current-in-plane
CKR	cross Kelvin resistor
CL	cathodoluminescence
CMOS	complementary metal-oxide-semiconductor
CMP	chemical mechanical polishing
CMS	core multishell
CO	carbon monoxide
COD	catastrophic optical damage
COP	crystal originated particle
CP	charge pumping
CPM	constant photocurrent method
CPP	current-perpendicular-to-plane
CPU	central processing unit
CRA	cast recrystallize anneal
CSD	chemical solution deposition
CSO	chip scale optic
CSP	chip scale packaging
CSS	close-space sublimation
CTE	coefficient of thermal expansion
CTEM	conventional TEM
CTL	charge transport layer
CVD	chemical vapor deposition
CVT	chemical vapor transport
CW	continuous wave
CZ	Czochralski

D

DA	Drude approximation
DAG	direct alloy growth
DAP	donor–acceptor pair

DBE	donor-bound exciton
DBP	dual beam photoconductivity
DBR	distributed Bragg reflector
DC	direct current
DCA	direct chip attach
DCPBH	double-channel planar buried heterostructure
DCXRD	double crystal x-ray diffraction
DEZ	diethylzinc
DFB	distributed feedback
DFT	density functional theory
DH	double heterostructure
DIL	dual-in-line
DiP	dual-in-line-package
DLHJ	double-layer heterojunction
DLPH	double-layer planar heterostructure
DLTS	deep-level transient spectroscopy
DME	detector materials engineering
DMFT	dynamical mean field theory
DMS	dilute magnetic semiconductor
DOS	density of states
DP	Dulong–Petit rule
DPI	dual polarization interferometric
DPN	dip-pen nanolithography
DQE	detective quantum efficiency
DQW	double quantum well
DRAM	dynamic random-access memory
DSC	differential scanning calorimeter
DSIMS	dynamic SIMS
DSP	digital signal processor
DSSR	double split-ring resonator
DTGS	deuterated triglycerine sulfate
DUT	device under test
DVD	digital versatile disk
DW	domain wall
DWB	dual-waveband detector
DWDM	dense wavelength-division multiplexing

E

e-APD	electron avalanche photodiode
EAM	electroabsorption modulator
EBIC	electron beam-induced current
EBL	electron beam lithography
ECA	electrically conductive adhesives
ECB	electrically controlled birefringence
ECC	exchange-coupled-composite
ECR	electron cyclotron resonance
ECRPECVD	electron cyclotron PECVD
EDFA	erbium-doped fiber amplifier
EDMR	electrically detected magnetic resonance
EDS	energy dispersive x-ray spectroscopy
EDX	energy dispersive x-ray
EELS	electron energy loss spectroscopy
EEPROM	electrically erasable programmable read-only memory
EFG	edge-defined film-fed growth
EHP	electron–hole pair
EIA	Electronic Industries Alliance
EL	electroluminescence
ELISA	enzyme-linked immunosorbent assay
ELO	epitaxial lateral overgrowth
ELOG	epitaxial layer overgrowth
EMA	effective media approximation
EMR	extraordinary magnetoresistance
ENDOR	electron–nuclear double resonance
EOT	equivalent oxide thickness
EP	exciton polariton
EPD	etch pit density
EPR	electron paramagnetic resonance
ESA	excited state absorption
ESCA	electron spectroscopy for chemical analysis
ESR	electron spin resonance
ETB	empirical tight-binding approximation
EXAFS	extended x-ray absorption fine structure

F

FAB	fabrication facility
FBG	fiber Bragg grating
FBZ	first Brillouin zone
FCA	free-carrier absorption
fcc	face-centered cubic
FeRAM	ferroelectric random access memory
FET	field-effect transistor
FF	fill factor
FFT	fast Fourier transformation
FHD	flame hydrolysis deposition
FIB	focused ion beam
FIC	fast ion conductor
FinFET	fin field effect transistor
FM	ferromagnetic
FMR	ferromagnetic resonance
FOD	flame oxidation deposition
FOM	figure of merit
FPA	focal plane array
FPD	flow pattern defect
FRAM	ferroelectric random-access memory
FSG	fluorinated silicate glass
FSS	frequency selective surface
FT	Fourier transform
FTIR	Fourier-transform infrared
FTO	fluorine-doped tin oxide
FTPS	Fourier-transform photocurrent spectroscopy
FWHM	full width at half maximum
FX	free exciton
FZ	floating-zone

G

GC	glassy carbon
GDA	generalized Drude approximation
GDMS	glow discharge mass spectrometry
GDOES	glow discharge optical emission spectroscopy
GM	grid method
GMR	giant magnetoresistance

GNR	graphene nanoribbon
GOI	gate oxide integrity
GRL	granular recording layer
GSMBE	gas-source molecular beam epitaxy

H

HAADF	high-angle annular dark field
HAMR	heat assisted magnetic recording
HARP	high-gain avalanche rushing photoconductor
HAZ	heat affected zone
HBT	heterojunction bipolar transistor
hcp	hexagonal close-packed
HDC	horizontal directional solidification crystallization
HDMR	heated dot magnetic recording
HEMT	high electron mobility transistor
HFET	heterojunction field effect transistor
HIP	hot isostatic pressing
HIS	high impedance surface
HMS	head-media separation
HNP	high nitrogen pressure
HOLZ	high-order Laue zone
HOMO	highest occupied molecular orbital
HOT	higher operating temperature
HPB	high-pressure Bridgman
HPD	high-pressure depletion
HREM	high-resolution electron microscopy
HRTEM	high-resolution transmission electron microscopy
HRXRD	high-resolution x-ray diffraction
HSG	hemispherical-grain
HSRS	higher silane-related species
HV	high vacuum
HVPE	hydride vapor phase epitaxy
HWE	hot-wall epitaxy

I

I/O	input/outputs
IBI	isobutyl iodide
IC	integrated circuit
ICSD	inorganic crystal structure database
ICTS	isothermal capacitance transient spectroscopy
IDCA	integrated detector – cooler assembly
IDE	interdigitated electrode
IFIGS	interface-induced gap states
IFTOF	interrupted field time-of-flight
IgG	immunogamma globulin
IMC	integrated magnetic concentrator
IMM	inverted metamorphic structure
IMP	interdiffused multilayer process
IMPATT	impact avalanche transit time
iOLEDS	inverted OLED
iPCM	interfacial phase-change memory
IPEYS	internal photoemission yield spectroscopy
IPS	in-plane switching
IQE	internal quantum efficiency
IR	infrared
ISB	intersubband
ISE	ion-selective electrode
ISHE	inverse spin-Hall effect
ITO	indium tin oxide
IZO	indium zinc oxide

J

JFET	junction field-effect transistor

K

KCR	Kelvin contact resistance
KKR	Kramers–Kronig relation
KWW	Kohlrausch–Williams–Watts

L

LA	longitudinal acoustical
LAST	lead–antimony–silver–telluride
LC	liquid crystal
LCD	liquid crystal display
LD	laser diode
LDA	local density approximation
LDD	lightly doped drain
LEC	liquid encapsulated Czochralski
LED	light-emitting diode
LEEBI	low-energy electron beam irradiation
LEIS	low-energy ion scattering
LEL	lower explosive limit
LEO	lateral epitaxial overgrowth
LF	low frequency
LFN	low-frequency noise
LL	Luttinger liquid
LLS	laser light scattering
LMA	law of mass action
LMR	longitudinal magnetic recording
LMTO	linearized muffin-tin orbitals
LO	longitudinal optical
LOCOS	local oxidation of silicon
LPCVD	low pressure chemical vapor deposition
LPD	light point defect
LPE	liquid phase epitaxy
LPEE	liquid phase electroepitaxy
LSDA	local spin density approximation
LSTD	laser light scattering tomography defect
LTCC	low-temperature co-fired ceramic
LUMO	lowest unoccupied molecular orbital
LWIR	long-wavelength infrared

M

MAA	monomer methacrylic acid
MBE	molecular beam epitaxy
MC	Monte Carlo
MCCZ	magnetic field applied continuous CZ

MCL	magnetic cap layer
MCT	mercury cadmium telluride
MCZ	Czochralski growth with an applied magnetic field
MDF	model dielectric function
MDP	molecularly doped polymer
MDSC	temperature-modulated differential scanning calorimetry
MEED	medium-energy electron diffraction
MEM	micro-electromechanical
MEMS	micro-electromechanical system
MESFET	metal–semiconductor field-effect-transistor
MF	meeting front
MFCs	mass flow controllers
MHJ	mesa heterojunction
MI-M-I-M	metal-insulator-metal-insulator-metal
MIGS	metal-induced gap state
MIM	metal–insulator–metal
MIP	molecularly imprinted polymer
MIPS	millions of instructions per second
MIS	metal–insulator–semiconductor
ML	monolayer
MLC	multilayer ceramic
MLHJ	multilayer heterojunction
MO	metal-organic
MOCVD	metal-organic chemical vapor deposition
MOD	metal-organic deposition
MODFET	modulation-doped FET
MOMBE	metal-organic molecular beam epitaxy
MOPA	master oscillator power amplifier
MOS	metal–oxide–semiconductor
MOSFET	metal–oxide–semiconductor field-effect transistor
MOSS	multi-beam optical stress sensor
MOVPE	metal-organic vapor phase epitaxy
MPB	morphotropic phase boundary
MPC	modulated photoconductivity
MPU	microprocessor unit
MQD	multiple quantum disk
MQW	multiple quantum well
MR	magnetoresistance
MRAM	magnetic random-access memory
MRI	magnetic resonance imaging
MS	metal-semiconductor
MSG	microstructure glass
MTC	minimal thermal conductivity
MTF	modulation transfer function
MTJ	magnetic tunneling junction
MUTIS	metal-ultrathin insulator–semiconductor
MWIR	medium-wavelength infrared
MWNT	multiwall nanotube

N

NBTI	negative-bias temperature instability
NCA	nanochannel alumina
NCG	nanochannel glass
NDM	negative differential mobility
NDR	negative differential resistance
near-IR	near-infrared
NEPh	noise equivalent photons
NETD	noise equivalent temperature difference
NeXT	nonthermal energy exploration telescope
NHE	normal hydrogen electrode
NHF	nonreversing heat flow
NMOS	n-type-channel metal-oxide-semiconductor
NMR	nuclear magnetic resonance
NNH	nearest-neighbor hopping
NREL	National Renewable Energy Laboratory
NTC	negative temperature coefficient
NTCR	negative temperature coefficient of resistance
NTD	neutron transmutation doping
NW	nanowire
NW LED	nanowire light-emitting diode
NWQD	nanowire quantum dot

O

OCV	open circuit potential
ODMR	optically detected magnetic resonance
OLED	organic light-emitting diode
OMVPE	organometallic vapor phase epitaxy
OPSEL	optically pumped surface-emitting laser
OPV	organic photovoltaic device
OSF	oxidation-induced stacking fault
OSG	organo-silicate glass
OSL	optically stimulated luminescence
OZM	overlap zone-melting

P

PA	photoinduced absorption
PACE-I	producible alternative to CdTe epitaxy
PAMBE	plasma-assisted molecular beam epitaxy
PAni	polyaniline
PAS	positron annihilation spectroscopy
PBG	photonic band gap
PCA	principal component analysis
PCB	printed circuit board
PCM	phase-change memory
PDP	plasma display panel
PDS	photothermal deflection spectroscopy
PE BJT	polysilicon emitter bipolar junction transistor
PECVD	plasma-enhanced chemical vapor deposition
PEM	polymer electrolyte membrane
PES	photoemission spectroscopy
PET	positron emission tomography
PGEC	phonon-glass-electron crystal
PICTS	photoinduced current transient spectroscopy
PIDC	photoinduced discharge

PL	photoluminescence
PLC	planar lightwave circuit
PLD	pulsed laser deposition
PM	particulate matter
PMOS	p-type MOS
PMR	perpendicular magnetic recording
POF	plastic optical fiber
ppm	parts per million
PR	photoreflectance
PRAM	phase-change random access memory
PRML	partial response maximum likelihood
PRT	platinum resistance thermometer
PSL	photo-stimulated luminescence
PTC	positive temperature coefficient
PTCR	positive temperature coefficient of resistance
PTS	photon-trapping structure
PV	photovoltaic
PVD	physical vapor deposition
PVT	physical vapor transport
PZC	point of zero charge
PZT	lead zirconate titanate

Q

QA	quench anneal
QC	quantum cascade
QCL	quantum cascade laser
QCM	quartz crystal microbalance
QCSE	quantum-confined Stark effect
QD	quantum dot
QDIP	quantum dot infrared photodetector
QE	quantum efficiency
QFL	quasi-Fermi level
QHE	quantum Hall effect
QW	quantum well
QWIP	quantum well intersubband photodetector

R

RA	resistance × area
RAIRS	reflection adsorption infrared spectroscopy
RBS	Rutherford backscattering spectrometry
RCLED	resonant cavity light emitting diode
RDF	radial distribution function
RDS	reflection difference spectroscopy
RE	rare earth
RED	reflection electron diffraction
RF	radio frequency
RGO	reduced graphene oxide
RH	relative humidity
RHEED	reflection high-energy electron diffraction
RHF	reversing heat flow
RIBE	reactive ion beam etching
RIE	etching reactive ion etching
RIL	resistive interface layer
RIU	refractive index units
RKKY	Ruderman–Kittel–Kasuya–Yoshida
RMS	root mean square
ROIC	readout integrated circuit
RRR	residual resistivity ratio
RTA	rapid thermal annealing
RTD	resistance temperature device
RTS	random telegraph signal

S

S-R-H	Shockley–Read–Hall
SA	self-assembly
SA-VPE	selective-area vapor-phase epitaxy
SAF	synthetic antiferromagnet
SAM	self-assembled monolayer
SAW	surface acoustic wave
SAXS	small-angle x-ray scattering
SCE	saturated calomel electrode
SCH	separate confinement heterostructure
SCLC	space-charge-limited current
SCM	step-coverage method
SCVT	seeded chemical vapor transport
SD	sputtering deposition
SDW	spin density wave
SE	spectroscopic ellipsometry
SEM	scanning electron microscopy
SF	staking fault
SFM	synthetic ferrimagnetic media
SG	selenide glass
SGE	spin-Gunn effect
SHE	spin-Hall effect
SiC	silicon carbide
SIMS	secondary ion mass spectrometry
SiON	silicon oxynitride
SIPBH	semi-insulating planar buried heterostructure
SK	Stranski–Krastanov
SL	superlattice
SL-APD	superlattice avalanche photodiode
SMD	shifted-modulation-doped
SMR	spin-Hall magnetoresistance
SNR	signal-to-noise ratio
SO	small outline
SOA	semiconductor optical amplifier
SOC	system-on-a-chip
SOFC	solid oxide fuel cell
SOI	silicon-on-insulator
SPAD	single-photon avalanche diode
SPC	spin-polarized current
SPE	spin-Peltier effect
SPECT	single-photon emission computed tomography
SPGE	spin photo-galvanic effect
SPR	surface plasmon resonance
SPRI	surface plasmon resonance imaging
SPV	surface photovoltage
SPVT	seeded physical vapor transport
SQW	single quantum well

SRAM	static random-access memory
SRR	split-ring resonator
SSE	spin-Seebeck effect
SSG	superstructure grating
SSIMS	static secondary ion mass spectrometry
SSMBE	solid source molecular beam epitaxy
SSPC	steady-state photoconductivity
SSPG	steady-state photocarrier grating
SSR	solid-state recrystallization
SSRM	scanning spreading resistance microscopy
STEBIC	scanning transmission electron beam induced conductivity
STHM	sublimation traveling heater method
STI	shallow trench isolation
STM	scanning–tunnelling microscopy
STN	super-twisted nematic
STT	spin-transfer torque
SUL	soft underlayer
SUMIT	simultaneous unipolar multispectral integrated technology
SVP	saturated vapor pressure
SWC	spin-wave current
SWCN	single-wall carbon nanotube
SWIR	short-wavelength infrared
SWNT	single-wall nanotube

T

TA	transverse acoustical
TAB	tab automated bonding
TADF	thermally activated delayed fluorescence
TAT	trap-assisted tunneling
TCE	thermal coefficient of expansion
TCO	transparent conducting oxide
TCR	temperature coefficient of resistivity
TCRI	temperature coefficient of the refractive index
TD	threading dislocation
TDD	threading dislocation density
TDDB	time-dependent dielectric breakdown
TDI	time-delay and integration
TE	transverse electric
TEC	thermal expansion coefficient
TED	transmission electron diffraction
TEM	transmission electron microscopy
TEM-CL	TEM-cathodoluminescence
TET	two-electron transition
TFSCLC	trap-free space-charge-limited current density
TFT	thin film transistor
THM	traveling heater method
TL	thermoluminescence
TLC	trap-limited conduction
TLM	transmission line measurement
TM	transverse magnetic
TMR	tunnel magnetoresistance
TN	twisted nematic
TNM	Tool–Narayanaswamy–Moynihan
TO	transverse optical
TOF	time-of-flight
ToFSIMS	time-of-flight SIMS
TOS	transparent oxide semiconductor
TPB	triple phase boundary
TPC	transient photoconductivity
TPV	thermo-photovoltaic
TRAPATT	trapped plasma avalanche-triggered transit
TRMC	time-resolved microwave conductivity
TSL	thermally stimulated luminescence
TVAN	twisted vertically aligned

U

UAPD	unipolar avalanche photodiode
UBM	under bump metallurgy
UD	uniformly doped
UHV	ultrahigh vacuum
UHV-ECR-RMS	ultrahigh vacuum-electron cyclotron resonance-radio-frequency magnetron sputtering
ULSI	ultra-large-scale integration
UMS	ultrasonic micro-spectroscopy

V

VAN	vertically aligned nematic
VAP	valence-alternation pairs
VB	valence band
VCSEL	vertical cavity surface emitting laser
VCZ	vapor-pressure-controlled Czochralski
VD	vacuum distillation
VE	vacuum evaporation
VECSEL	vertical external cavity surface emitting laser
VGA	visual graphics array
VGF	vertical gradient freezing
ViGS	virtual gap states
VIS	visible
VLS	vapor-liquid-solid
VLSI	very-large-scale integration circuit
VOC	volatile organic compound
VOx	vanadium oxide
VPE	vapor phase epitaxy
VRH	variable-range hopping
VS	vapor-solid
VTD	vapor transport deposition
VUV	vacuum ultraviolet
VUVG	vertical unseeded vapor growth
VW	Volmer–Weber

W

WAT	weak absorption tail
WDM	wavelength division multiplexing
WDX	wavelength dispersive x-ray

WEEE	waste in electrical and electronic equipment
WKB	Wentzel–Kramers–Brillouin
WSP	wafer scale packaging
WXI	wide-band x-ray imager

X

XANES	x-ray absorption near-edge structure
XAS	x-ray absorption spectroscopy
XEBIT	x-ray-sensitive electron beam image tube
XGA	extended graphics array
XPS	x-ray photoelectron spectrometry
XRD	x-ray diffraction
XRF	x-ray fluorescence
XRSP	x-ray storage phosphor

Y

YL	yellow luminescence

Z

ZTCR	zero temperature coefficient of resistance

1. Perspectives on Electronic and Photonic Materials

Tim Smeeton, Colin Humphreys

Electronic and photonic materials have a tremendous impact on the modern world. They include a wide range of material classes and are developed through a deeply interdisciplinary combination of physics, chemistry, materials science, and engineering. In this introductory chapter, we give some perspectives on this exciting and ever-changing field. We give an example of the tremendous integration of different materials used in today's consumer products, and then take a historical look at the development of some key semiconductor materials and devices from inception to today. Focusing in particular on the development of the transistor and integrated circuit and some of the key electronic and photonic applications of compound semiconductors, we take advantage of the long-distance view to point out some unifying themes across the wide portfolio of materials while appreciating their unique features.

1.1 **Tremendous Integration** 2
1.2 **The Silicon Age** 3
1.2.1 The Transistor and Early Semiconductor Materials Development 3
1.2.2 The Integrated Circuit 5
1.3 **The Compound Semiconductors** 7
1.3.1 High-Speed Electronics 8
1.3.2 Light Emitting Devices 9
1.3.3 The III-Nitrides 11
References 14

It can be easy to forget how remarkable electronic and photonic materials are. Take the light emitting diode (LED) as an example: An electric current is passed through a tiny stack of layers of slightly different materials and brilliant colored light is emitted. Of course, the LED is carefully designed, and there are powerful theories to explain the behavior, but that should not detract from the initial moment of wonderment that it works at all. Or that a room can be lit by photons generated in the active region of an LED which includes just a few thousand cubic micrometers of material – about the same volume as a grain of flour.

We all literally *see* the output from LEDs every day, from television screens, vehicle lights, or lighting luminaires. Most other electronic and photonic materials and devices are less conspicuous to our senses but together they have played an indisputable role in defining the way people live, work, and communicate in the twenty-first century: from microprocessors containing billions of transistors which provide immense computational power to laser diodes and radiofrequency transceivers which enable trans-global and wireless communication. These complex devices are only available today thanks to the work of countless researchers over the past century who have identified, designed and understood a vast array of materials, produced them with extraordinary purity, quality and economy, and deployed them in device designs which harness their power.

The detailed chapters in this handbook provide comprehensive introductions to the huge range of technical fields which electronic and photonic materials now occupy, written by world experts. In this introductory chapter, we have the luxury of stepping back from the details of these complex fields and reviewing the whole, in search of some perspective. First, we take a look at the complex mixture of electronic and photonic materials which are in use today and reflect on some common themes. Next, the majority of the chapter is dedicated to a review of the development of a handful of the most important semiconductor materials and devices since the early breakthroughs of the 1940s: the birth of electronics in germanium and silicon; the integrated circuit (IC); some of the key electronic and optoelectronic uses of III–V semiconductors; and the recent rise of III-nitride materials. These reviews serve as a tribute to breakthroughs of the last 75 years, hopefully providing some perspective of how the science and technology of these materials has come to its current state and perhaps providing some encouragement and inspiration to those who are striving to develop materials today.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_1

1.1 Tremendous Integration

A starting point for perspective is to appreciate the complex mixture of materials and devices that are routinely integrated together. For a case study, we need look no further than something which is probably within touching distance as you read this sentence: a smartphone. One of the defining technologies of the twenty-first century, it has been enabled by an astonishing convergence of electronic and photonic materials and devices. No two models are quite the same, but let us teardown a hypothetical handset and its supporting infrastructure, laying forward references to the detailed chapters in this book as we go. As a user, we see text, images and video from a high-resolution liquid crystal display (LCD) in which blue light emitted from indium gallium nitride (InGaN) LEDs (Chaps. 31, 35, and 40) excites a rare-earth-doped yttrium aluminum garnet (YAG) phosphor (Chap. 38) to produce white light which, in turn, is color-filtered into red, green, and blue components and modulated at the pixel level by liquid crystal cells (Chap. 36) controlled by silicon or indium gallium zinc oxide (IGZO) thin film transistors (TFTs) (Chap. 44). We have touch control thanks to capacitive sensors which use an optically transparent and electrically conducting oxide such as indium tin oxide (ITO) (Chap. 58). Low power consumption multicore processors, RAM, and flash memories – all silicon – provide the computing power and storage (Chap. 21). The device is powered by a high energy-density lithium ion battery (Chap. 11). Wireless connectivity is provided between gallium arsenide (GaAs) based transistor radio transceivers on the handset and base stations equipped with silicon, silicon carbide (SiC), or aluminum gallium nitride (AlGaN) transistor amplifiers. Beyond that, data communication – which gives the smartphone its purpose – is via silica (SiO_2) optical fibers (Chap. 41) using indium phosphide (InP)-based laser diodes, optical receivers, switches and amplifiers (Chap. 35) to data centers based on high-performance silicon processors. And we can go further still: The data centers consume electrical power distributed through more power transistors (silicon, SiC and AlGaN) and perhaps originally generated by crystalline silicon, amorphous silicon, copper indium gallium selenide (CIGS) or cadmium telluride (CdTe) photovoltaic cells (Chap. 43).

The list of core materials is already well into double figures, but we have still only scratched the surface. If we consider just the families of semiconductor components in a little more detail, these include multiple alloy compositions (Chap. 30) with optimized combinations of impurity doping (Chap. 2). Not even to mention dozens – probably hundreds – of different electrode, interconnect, dielectric and packaging (Chaps. 27, 29, and 53) materials.

We could go on, but surely the point is made: A remarkably complex combination of materials and devices is involved. It is a phenomenal achievement of technology and business – we could argue a wonder of the modern world – that this dense concentration of different materials and technologies can be brought together with such synergy in a reliable, mass-produced product weighing less than 150 g that can be sold at a price accessible to billions of people. It is often the product design or the software of these devices which steal the limelight in marketing, but it is the electronic and photonic materials which are the unassuming heroes!

Electronic and photonic materials include almost all types and forms of material:

- Inorganic compounds (semiconductors, dielectrics)
- Organic molecules (liquid crystals, semiconductors)
- Ionic crystals (batteries), polymers (packaging, lithography masks)
- Metals/alloys (electrodes, interconnect wiring)
- Single crystal
- Polycrystalline
- Amorphous
- Thin film
- Nanostructured
- Two-dimensional.

Despite this breadth, there is at least one strong unifying theme across all: the need for extraordinary precision in the design and fabrication of the materials. These materials demand combinations of high purities, low defect densities, and precise chemical composition that probably exceed the requirements of any other area of technology. These requirements have been met by a unique convergence of physics, chemistry, materials science and engineering which brings together the core theoretical understanding with the means for fabrication.

Two aspects, in particular, are vital for delivering the necessary precision: exact material fabrication methods and detailed understanding of material structure and properties. In many cases, the ultimate device performance is determined by how far we fall short of fabricating the *perfect* material. If the number of crystalline defects is too high, or if there are too many impurities, or if an interface is too rough, or if ..., a device may fail. Minute deviations from a target specification can have catastrophic effects. Taking LED or laser diode fabrication as an example, a change

in composition of material or doping accounting to much less than 1% of a layer just a few nanometers thick can be the difference between a viable device and a costly failure. In nearly all device fabrication steps, from bulk growth methods for fabrication of substrates with near-perfect quality, to thin film growth methods to deposit precise material compositions with atomic-scale precision, to processing technology to pattern device features at the nanometer scale, and through to packaging and integration methods, material production techniques are the lifeblood of the field. The precision of fabrication techniques needs to be at least matched by the precision with which we can determine the structure of the resulting materials. This means that the structural and chemical characterization tools and techniques are an inseparable aspect of the story of all electronic and photonic materials. Accurate understanding of the quality (or otherwise) of the materials that we make – obtained from microscopy or other tools – has been a continuous driver or enabler for progress in material and device development.

When the material fabrication expertise and understanding of structure and properties are combined with deep understanding of the physics of device design, the result is the remarkable devices that can enrich our lives.

So, today's electronic and photonic materials span a tremendous range of fields and disciplines. How did this myriad of technologies successfully emerge and grow? Why have the materials that are in use today become dominant over others? In the next sections, we will give some very selective answers to these questions, hoping to provide a perspective of how some technologies have reached the current state. In particular, we have chosen to focus on a few sectors of semiconductor development to explain some of the early development of what became mainstream technologies. Often the key messages are learnt in these early years – before a technical field fragments into highly specialized sectors – and may provide the most valuable perspective.

1.2 The Silicon Age

1.2.1 The Transistor and Early Semiconductor Materials Development

Electronics is the control of electrons to produce useful properties; electronic materials are the media in which this manipulation takes place. The foundations for today's electronics were laid in 1947 – 50 years after J.J. Thompson had discovered the electron – with the first demonstration of the semiconductor transistor effect. Building on quantum theories to explain semiconductor behavior from the 1930s [1.1, 2] and improvements in purity of candidate semiconductor materials, John Bardeen and Walter Brattain used germanium to build and demonstrate the first *semiconductor triode*. Later, to be named the point contact transistor to reflect its transresistive properties, this success at Bell Laboratories was obtained just a few years after a research group led by William Shockley was established to focus on understanding semiconducting materials. It was to earn Brattain, Bardeen, and Shockley the 1956 Nobel Prize for Physics.

The first point-contact transistor was based around three contacts onto an n-doped germanium block: When a small current passed between the *base* and *emitter*, an amplified current would flow between the *collector* and *emitter* [1.3]. The emitter and collector contacts needed to be located very close to one another (50–250 μm), and this was achieved by evaporating gold onto the corner of a plastic triangle, cutting the film with a razor blade, and touching this onto the germanium – the two isolated strips of gold serving as the two contacts [1.4]. At about 1 cm in height, based on relatively impure polycrystalline germanium and adopting a different principle of operation, the device bears barely any resemblance to today's IC electronics components. Nonetheless, it was the first implementation of a solid-state device capable of modulating (necessary for signal amplification in communications) and switching (needed for logic operations in computing) an electric current. In a world whose electronics were delivered by the thermionic vacuum tube, the transistor was immediately identified as a component which could be *employed as an amplifier, oscillator, and for other purposes for which vacuum tubes are ordinarily used* [1.3].

In spite of this, after the public announcement of the invention at the end of June 1948, the response of both the popular and technical press was somewhat muted. It was after all still *little more than a laboratory curiosity* [1.5] and ultimately point-contact transistors were never suited to mass production. The individual devices differed significantly in characteristics, the noise levels in amplification were high and they were rapidly to be superseded by improved transistor types.

A huge range of transistor designs have been introduced from the late 1940s to now. These successive generations either drew upon or served as a catalyst

for a range of innovations in semiconductor materials processing and understanding. There are many fascinating differences in device design but from a materials science point of view the three most striking differences between the first point-contact transistor and the majority of electronics in use today are the choice of semiconductor, the purity of this material, and its crystalline quality. Many of the key electronic materials technologies of today derive from the developments in these fields in the very early years of the semiconductor industry.

Both germanium and silicon had been produced with increasing purity throughout the 1940s [1.6]. Principally because of germanium's lower melting temperature (937 °C compared with 1415 °C) and lower chemical reactivity, its preparation had always proved easier and was therefore favored for the early device manufacture such as the first transistor. However, the properties of silicon make it a much more attractive choice for solid-state devices. Although germanium is expensive and rare, silicon is, after oxygen, the second most abundant element. Silicon has a higher breakdown field and a greater power handling ability; its semiconductor band gap (1.1 eV at 300 K) is substantially higher than that of germanium (0.7 eV), so silicon devices are able to operate over a greater range of temperatures without intrinsic conductivity interfering with performance.

These two materials competed with one another in device applications until the introduction of novel doping techniques in the mid-1950s. Previously, p- and n-doping had been achieved by the addition of dopant impurities to the semiconductor melt during solidification. A far more flexible technique involved the diffusion of dopants from the vapor phase into the solid semiconductor surface [1.7]. It became possible to dope with a degree of two-dimensional precision when it was discovered that silicon's oxide served as an effective mask to dopant atoms and that a photoresist could be used to control the etching away of the oxide [1.8, 9]. Successful diffusion masks could not be found for germanium and it was soon abandoned for mainstream device manufacturing. Dopant diffusion of this sort has since been superseded by the implantation of high-energy ions which affords greater control and versatility.

Shockley was always aware that the material of the late 1940s was nothing like pure enough to make reliable high performance commercial devices. Quantum mechanics suggested that to make a high quality transistor out of the materials, it was necessary to reduce the impurity level to about one part in 10^{10}. This was a far higher degree of purity than existed in any known material. However, William Pfann, who worked at Bell Laboratories, came up with the solution. He invented a technique called zone-refining to solve this problem, and showed that repeated zone refining of germanium and silicon reduced the impurities to the level required. The work of *Pfann* is not widely known but was a critical piece of materials science that enabled the practical development of the transistor [1.10, 11].

At a similar time, great progress was being made in reducing the crystalline defect density of semiconducting materials. Following initial hostility by some of the major researchers in the field, it was rapidly accepted that transistor devices should adopt single crystalline material [1.12]. Extended single crystals of germanium several centimeters long and up to 2 cm in diameter [1.12, 13] and later similar silicon crystals [1.14] were produced using the *Czochralski* technique of pulling a seed crystal from a high purity melt [1.15]. The majority of material in use today is still derived from this route. To produce silicon with the very lowest impurity concentration, an alternative method called float zoning was developed where a polycrystalline rod was converted to a single crystal by the passage of a surface tension confined molten zone along its length [1.16–18]. No crucible is required in the process, so there are fewer sources of impurity contamination. Float zoning is used to manufacture some of the purest material in current use [1.19]. The early Czochralski material contained dislocation densities of $10^5 - 10^6\,\mathrm{cm}^{-2}$ but by the start of the 1960s dislocation free material was obtained [1.20–23]. Initially, most wafers were on the silicon (111) plane, which was easiest to grow, cut, and polish [1.24]. For field-effect devices, which are discussed below, use of the (100) plane was found to offer preferable properties, so this was introduced in the same decade. The impurity concentration in dislocation-free silicon has been continually reduced up to the present day and wafer diameters have increased almost linearly (though accelerating somewhat in recent years) from about 10 mm in the early 1960s to the *dinner plate* 300 mm which dominates today [1.19] and with 450 mm tentatively in the pipeline within the next decade. These improvements represent one of the major achievements in the growth and processing of semiconductor materials.

A series of generations of transistors followed in rapid succession after Brattain and Bardeen's first triumph. Here, we only mention a few of the major designs whose production has traits in common with technology today. Early in 1948, *Shockley* developed a detailed formulation of the theory of p–n junctions that concluded with the conception of the junction transistor [1.25, 26]. This involved a thin n-doped base layer sandwiched between p-doped emitter and collector layers (or vice versa). This p–n–p (n–p–n) structure is the

simplest form of the bipolar transistor (so-called because of its use of both positive and negative charge carriers), a technology that remains important in analogue and high-speed digital ICs today. In April 1950, by successively adding arsenic and gallium (n- and p-type dopants, respectively) impurities to the melt, n–p–n junction structures with the required p-layer thickness ($\approx 25\,\mu m$) were formed from single crystal germanium. When contacts were applied to the three regions, the devices behaved much as expected from *Shockley's theory* [1.25, 27]. Growth of junction transistors in silicon occurred shortly afterwards and they entered production by Texas Instruments in 1954 [1.12].

By the later years of the 1950s, the diffusion doping technique was used to improve the transistor's speed response by reducing the thickness of the base layer in the diffused base transistor [1.28]. This began the trend of manufacturing a device in situ on a substrate material, so, in a sense, it was the foundation for all subsequent microelectronic structures. Soon afterwards, epitaxial growth techniques were introduced (what would today be described as vapor phase epitaxy (VPE)) which have since become central to both silicon and compound semiconductor technology. Gas phase precursors were reacted to produce very high quality and lightly doped crystalline silicon on heavily doped substrate wafers to form epitaxial diffused transistors. Even though the collector contact was made through the thickness of the wafer, the use of highly doped (low resistance) wafers reduced the series resistance and therefore increased the frequency response [1.29].

For some years, the highest performance devices were manufactured using the so-called *mesa* process where the emitter and diffused base were raised above the collector using selective etching of the silicon [1.25]. The planar process (which was at the heart of nearly all device production until the last few years) was subsequently developed, in which the p–n junctions were all formed inside the substrate using oxide masking and diffusion from the surface. This resulted in a flat surface to which contacts could be made using a patterned evaporated film [1.30]. This processing technique was combined with some exciting thoughts at the end of the 1950s and led to the application of transistor devices and other components in a way which was to transform the world, that is, the IC.

1.2.2 The Integrated Circuit

With the benefit of hindsight, the IC concept is quite simple. The problem faced by the electronics industry in the 1950s was the increasing difficulty of physically fitting into a small device all of the discrete electronic components (transistors, diodes, resistors, and capacitors), and then connecting them together. It was clear that this problem would eventually limit the complexity, reliability, and speed of circuits which could be created. Transistors and diodes were manufactured from semiconductors but resistors and capacitors were best formed from alternative materials. Even though they would not deliver the levels of performance achievable from the traditional materials, functioning capacitors and resistors *could* be manufactured from semiconductors, so, in principle, all of the components of a circuit could be prepared on a single block of the semiconducting material. This reasoning had been proposed by Englishman G.W.A. Dummer at a conference in 1952 [1.31], but small-scale attempts to realize circuits had failed, largely because they were based on connecting together layers in grown-junction transistors [1.32]. In 1958, however, Jack Kilby successfully built a simple oscillator and *flip–flop* logic circuits from components formed in situ on a germanium block and interconnected to produce circuits. He received the Nobel Prize in 2000 for *his part in the invention of the IC*.

Kilby's circuits were the first built on a single semiconductor block, but by far the majority of the circuit's size was taken up by the wires connecting together the components. Robert Noyce developed a truly IC in the form that it was later to be manufactured. While Kilby had used the mesa technique with external wiring, Noyce applied the planar technique to form transistors on silicon and photolithographically defined gold or aluminum interconnects. This was more suited to batch processing in production and was necessary for circuits with large numbers of components.

In the 1960s, a new transistor design – the metal oxide semiconductor field effect transistor (MOSFET) – was introduced that replaced the bipolar device used in the first chips [1.33]. In this device, a *gate* was deposited onto a thin insulating silicon oxide layer on the silicon. The application of a voltage to the gate resulted in an inversion layer in the silicon below the oxide, thereby, modifying the conducting channel between *source* and *drain* contacts. This structure was a p-MOS device (current transfer between the collector and emitter was by hole conduction) grown on (111) silicon using an aluminum gate. Earlier attempts at such a device had failed because of trapped impurities and charges in the gate oxide – this new structure had reduced the density of these to below tolerable levels but the device still could not compete with the bipolar transistors of its time [1.24]. By 1967, however, (100) silicon (which offered lower densities of states at the Si/SiO_2 interface) was used together with a polycrystalline silicon gate to construct a more effective and more easily processed device with advantages over the bipolar transistor. In the early 1970s, the n-MOS device,

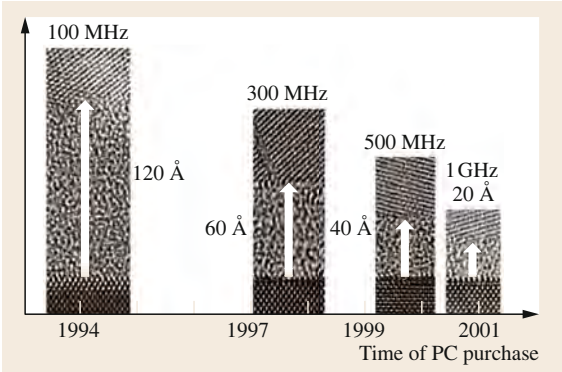


Fig. 1.1 The relentless march to zero thickness. HRTEM images of SiO_2 gate oxide thickness used in personal computers before high-k gate oxides were introduced. (After [1.34], with permission from Cambridge University Press)

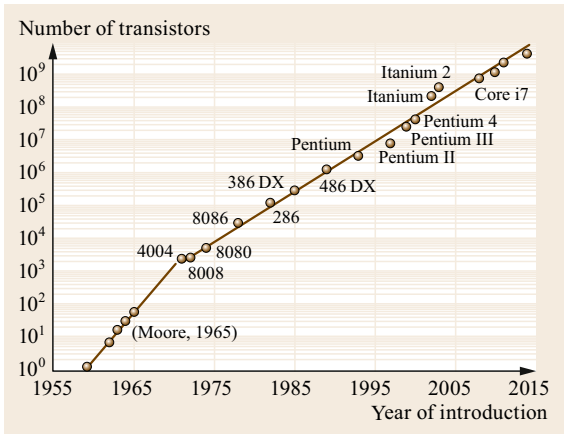


Fig. 1.2 The realization of Moore's law by commercial Intel processors: the logarithmic increase in the number of transistors in each processor chip

which was even less tolerant of the positive gate oxide charges, was realized thanks to much improved cleanliness in the production environment. With conduction occurring by the transfer of electrons rather than holes, these were capable of faster operation than similar p-MOS structures (the mobility of electrons in silicon is about three times that of holes). By the 1980s, these two devices, were combined in the complementary MOS (CMOS) device which afforded much lower power consumption and simplified circuit design [1.35].

The MOSFET is still a common structure in microelectronics today. Of course, now it can be much smaller, with transistor dimensions well below 100 nm, enabling more transistors within a single chip. The smaller dimensions are achieved through improvements in optical lithographic patterning methods. The minimum dimension of components which can be lithographically patterned on an IC is ultimately limited by the wavelength of radiation used in the process and this has continually been decreased over the past few decades. In the late 1980s, wavelengths of 365 nm were employed; by the late 1990s, 248 nm were common; and today, 193 nm is being used; sometimes coupled with *double-patterning* processes to extend to the smallest dimensions. Research into extreme ultraviolet lithography (e.g., at 13.5 nm wavelength) continues for potential use in future generation processes.

Providing good performance from transistors with smaller dimensions required significant advances in materials. For example: better control of doping; use of copper interconnects instead of aluminum; and *low-k* dielectric insulating layers between interconnects instead of silicon dioxide. Perhaps, the most disruptive modification was the replacement of the *native* silicon oxide as the gate insulator by *high-k* dielectric material such as hafnium oxide (Chap. 27). Smaller MOSFET dimensions had required ever thinner silicon gate oxide thicknesses – as shown clearly in Fig. 1.1 – eventually culminating in gate insulator layers just a few atoms thick; sufficiently thin that electron quantum mechanical tunneling through it became a significant power drain. Use of *high-k* material such as hafnium oxide, which has a significantly higher dielectric constant than silicon oxide ($k \approx 22$, compared with $k \approx 3.9$), enabled a return to thicker, tunneling-resistant, gate insulators for devices launched in the early 2000s. This improvement required huge commitment in research and development, including development and production scale-up of new methods to deposit sufficiently perfect *high-k* material (e.g., atomic layer deposition, ALD).

A most recent development in transistor technology has seen the abandonment of the traditional planar geometry and the first widespread use of three-dimensional (3-D) transistor geometries to further reduce leakage problems associated with the extreme proximity of source and drain. For example, the fin field effect transistor (FinFET), where the transistor channel is formed in a several nanometers wide ridge that protrudes above the planar substrate surface and is covered on its three exposed edges by gate material.

The development in complexity and performance of silicon devices, significantly due to materials science progress, is unparalleled in the history of technology. Never before could improvements be measured in terms of a logarithmic scale for such a sustained period. This is often seen as the embodiment of *Moore's law*. Noting a doubling of the number of components fitted onto ICs each year between 1959 and 1965, *Moore* predicted that this rate of progress would continue until at least

10 years later [1.36]. From the early 1970s, a modified prediction of doubling the number of components every couple of years has been sustained to the current day. Since the goals for innovation have often been defined assuming the continuation of the trend, it should perhaps be viewed more as a self-fulfiling prophecy. A huge variety of statistics relating to the silicon microelectronics industry follow a logarithmically scaled improvement from the late 1960s to the current day: the number of transistors shipped per year (increasing); average transistor price (decreasing); and number of transistors on a single chip (increasing) are examples [1.37]. The final member of this list is plotted in Fig. 1.2.

1.3 The Compound Semiconductors

It has been said that silicon is to electronics what steel is to mechanical engineering [1.38]. Steel is very effectively used for most of the world's construction but there are some tasks which it is incapable of performing and others for which alternative structural materials are better suited. In the same way, there are some crucial applications – such as optoelectronics and very high speed electronics – that silicon cannot usually deliver but which a wide range of compound semiconductors are better equipped to perform.

Silicon's band gap is indirect (an electron–hole recombination across the band gap must be accompanied by an interaction with a phonon in the lattice) that severely limits the potential efficiency of light emission from the material. Many of the important compound semiconductors, such as the alloys $Al_xGa_{1-x}As$, $In_xGa_{1-x}N$, $Al_yGa_{1-y}N$, and $Al_xIn_yGa_{1-x-y}P$, exhibit direct band gaps (no phonon interaction is required), so they can efficiently emit brilliant light in LEDs and laser diodes. Furthermore, in these alloy systems, where the band gap can be adjusted by changing the composition, there is a means of selecting the energy released when an electron and hole recombine across the gap and therefore controlling the wavelength of the photons emitted. From the $Al_xIn_yGa_{1-x-y}As$ and $Al_yIn_xGa_{1-x-y}N$ alloy systems, there is, in principle, a continuous range of direct band gaps from deep in the infrared (InAs; $\lambda = 3.5\,\mu m$) to far into the ultraviolet (AlN; $\lambda = 200\,nm$). The semiconductor band gaps of these materials and the corresponding photon wavelengths are put into context with the visible spectrum in Fig. 1.3.

Compounds are also very useful in high speed electronics applications. One of the determining factors in the speed of a transistor is the velocity of the charge carriers in the semiconductor. In GaAs, the electron drift velocity is much higher than in silicon, so its transistors are able to operate at significantly higher frequencies. The electron velocity in InAs is higher still. Furthermore, in the same way that silicon was preferred to germanium, devices manufactured using semiconductors such as GaN, which have much wider band gaps than silicon (3.4 eV compared with 1.1 eV), are capable of operating in much higher temperature environments.

Apart from these advantageous properties of compound semiconductors, the use of different alloy compositions, or totally different semiconductors, in a single device introduces entirely new possibilities. In silicon, most device action is achieved by little more than careful control of dopant impurity concentrations. In structures containing thin layers of semiconductors with different band gaps (heterostructures), there is the potential to control more fundamental parameters such as the band gap width, mobilities, and effective masses of the carriers [1.38]. In these structures, important new features become available which can be used by the device designer to tailor specific desired proper-

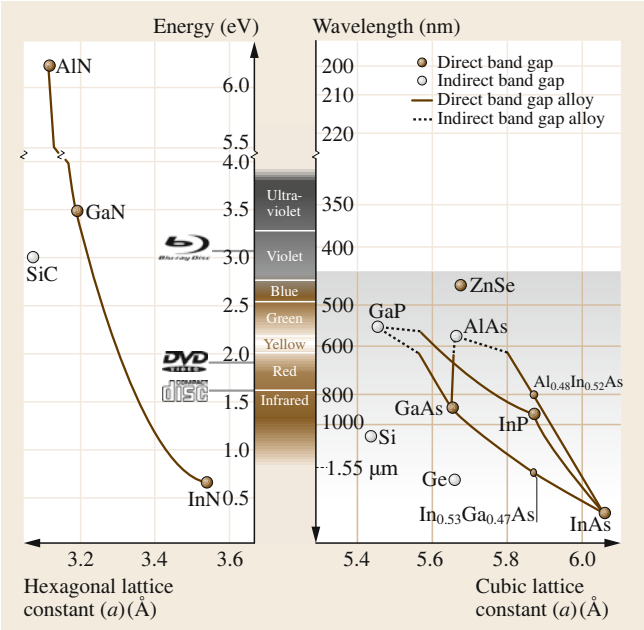


Fig. 1.3 Parameter perspective: band gaps and lattice parameters of selected semiconductors. The important wavelengths for optical storage (CD, DVD, and Blu-Ray) and the 1.55 μm used for efficient data transmission through optical fibers are labeled

ties. Hebert Kroemer and Zhores Alferov shared the Nobel prize in 2000 *for developing semiconductor heterostructures used in high-speed- and optoelectronics*.

We will mainly consider the compounds formed between elements in group III of the periodic table and those in group V (the III–V semiconductors); principally those based around GaAs and InP which were developed over much of the last 40 years, and GaN and its related alloys which have been most heavily studied only during the last two decades. Other families are given less attention here, though they also have important applications (e.g., the II–VI materials in optoelectronic applications). It can be hazardous to try and consider the *compound semiconductors* as a single subject. Though lessons can be learnt from the materials science of one of the compounds and transferred to another, each material is unique and must be considered on its own (i. e., of course, the purpose of the specialized chapters which follow in this handbook).

It is worth repeating that the power of the compound semiconductors lies in their use as the constituent layers in heterostructures. The principal contribution from chemistry and materials science to enable successful devices has been in the manufacture of high-quality bulk single crystal substrates and the development of techniques to reliably and accurately produce real layered structures on these substrates from the plans drawn up by a device theorist. In contrast to silicon, the compound semiconductors include volatile components, so encapsulation has been required for the synthesis of low-defect InP and GaAs substrates such as in the liquid encapsulated Czochralski technique [1.39, 40]. The size and crystalline quality of these substrates lag some way behind those available in silicon. Crucial to the commercialization of electronic and optoelectronic heterostructures were the improvements over the last few decades in the control of epitaxial growth available to the crystal grower. The first successful heterostructures were manufactured using deposition onto a substrate from the liquid phase (liquid phase epitaxy (LPE)) – *a beautifully simple technology but with severe limitations* [1.38]. However, the real heterostructure revolution had to wait for the 1970s and the introduction of molecular beam epitaxy (MBE) and metallorganic chemical vapor deposition (MOCVD) – also known as metallorganic vapor phase epitaxy (MOVPE) provided that the deposition is epitaxial.

MBE growth occurs in an ultrahigh vacuum with the atoms emitted from effusion cells forming *beams* which impinge upon and form compounds at the substrate surface. It derives from pioneering work at the start of the 1970s [1.41]. MOCVD relies on chemical reactions occurring on the substrate involving metallorganic vapor phase precursors and also stems from initial work at this time [1.42]. In contrast to LPE, these two techniques permit the combination of a wide range of different semiconductors in a single structure and offer a high degree of control over the local composition, in some cases on an atomic layer scale. The successful heterostructure devices of the late 1970s and 1980s would not have been achievable without these two tools and they still dominate III–V device production and research today.

1.3.1 High-Speed Electronics

The advantages of the III–V materials over silicon for use in transistors capable of operating at high frequencies were identified early in the semiconductor revolution [1.43]. Shockley's first patent for p–n junction transistors had included the proposal to use a wide-gap emitter layer to improve performance, and in the 1950s, *Kroemer* presented a theoretical design for a heterostructure transistor [1.44]. Some years later, the structure of a GaAs metal semiconductor FET (MESFET) was proposed and realized soon afterwards [1.45, 46]. In these devices, a Schottky barrier surface potential was used to modulate the conductivity of the GaAs channel. One of the earliest applications of the III–Vs was the low-noise amplifiers in microwave receivers that offered substantial improvements relative to the silicon bipolar transistors of the time. The devices were later used to demonstrate subnanosecond switching in monolithic digital ICs [1.47]. Today, they form the core of the highest speed digital circuits and are used in high speed electronics in microwave radar systems and wireless communications which incorporate monolithic ICs.

For at least 30 years, there have been repeated attempts to replicate the MOSFET, the dominant transistor form in silicon ICs, on GaAs material. These attempts have been frustrated by the difficulty of reproducibly forming a high quality stoichiometric oxide on GaAs. In direct analogy with the initial failure of constructing working n-MOSFETs on silicon, the GaAs devices have consistently been inoperable because of poor quality gate oxides with a high density of surface states at the GaAs–insulator interface [1.40]. One of the research efforts focused on realizing this device was, however, to be diverted and resulted in the discovery of probably the most important III–V electronic device, that is, the high electron mobility transistor (HEMT).

The background to this invention lies in the beautiful concept of modulation doping of semiconductors which was first demonstrated in 1978 [1.48]. One of the tenets of undergraduate semiconductor courses is the demonstration that as the dopant density in a semiconductor increases, the mobility of the carriers is reduced,

because the carriers are scattered more by the ionized dopants. It was found that in a multilayer of repeating n-AlGaAs layers and undoped GaAs layers, the electrons supplied by donor atoms in the AlGaAs, moved into the adjacent potential wells of the lower-band gap GaAs layers. In the GaAs these suffered from substantially less ionized impurity scattering and therefore demonstrated enhanced mobility.

While working in a group attempting to create GaAs MOSFETs (and seemingly despairing at the task [1.49]), *Takashi Mimura* heard of these results and conceived of a field effect transistor where the conducting channel exploited the high mobility associated with a modulation doped structure. In essence, a doped AlGaAs layer was formed above the undoped GaAs channel of the transistor. Donated carriers gathered in the GaAs immediately below the interface where they did not suffer from as much ionized impurity scattering and so their mobility would approach that of an ultrapure bulk semiconductor. The current was conducted from the source to the drain by these high mobility carriers and so the devices were able to operate in higher frequency applications [1.49]. Realization of the structure required a very abrupt interface between the GaAs and AlGaAs and was considered beyond the capability of MOCVD of the time [1.49]. However, following the advances made in MBE procedures during the 1970s, the structure was achieved by that technique within a few months of the original conception [1.49, 50]. The first operational HEMT chips were produced on December 24, 1980: By pleasing coincidence this was the anniversary of Brattain and Bardeen's demonstration of their point-contact resistor to the management of Bell Labs in 1947. Structures based on the same principle as Mimura's device were realized in France very shortly afterwards [1.51].

The commercialization of the HEMT became significant in the late 1980s, thanks to broadcasting satellite receivers. The improved performance of the devices compared with the existing technology allowed the satellite parabolic dish size to be reduced by at least a factor of two. Structures similar to these have since played a crucial role in the massive expansion in mobile telephones.

The evolution in HEMT structures since the early 1980s is a fine example of how fundamental compound semiconductor properties have been exploited as the technology has become available to realize new device designs. The electron mobility in InAs is much higher than that in GaAs and rises as the indium content in $In_xGa_{1-x}As$ is increased [1.52]. The introduction of an InGaAs (as opposed to GaAs) channel to the HEMT structure, to create the so-called pseudomorphic HEMT (pHEMT), resulted both in increased electron mobility and a higher density of carriers gathering from the doped AlGaAs layer (because of the larger difference in energy between the conduction band minima of InGaAs and AlGaAs than that between GaAs and AlGaAs) [1.39]. The indium content and thickness of the channel are limited by the lattice mismatch with the GaAs (Fig. 1.3). If either is increased too much, then misfit dislocations are formed within the channel. The restriction is reduced by growing lattice matched structures on InP, rather than GaAs, substrates. $Al_{0.48}In_{0.52}As$ and $In_{0.53}Ga_{0.47}As$ are both lattice matched to the InP (Fig. 1.3) and their conduction band minimum energies are well separated, so that in the InGaAs below the interface between the two compounds a high density of electrons with a very high mobility is formed. Compared to the pHEMTs, these InP-based HEMTs exhibited significant improvements, were shown to exhibit gain at over 200 GHz, and became established as the leading transistor for millimeter-wave low noise applications such as radar [1.39].

1.3.2 Light Emitting Devices

LEDs and laser diodes exploit the direct band gap semiconductors to efficiently convert an electric current into photons of light. Work on light emission from semiconductor diodes was carried out in the early decades of the twentieth century [1.53], but the start of the modern era of semiconductor optoelectronics traces from the demonstration of lasing and LED behavior from p–n junctions in GaAs [1.54, 55] and $GaAs_{1-x}P_x$ [1.56]. The efficiency of these LEDs was low and the lasers had large threshold currents and only operated at low temperatures. A year later, in 1963, *Kroemer* and *Alferov* independently proposed the concept of the double heterostructure (DH) laser [1.57, 58]. In the DH device, a narrow band gap material was to be sandwiched between layers with a wider gap so that there would be some degree of confinement of carriers in the *active layer*. By the end of the decade, DH devices had been constructed that exhibited continuous lasing at room temperature [1.59, 60]. Alferov's laser was grown by LPE on a GaAs substrate with a 0.5 μm GaAs active layer confined between 3 μm of $Al_{0.25}Ga_{0.75}A$ on either side. The launch of the compact disk in 1982 saw this type of device, or at least its offspring, becoming taken for granted in the households of the world.

One of the major challenges in materials selection for heterostructure manufacture has always been avoiding the formation of misfit dislocations to relieve the strain associated with lattice parameter mismatch between the layers. $Al_xGa_{1-x}As$ exhibits a direct band gap for $x < 0.45$ and the early success and sustained

dominance of the AlGaAs/GaAs system derive significantly from the very close coincidence of the AlAs and GaAs lattice parameters (5.661 Å and 5.653 Å; Fig. 1.3). This allows relatively thick layers of AlGaAs with reasonably high aluminum content to be grown lattice matched onto GaAs substrates with no misfit dislocation formation. The use of the quaternary alloy solid solution $In_xGa_{1-x}As_yP_{1-y}$ was also suggested in 1970 [1.59] to offer the independent control of lattice parameters and band gaps. Quaternaries based on three group III elements have since proved very powerful tools for lattice matching within heterostructures. $(Al_xGa_{1-x})_{0.5}In_{0.5}P$ was found to be almost perfectly lattice matched to GaAs and additionally have a very similar thermal expansion coefficient (which is important to avoid strain evolution when cooling after growth of heteroepitaxial layers at high temperatures). By varying x in this compound, direct band gaps corresponding to light between red and green could be created [1.61]. Lasers based on this alloy grown by MOCVD are a common choice for the red wavelengths (650 nm) used in DVD reading (Fig. 1.3).

Obtaining lattice matching is not so crucial for layers thinner than the critical thickness for dislocation production and can be less of an issue these days because of probably the most important development in the history of optoelectronic devices, that is, the introduction of the quantum well. In some way a quantum well structure is an evolution of the double heterostructure but with a very much thinner active layer. It is the chosen design for most solid-state light emitting devices today. With the accurate control available from MBE or MOCVD, and following from some early work on superlattices [1.62], very thin layers of carefully controlled composition could be deposited within heterostructure superlattice stacks. It became possible to grow GaAs layers much less than 10 nm thick within AlGaAs–GaAs heterostructures. The carriers in the GaAs were found to exhibit quantum mechanical confinement within the one-dimensional potential well [1.63, 64]. Lasing from $GaAs/Al_{0.2}Ga_{0.8}As$ quantum wells was reported the following year [1.65], but it was a few years before the emission matched that achievable from DH lasers of the time [1.66] and the quantum well laser was further advanced to significantly outperform the competition by researchers in the 1980s [1.67].

The introduction of heterostructures with layer thicknesses on the nanometer or atomic scale represents the final stage in scaling down of these devices. Similarly, Brattain and Bardeen's centimeter-sized transistor has evolved into today's microprocessors with far submicron FETs whose gate oxide thicknesses are measured in Angstroms. Throughout this evolution, materials characterization techniques (Chaps. 17–20) have contributed heavily to the progress in our understanding of electronic materials and these techniques deserve a brief detour here. As the dimensions have been reduced over the decades, the cross-sectional images of device structures published in the literature have progressed from optical microscopy [1.28], to scanning electron microscopy (SEM) images [1.66], to today's high- or atomic-resolution transmission electron microscopy (TEM) analysis of ultrathin layers. The challenge of obtaining relevant and representative data from the smaller volume of material, which is available in higher magnification analysis, has been met during the past decade or so by widespread adoption of highly site specific specimen preparation techniques based around focused ion beam (FIB) milling. For each new material family, understanding of defects and measurement of their densities (e.g., by TEM and x-ray topography) have contributed to improvements in quality. Huge improvements in x-ray optics have seen high-resolution x-ray diffraction techniques develop to become a cornerstone of heterostructure research and production quality control [1.68]. Scanning-probe techniques such as scanning tunneling and atomic force microscopy have become crucial to the understanding of MBE and MOCVD growth. Chemically sensitive techniques such as secondary ion mass spectroscopy and Rutherford backscattering have improved to provide information on doping concentrations and compositions in layered structures with excellent depth resolution. Chemically sensitive measurements have truly reached the atomic scale in the past decade through the exploitation of new generations of aberration-corrected and monochromated transmission electron microscopes and three-dimensional atom probe microscopy (3DAP). The disruptive improvement of spatial resolution enabled by the new generation TEMs continues to yield invaluable insight into atomic-scale material, interface, and device structures, especially when combined with chemically sensitive techniques such as electron energy loss spectroscopy (EELS). Meanwhile, the application of 3DAP to semiconducting and insulating materials and devices provides, for the first time, the *holy grail* of 3-D, atom-by-atom chemically resolved maps. By way of example, we show an atomic-scale chemical map of the critical layers in a III-nitride laser diode structure in Fig. 1.4, revealing crucial information on the distribution of alloy and dopant atoms. The materials characterization process remains a critical element of electronic and photonic materials research.

A primary commercial driver for semiconductor laser diode research has been to produce more effective emitters of infrared wavelengths for transmission of data along optic fibers. Devices based on InP sub-

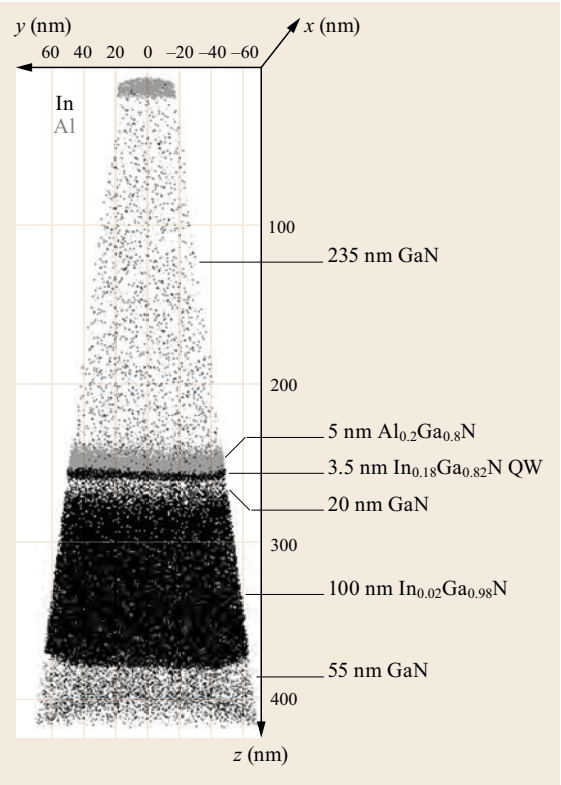


Fig. 1.4 Reconstructed atom map of the active region of a III-nitride laser diode structure determined by 3-D atom probe microscopy (3DAP). Indium atoms are displayed as *black* (25% displayed); aluminum atoms are displayed as *gray* (25% displayed); other atoms are not displayed. (After [1.69], with permission from AIP Publishing)

strates have proven to be extremely effective because of its fortuitous lattice parameter match with other III–V alloys which have band gaps corresponding to the low-absorption *windows* in optic fibers. While remaining lattice matched to InP, the $In_xGa_{1-x}As_yP_{1-y}$ quaternary can exhibit band gaps corresponding to infrared wavelengths of 1.3 and 1.55 μm at which conventional optic fibers absorb the least of the radiation (the absolute minimum is for 1.55 μm). Room temperature continuous lasing of 1.1 μm radiation was demonstrated from the material in 1976 [1.70], and InP-based lasers and photodiodes have played a key role in the optical communications industry since the 1980s [1.39].

A second major driver for laser diode research since the 1970s was to produce laser components for reading or writing optical storage media. We have already mentioned the AlGaAs infrared ($\lambda = 780$ nm) emitters used to read compact discs (CD) and the AlGaInP red ($\lambda = 650$ nm) lasers in DVD devices (Fig. 1.3). As lasers emitting the shorter wavelength became available, the optical disc's surface pits (through which bits of data are stored) could be made smaller and the storage density increased. The pursuit of shorter wavelength laser diodes for higher density optical storage supported extensive development of wide band gap II–VI compounds, principally ZnSe, for their potential in green ($\lambda \approx 520$ nm) and blue ($\lambda \approx 440$ nm) wavelengths. Laser operation in this part of the visible spectrum proved difficult to realize [1.71] and, although a blue/green laser was demonstrated in the early 1990s following improvements in the p-doping of ZnSe [1.72], the devices were prone to rapid deterioration during operation and were never commercialized. Ultimately, the winning laser diode technology for the third generation of optical storage laser was based on III-nitride materials following breakthroughs in the 1990s that are introduced in the next section. Blue/violet ($\lambda = 405$ nm) emitting $In_xGa_{1-x}N$-based laser diodes enabled Blu-ray (BD), the third and current generation of optical storage. CD, DVD, and BD technologies were introduced in 1982, 1995, and 2006, respectively, providing data storage per layer of approximately 0.7 GB, 4.7 GB, and 25 GB; a growth of data capacity five times per decade.

It is likely that BD also marks the final phase of the consumer optical disk era as network-based data distribution and storage replace the widespread use of optical discs. This is a good time, therefore, to note that for over 30 years the rapid advances in home entertainment (music, films, and console games) and computing (software distribution and data storage) were enabled in part by laser diode materials and device technology. The technology race for optical storage laser diodes leaves a legacy of extraordinary laser component technologies that inevitably evolve to bring value to new sectors, for example, the imminent growth of laser diode-based image projectors and automotive headlights. Of course, the growth in network data distribution which has caused the consumer optical disk era to peak is enabled by the laser diodes used in network data communication, a wonderful example of the ebb and flow of technology dominance.

1.3.3 The III-Nitrides

In the final section of this chapter, we focus on the III-nitride materials, such as GaN, $In_xGa_{1-x}N$, and $Al_yGa_{1-y}N$ and their devices, which have been subject to some of the most widespread research and development among all electronic and photonic materials since the mid-1990s. III-nitride LEDs and laser diodes have revolutionized the lighting, display, and optical storage industries during the past decade. Electronic devices based on these materials are expected to have a strong

impact on electrical power distribution and electric vehicles in the next.

The relevance of the $In_xGa_{1-x}N$ alloy for visible light emitting devices is clear from Fig. 1.3. The InN and GaN direct band gaps correspond to wavelengths straddling the visible spectrum, and the alloy potentially offers access to all points in-between. The early commercially successful InGaN light emitters were marketed by Nichia Chemical Industries following the research work of *Shuji Nakamura* who demonstrated the first blue InGaN DH LEDs [1.73], blue InGaN quantum well LEDs, and laser diodes soon after [1.74]. Nakamura, together with other pioneers in III-nitride growth, Isamu Akasaki, and Hiroshi Amano received the 2014 Nobel Prize for Physics for their contributions to blue LED technology. Meanwhile, the wide band gap and other advantageous properties of GaN and $Al_yGa_{1-y}N$ alloys provide new opportunities for high-frequency and power transistors for high voltage switching and deep ultraviolet light emitters.

The development of III-nitride materials and devices has much in common with the early research of other III–V systems. For example, MOCVD and MBE growth technologies were both successfully used for III-nitride device growth (MOCVD has become the preferred technology for mass manufacturing of light emitting devices); both electronic and optoelectronic device applications were explored from an early stage; and one of the obstacles limiting early device development was achieving sufficiently high p-type carrier densities. However, in some ways III-nitrides presented new challenges: Although *conventional* III–Vs all share the same cubic crystallographic structure, the nitrides most readily form in a hexagonal allotrope that has strong polar and piezoelectric properties; the vast majority of III-nitride device development and manufacturing has used heteroepitaxial growth on highly lattice-mismatched substrates leading to devices with very high defect densities, in contrast to a reliance on homoepitaxy or closely lattice matched substrates in previously successful III–V systems where much lower defect densities are standard; and many III-nitride devices have been found to be astonishingly more tolerant to the presence of these crystalline defects than other III–Vs. We will focus briefly on two fascinating aspects of the story of the III-nitrides today: the variety of substrate materials in mainstream use, and the remarkable – and still not fully understood – properties of the $In_xGa_{1-x}N$ alloy.

First, the plurality of substrate materials in use for LEDs. Synthesis of large bulk GaN crystals – the obvious substrate choice from a technical perspective – is much more difficult than for other III–Vs such as GaAs or InP. Methods including ammonothermal growth and hydride VPE are technically viable routes to produce high quality GaN substrates today. However, in the early stages of III-nitride development, there were no large area GaN substrates available and heteroepitaxial growth of GaN on foreign substrates was the only option for commercial LED production. Even today, GaN substrate costs are very high owing to combinations of high temperature, high pressures, slow growth rates, and expensive precursor materials in the manufacturing process.

A consequence of the initial necessity of heteroepitaxy, and ongoing commercial barriers to the use of GaN substrates, is that a total of four substrate materials are in use for commercial production of InGaN-based blue LEDs: sapphire (α-Al_2O_3), silicon carbide, silicon, and GaN. Substrate choice is not purely based on technical factors – commercial considerations are strong drivers – but in the context of conventional compound semiconductors, it is extraordinary that four completely different substrates can be used to produce LEDs with approximately competitive device performance (say, competitive within a factor of two).

Current LED production is dominated by use of sapphire substrates, with silicon carbide substrates a distant second and silicon and GaN substrates just entering the market in recent years. Sapphire is by no means an ideal choice. First, it is electrically insulating which means that electrical contacts cannot be made to the device through the substrate material, thus, requiring LED designs with both n- and p-contacts on the same surface. Second, the refractive indices of sapphire and GaN are significantly different, which lead to low efficiency of extraction of light from the LED chip unless special measures are applied, such as the removal of the substrate or 3-D patterning of the sapphire surface prior to GaN growth. Third, and perhaps most importantly, sapphire has a large lattice mismatch of approximately 16% with GaN [1.75]. This large lattice mismatch is relieved by the formation of misfit dislocations at the GaN/Al_2O_3 interface which lead to threading dislocations propagating through the GaN into the active layers of the devices. The key discovery for reducing the defect densities to tolerable levels was the use of *buffer* layers at the interface with the sapphire [1.75]. Dislocation densities of approximately $10^9\,cm^{-2}$ were adequate for the first commercial LEDs and countless refinements to buffer layer growth have led to typical threading dislocation densities of approximately 10^7–$10^8\,cm^{-2}$ in devices today. In principle, silicon carbide has advantages over sapphire: it can be electrically conducting, the substrate can be readily etched to form chip shapes that improve light extraction efficiency, and it has a smaller lattice mismatch with GaN. However, despite these significant differences,

devices grown on sapphire and silicon carbide have remained broadly competitive with one another. Cost and availability considerations led to sapphire becoming the mainstream substrate choice for LEDs during the early 2000s.

More recently, silicon and GaN have emerged as alternative substrate choices, for quite distinct reasons. Silicon substrates are pursued as a route to reduce LED manufacturing costs compared with incumbent production on sapphire. A key driver is the lower cost of silicon wafers, especially in large wafer formats (at least 150 mm diameter) which can be processed by otherwise unused and fully depreciated wafer processing lines previously established for silicon electronics processing. In addition to the lower substrate cost, GaN-on-silicon enables cost reductions to be made in processing, packaging, and phosphors. For example, GaN-on-silicon facilitates chip scale packaging (CSP), chip scale optics (CSO), and die level integration. CSP has been used for silicon electronics for many years and this technology can now be used for GaN-on-silicon. CSP also reduces the phosphor usage, gives improved thermal performance and is a route to wafer scale packaging (WSP). The main technical challenges to delivering viable device performance and yield are large lattice mismatch (again) and large difference in thermal expansion coefficient compared with GaN, which leads to cracking and bowing during cool down after growth unless careful strain engineering is used during growth [1.76]. In contrast, GaN substrates are substantially more expensive than sapphire, and only viable in smaller formats (typically 50–75 mm diameter). However, LEDs grown on GaN substrates benefit from much lower threading dislocation densities and consequently deliver some of the best device efficiencies, especially for small chips operated at high drive currents [1.77]. The next few years will reveal if silicon substrates or GaN substrates can substantially displace the incumbent sapphire and silicon carbide-based production.

Throughout the history of InGaN LED research, there has been ongoing fascination and controversy regarding the precise mechanism of light emission from InGaN quantum wells. The original driver behind this was the extraordinarily high light emission efficiency of InGaN quantum wells, despite the presence of dislocation densities as high as $10^9\,\mathrm{cm}^{-2}$. Obtaining similar light emission efficiencies in other III–V materials typically required dislocation densities to be several orders of magnitude lower.

For a few years, a popular explanation for the high efficiency of light emission from InGaN quantum wells was that strong fluctuations in the indium concentration in the quantum wells created low-energy quantum-dot like sites that localized electrons and holes and reduced the deleterious impact of dislocations on the light emission process. This explanation was based primarily on TEM analysis of InGaN quantum wells which indicated nanometer-scale regions with greatly higher indium fraction than the remainder of the quantum well – so called *indium clusters*. Many attempts were made to control and optimize this clustering process to increase light emission efficiency. However, since we demonstrated that the apparent indium concentration fluctuations could be a misleading artifact of TEM analysis [1.78, 79], and that high light emission efficiency could be obtained from InGaN quantum wells with random indium distribution [1.80], this theory is no longer widely supported. The origin of localization is now supposed to be due to much subtler features of the InGaN quantum wells.

Recently, the controversy concerning InGaN light emission has evolved to also explaining the phenomenon of *efficiency droop* in InGaN LEDs. *Droop* refers to the observed reduction in efficiency of state-of-the-art LEDs when they are operated at high current densities. There is great pressure to understand and reduce droop because of the potential cost benefit from using smaller LED chips with higher operating currents (i. e., high current densities). The subject of dozens of high profile publications and review articles over the last 10 years, many theories have been vigorously proposed for lower efficiencies at high current densities: increased Auger recombination of electrons and holes; overflow of electrons from the LED active region; the delocalization of carriers; and a complex role of large-scale crystalline defects.

Although these important technical debates are still largely unresolved, the global impact of InGaN LEDs cannot be disputed. This is a good example of how commercial success of electronic and photonic devices can easily occur before full technical understanding of the materials is established. Twelve years ago, in the introduction to the 1st Edition of this handbook, we wrote of the potential for III-nitride LEDs to replace traditional light sources such as incandescent light bulbs, but noted the need to reduce costs and increase the output power from individual LEDs to reach this goal. These commercial and technical barriers were emphatically overcome and the potential is now being realized. Production costs have been driven down by massive economies of scale during epitaxial growth, improvements in yield, cheaper device processing, and improved device performance which mean fewer, smaller LEDs are required for a given application. Tremendous improvements in LED performance – in particular, higher efficiency and high power light output – have been delivered through countless improvements in all aspects of the LED technology, including the core semiconductor material

fabrication. As a consequence, InGaN blue- or violet-emitting LEDs, combined with phosphors that produce white light emission, are now cornerstone technologies in domestic, retail, industrial and outdoor lighting, vehicle headlights, and LCD backlighting. These solid-state technologies provide large energy savings compared with conventional light sources. Every day for the foreseeable future, we will directly sense – through our eyes – the light generated by these devices. Next generation LED lighting may be even more efficient, by omitting phosphors and generating white light by mixing red, yellow, green, and blue LEDs. Such lighting would be color tunable and could mimic sunlight indoors. There is increasing evidence that such optimized lighting would be good for our health, increase productivity at work, and even improve examination results in schools.

In this introductory chapter, we have glimpsed just one sector of the electronic and photonic materials landscape, focusing on just a few semiconductor families. We have given examples of the complex interdependence and balance of power between different materials and technologies: the astonishing integration of materials which we take for granted in a budget smartphone; the switch from germanium to silicon in the early years of the transistor; the dominance and relentless performance improvements in silicon ICs over several decades; the laser-diode enabled growth of consumer optical storage to its peak; and the co-existence of dramatically different approaches to make LEDs which are revolutionizing the way we light the world. There is a complex fusion of science and engineering excellence, ingenious breakthroughs, serendipity, and commercial considerations. But underpinning it all is the need to fundamentally understand and precisely control these fascinating materials. And this is all the more exciting because the pace of progress is so high, and the influence of the materials is so great, thanks to the electronic and photonic materials being the enabling technology in so many aspects of modern life. The detailed chapters in the rest of this book are a valuable resource to support further advances in this vital field.

References

1.1 A.H. Wilson: Proc. R. Soc. Lond. A **133**, 458 (1931)
1.2 A.H. Wilson: Proc. R. Soc. Lond. A **134**, 277 (1931)
1.3 J. Bardeen, W.H. Brattain: Phys. Rev. **74**, 230 (1948)
1.4 J. Bardeen: *Nobel Lecture, Physics, 1942–1962* (Elsevier, Amsterdam 1956), http://www.nobelprize.org/nobel_prizes/physics/laureates/1956/shockley-lecture.html
1.5 E. Braun, S. MacDonald: *Revolution in Miniature*, 2nd edn. (Cambridge Univ. Press, Cambridge 1982)
1.6 W.H. Brattain: *Nobel Lecture, Physics, 1942–1962* (Elsevier, Amsterdam 1956), http://www.nobelprize.org/nobel_prizes/physics/laureates/1956/brattain-lecture.html
1.7 C.S. Fuller: Phys. Rev. (Ser. 2) **86**, 136 (1952)
1.8 I. Derick, C.J. Frosh: US Patent, 2802760 (1955)
1.9 J. Andrus, W.L. Bond: US Patent, 3122817 (1957)
1.10 W.G. Pfann: Trans. Am. Inst. Mech. Eng. **194**, 747 (1952)
1.11 W.G. Pfann: *Zone Melting* (Wiley, New York 1958)
1.12 G.K. Teal: IEEE Trans. Electron Devices **23**, 621 (1976)
1.13 G.K. Teal, J.B. Little: Phys. Rev. (Ser. 2) **78**, 647 (1950)
1.14 G.K. Teal, E. Buehler: Phys. Rev. **87**, 190 (1952)
1.15 J. Czochralski: Z. Phys. Chem. **92**, 219 (1917)
1.16 P.H. Keck, M.J.E. Golay: Phys. Rev. **89**, 1297 (1953)
1.17 H.C. Theurer: US Patent 3 060 123 (1952)
1.18 H.C. Theurer: Trans. Am. Inst. Mech. Eng. **206**, 1316 (1956)
1.19 K.A. Jackson (Ed.): *Silicon Devices* (Wiley-VCH, Weinheim 1998)
1.20 W.C. Dash: J. Appl. Phys. **29**, 736 (1958)
1.21 W.C. Dash: J. Appl. Phys. **30**, 459 (1959)
1.22 W.C. Dash: J. Appl. Phys. **31**, 736 (1960)
1.23 G. Zeigler: Z. Naturforsch. **16a**, 219 (1961)
1.24 M. Grayson (Ed.): *Encyclopedia of Semiconductor Technology* (Wiley, New York 1984) p. 734
1.25 I.M. Ross: Bell Labs Tech. J. **2**(4), 3 (1997)
1.26 W. Shockley: Bell Syst. Tech. J. **28**(4), 435 (1949)
1.27 W. Shockley, M. Sparks, G.K. Teal: Phys. Rev. **83**, 151 (1951)
1.28 M. Tanenbaum, D.E. Thomas: Bell Syst. Tech. J. **35**, 1 (1956)
1.29 C.M. Melliar-Smith, D.E. Haggan, W.W. Troutman: Bell Labs Tech. J. **2**(4), 15 (1997)
1.30 J.A. Oerni: IRE Trans. Electron Devices **7**, 178 (1960)
1.31 J.S. Kilby: IEEE Trans. Electron Devices **23**, 648 (1976)
1.32 J.S. Kilby: *Nobel Lectures in Physics: 1996–2000* (Imperial College, London 2000), http://www.nobelprize.org/nobel_prizes/physics/laureates/2000/kilby-lecture.html
1.33 D. Kahng, M.M. Atalla: Silicon-silicon dioxide field induced surface devices, Proc. Solid State Res. Conf., Pittsburgh (1960)
1.34 F.H. Baumann, C.P. Chang, J.L. Grazul, A. Kamgar, C.T. Liu, D.A. Muller: C4.1, Mat. Res. Soc. Symp. **611**, 1 (2000)
1.35 J.T. Clemens: Bell Labs Tech. J. **2**(4), 76 (1997)
1.36 G.E. Moore: Electronics **38**(8) (1965)
1.37 G.E. Moore: IEEE Int. **1**, 20 (2003)
1.38 H. Kroemer: *Nobel Lectures in Physics: 1996–2000* (Imperial College, London 2000), http://www.nobelprize.org/nobel_prizes/physics/laureates/2000/kroemer-lecture.html
1.39 O. Wada, H. Hasegawa (Eds.): *InP-Based Materials and Devices* (Wiley, New York 1999)
1.40 C.Y. Chang, F. Kai: *GaAs High-Speed Devices* (Wiley, New York 1994)

1.41 A.Y. Cho: J. Vac. Sci. Technol. **8**, S31 (1971)
1.42 H.M. Manasevit: Appl. Phys. Lett. **12**, 156 (1968)
1.43 H.J. Welker: IEEE Trans. Electron Devices **23**, 664 (1976)
1.44 H. Kroemer: RCA Review **18**, 332 (1957)
1.45 C.A. Mead: Proc. IEEE **54**, 307 (1966)
1.46 W.W. Hooper, W.I. Lehrer: Proc. IEEE **55**, 1237 (1967)
1.47 R. van Tuyl, C. Liechti: IEEE Spectrum **14**(3), 41 (1977)
1.48 R. Dingle, H.L. Störmer, A.C. Gossard, W. Wiegmann: Appl. Phys. Lett. **33**, 665 (1978)
1.49 T. Mimura, S. Hiyamizu, T. Fujii, K. Nanbu: Jpn. J. Appl. Phys. **19**, L225 (1980)
1.50 T. Mimura: IEEE Trans. Microw. Theory Tech. **50**, 780 (2002)
1.51 D. Delagebeaudeuf, P. Delescluse, P. Etienne, M. Laviron, M. Chaplart, N.T. Linh: Electron. Lett. **16**, 667 (1980)
1.52 D. Chattopadhyay, S.K. Sutradhar, B.R. Nag: J. Phys. C **14**, 891 (1981)
1.53 E.E. Loebner: IEEE Trans. Electron Devices **23**, 675 (1976)
1.54 R.N. Hall, G.E. Fenner, J.D. Kingsley, T.J. Soltys, R.O. Carlson: Phys. Rev. Lett. **9**, 366 (1962)
1.55 M.I. Nathan, W.P. Dumke, G. Burns, F.H. Dill Jr., G. Lasher: Appl. Phys. Lett. **1**, 62 (1962)
1.56 N. Holonyak, M.B. Panish, P.W. Foy, S. Sumski: Appl. Phys. Lett. **1**, 82 (1962)
1.57 H. Kroemer: Proc. IEEE **51**, 1782 (1963)
1.58 Z.I. Alferov: *Nobel Lectures in Physics: 1996–2000* (Imperial College Press, London 2000), http://www.nobel.se/physics/laureates/2000/alferov-lecture.html
1.59 Z.I. Alferov, V.M. Andreev, D.Z. Garbuzov, Y.V. Zhilyaev, E.P. Morozov, E.L. Portnoi, V.G. Trofim: Fiz. Tekh. Poluprovodn. **4**, 1826 (1970)
1.60 I. Hayashi, G.E. Fenner, J.D. Kingsley, T.J. Soltys, R.O. Carlson: Appl. Phys. Lett. **17**, 109 (1970)
1.61 A. Žukauskas: *Introduction to Solid-State Lighting* (Wiley, New York 2002)
1.62 L. Esaki, R. Tsu: IBM J. Res. Dev. **14**, 61 (1970)
1.63 L.L. Chang, L. Esaki, R. Tsu: Appl. Phys. Lett. **24**, 593 (1974)
1.64 R. Dingle, W. Wiegmann, C.H. Henry: Phys. Rev. Lett. **33**, 827 (1974)
1.65 J.P. van der Ziel, R. Dingle, R.C. Miller, W. Wiegmann, W.A. Nordland: Appl. Phys. Lett. **26**, 463 (1975)
1.66 R. Dupuis, P.D. Dapkus, U. Holonyak Jr., E.A. Reack, R. Chin: Appl. Phys. Lett. **32**, 295 (1978)
1.67 Z.I. Alferov: Semiconductors **32**, 1 (1998)
1.68 D.K. Bowen, B.K. Tanner: *High Resolution X-ray Diffractometry and Topography* (Taylor Francis, London 1998)
1.69 S.E. Bennett, T.M. Smeeton, D.W. Saxey, G.D.W. Smith, S.E. Hooper, J. Heffernan, C.J. Humphreys, R.A. Oliver: J. Appl. Phys. **111**, 053508 (2012)
1.70 J.J. Hsieh, J.A. Rossi, J.P. Donneley: Appl. Phys. Lett. **28**, 709 (1976)
1.71 H.E. Ruda (Ed.): *Widegap II–VI Compounds for Opto-Electronic Applications* (Chapman Hall, London 1992)
1.72 M.A. Haase, J. Qiu, J.M. DePuydt, H. Cheng: Appl. Phys. Lett. **59**, 1272 (1991)
1.73 S. Nakamura, T. Mukai, M. Senoh: Appl. Phys. Lett. **64**, 1687 (1994)
1.74 S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku, Y. Sugimoto: Jpn. J. Appl. Phys. **35**, L74 (1996)
1.75 S. Nakamura, G. Fasol: *The Blue Laser Diode* (Springer, Berlin 1997)
1.76 D. Zhu, D.J. Wallis, C.J. Humphreys: Rep. Prog. Phys. **76**, 106501 (2013)
1.77 C.A. Hurni: Appl. Phys. Lett. **106**, 031101 (2015)
1.78 T.M. Smeeton, M.J. Kappers, J.S. Barnard, M.E. Vickers, C.J. Humphreys: Appl. Phys. Lett. **83**, 5419 (2003)
1.79 T.M. Smeeton, C.J. Humphreys, J.S. Barnard, M.J. Kappers: J. Mater. Sci. **41**, 2729 (2006)
1.80 M.J. Galtrey: J. Appl. Phys. **104**, 013524 (2008)

Part A Fundamental Properties

Part A Fundamental Properties

2 **Electrical Conduction in Metals and Semiconductors**
Safa Kasap, Saskatoon, Canada
Cyril Koughia, Saskatoon, Canada
Harry E. Ruda, Toronto, Canada

3 **Optical Properties of Electronic Materials: Fundamentals and Characterization**
Jan Mistrik, Pardubice, Czech Republic
Safa Kasap, Saskatoon, Canada
Harry E. Ruda, Toronto, Canada
Cyril Koughia, Saskatoon, Canada
Jai Singh, Darwin, Australia

4 **Magnetic Properties: From Traditional to Spintronic**
Charbel Tannous, Brest, France
Jacek Gieraltowski, Plouzané, France

5 **Defects in Monocrystalline Silicon**
Wilfried von Ammon, Roth, Germany
Andreas Sattler, München, Germany
Gudrun Kissinger, Frankfurt/Oder, Germany

6 **Diffusion in Semiconductors**
Derek Shaw, Hull, UK

7 **Photoconductivity in Materials Research**
Stephen Reynolds, Dundee, UK
Monica Brinza, Veldhoven, The Netherlands
Mohammed L. Benkhedir, Tebessa, Algeria
Guy J. Adriaenssens, Leuven, Belgium

8 **Electronic Properties of Semiconductor Interfaces**
Winfried Mönch, Duisburg, Germany

9 **Charge Transport in Disordered Materials**
Sergei Baranovskii, Marburg, Germany
Oleg Rubel, Hamilton, Canada

10 **Dielectric Response**
Leonard Dissado, Leicester, UK

11 **Ionic Conduction and Applications**
Harry Tuller, Cambridge, USA

2. Electrical Conduction in Metals and Semiconductors

Safa Kasap, Cyril Koughia, Harry E. Ruda

Electrical transport through materials is a large and complex field, and in this chapter we cover only a few aspects that are relevant to practical applications. We start with a review of the semi-classical approach that leads to the concepts of drift velocity, mobility and conductivity, from which Matthiessen's Rule is derived. A more general approach based on the Boltzmann transport equation is also discussed. We review the conductivity of metals and include a useful collection of experimental data. The conductivity of nonuniform materials such as alloys, polycrystalline materials, composites and thin films is discussed in the context of Nordheim's rule for alloys, effective medium theories for inhomogeneous materials, and theories of scattering for thin films. We also discuss some interesting aspects of conduction in the presence of a magnetic field (the Hall effect). We present a simplified analysis of charge transport in semiconductors in a high electric field, including a modern avalanche theory (the theory of *lucky* drift). The properties of low-dimensional systems are briefly reviewed, including the quantum Hall effect.

2.1 Fundamentals: Drift Velocity, Mobility and Conductivity 20
2.2 Matthiessen's Rule 22
2.3 Resistivity of Metals 23
2.3.1 General Characteristics 23
2.3.2 Fermi Electrons 25
2.4 Solid Solutions and Nordheim's Rule 26
2.5 Carrier Scattering in Semiconductors 28
2.6 The Boltzmann Transport Equation 29
2.7 Resistivity of Thin Polycrystalline Films 30
2.8 Inhomogeneous Media: Effective Media Approximation 32
2.9 The Hall Effect 35
2.10 High Electric Field Transport 37
2.11 Impact Ionization 38
2.12 Two-Dimensional Electron Gas 40
2.13 One-Dimensional Conductance 42
2.14 The Quantum Hall Effect 43
References 44

A good understanding of charge carrier transport and electrical conduction is essential for selecting or developing electronic materials for device applications. Of particular importance are the drift mobility of charge carriers in semiconductors and the conductivity of conductors and insulators. Carrier transport is a broad field that encompasses both traditional *bulk* processes and, increasingly, transport in low dimensional or quantized structures. In other chapters of this handbook, Baranovskii describes hopping transport in low mobility solids such as insulators, Morigaki deals with the electrical properties of amorphous semiconductors and Gould discusses in detail conduction in thin films. In this chapter, we outline a semi-quantitative theory of charge transport suitable for a wide range of solids of use to materials researchers and engineers. We introduce theories of *bulk* transport followed by processes pertinent to ultra-fast transport and quantized transport in lower dimensional systems. The latter covers such phenomena as the Quantum Hall Effect, and Quantized Conductance and Ballistic Transport in Quantum Wires that has potential use in new kinds of devices. There are many more rigorous treatments of charge transport; those by *Rossiter* [2.1] and *Dugdale* [2.2] on metals, and *Nag* [2.3] and *Blatt* [2.4] on semiconductors are highly recommended.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_2

2.1 Fundamentals: Drift Velocity, Mobility and Conductivity

Basic to the theory of the electronic structure of solids are the solutions to the quantum mechanical problem of an electron in a periodic potential known as Bloch waves. These wavefunctions are traveling waves and provide the physical basis for conduction. In the semi-classical approach to conduction in materials, an electron wavepacket made up of a superposition of Bloch waves can in principle travel unheaded in an ideal crystal. No crystal is ideal, however, and the imperfections cause scattering of the wavepacket. Since the interaction of the electron with the potential of the ions is incorporated in the Bloch waves, one can concentrate on the relatively rare scattering events which greatly simplifies the theory. The motion of the electrons between scattering events is essentially free (with certain provisos such as no interband transitions) subject only to external forces, usually applied electric or magnetic fields. A theory can then be developed that relates macroscopic and measurable quantities such as conductivity or mobility to the microscopic scattering processes. Principle in such a theory is the concept of *mean free time* τ which is the average time between scattering events. τ is also known as the *conductivity relaxation time* because it represents the time scale for the momentum gained from an external field to relax. Equivalently, $1/\tau$ is the average probability per unit time that an electron is scattered.

There are two important velocity quantities that must be distinguished. The first is the *mean speed* u or *thermal velocity* v_{th} which, as the name implies, is the average speed of the electrons. u is quite large being on the order of $\sqrt{3k_{\mathrm{B}}T/m_{\mathrm{e}}^*} \approx 10^5\,\mathrm{m/s}$ for electrons in a nondegenerate semiconductor and $\sqrt{2E_{\mathrm{F}}/m_{\mathrm{e}}^*} \approx 10^6\,\mathrm{m/s}$ for electrons in a metal, where k_{B} is Boltzmann's constant, T is the temperature, E_{F} is the Fermi energy of the metal, and m_{e}^* is the electron effective mass. The distance an electron travels between scattering events is called the free path. It is straightforward to show that the average or *mean free path* for an electron is simply $\ell = u\tau$. The second velocity is the mean or *drift velocity* $\boldsymbol{v}_{\mathrm{d}}$ (variables in boldface are vectors) which is simply the vector average over the velocities of all N electrons,

$$\boldsymbol{v}_{\mathrm{d}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{v}_i \,. \tag{2.1}$$

With no external forces applied to the solid, the electron motion is random and thus the drift velocity is zero. When subject to external forces like an electric field, the electrons acquire a net drift velocity. Normally, the magnitude of the drift velocity is much smaller than u so that the mean speed of the electron is not affected to any practical extent by the external forces. An exception is charge transport in semiconductors in high electric fields, where $|\boldsymbol{v}_{\mathrm{d}}|$ becomes comparable to u.

The drift velocity gives rise to an electric current. If the density of electrons is n then the current density $\boldsymbol{J}_{\mathrm{e}}$ is

$$\boldsymbol{J}_{\mathrm{e}} = -en\boldsymbol{v}_{\mathrm{d}} \tag{2.2}$$

where e is the fundamental unit of electric charge. For the important case of an applied electric field $\boldsymbol{E}$, the solutions of the semi-classical equations give a drift velocity that is proportional to $\boldsymbol{E}$. The proportionality constant is the *drift mobility* μ_{e}

$$\boldsymbol{v}_{\mathrm{d}} = -\mu_{\mathrm{e}}\boldsymbol{E} \,. \tag{2.3}$$

The drift mobility might be a constant or it might depend on the applied field (usually only if the field is large). Ohm's Law defines the conductivity σ of a material $\boldsymbol{J} = \sigma\boldsymbol{E}$ resulting in a simple relation to the drift mobility

$$\sigma = en\mu_{\mathrm{e}} \,. \tag{2.4}$$

Any further progress requires some physical theory of scattering. A useful model results from the simple assumption that the scattering randomizes the electron's velocity (taking into proper account the distribution of electrons and the Pauli Exclusion Principle). The equation of motion for the drift velocity then reduces to a simple form

$$\frac{\mathrm{d}\boldsymbol{v}_{\mathrm{d}}}{\mathrm{d}t} = \frac{\boldsymbol{F}(t)}{m_{\mathrm{e}}^*} - \frac{\boldsymbol{v}_{\mathrm{d}}}{\tau} \,, \tag{2.5}$$

where $\boldsymbol{F}(t)$ is the sum of all external forces acting on the electrons. The effect of the scattering is to introduce a frictional term into what otherwise would be just Newton's Law. Solutions of (2.5) depend on $\boldsymbol{F}(t)$. In the simplest case of a constant applied electric field, the steady-state solution is trivial,

$$\boldsymbol{v}_{\mathrm{d}} = \frac{-e\boldsymbol{E}\tau}{m_{\mathrm{e}}^*} \,. \tag{2.6}$$

The conductivity and drift mobility can now be related to the scattering time [2.5],

$$\mu_{\mathrm{e}} = \frac{e\tau}{m_{\mathrm{e}}^*} \quad \text{and} \quad \sigma = \frac{ne^2\tau}{m_{\mathrm{e}}^*} \,. \tag{2.7}$$

More sophisticated scattering models lead to more accurate but more complicated solutions.

The simple expression (2.7) can be used to explain qualitative features of conduction in materials once a physical origin for the scattering is supplied. For any scattering site, the effective area is the *scattering cross-section* S as depicted in Fig. 2.1. The scattering cross section is related to the mean free path since the volume $S\ell$ must contain one scattering center. If there are N_S scattering centers per unit volume then

$$\ell = \frac{1}{SN_S} , \tag{2.8}$$

or, rewriting in terms of τ,

$$\tau = \frac{1}{SN_S u} . \tag{2.9}$$

Once the cross-section for each physically relevant scattering mechanism is known then the effect on the scattering time and conductivity is readily calculated.

An overly restrictive assumption in the above analysis is that the electron's velocity is completely randomized every time it is scattered. On the other hand, suppose that ν collisions are required to completely destroy the directional velocity information. That is only after an average of ν collisions do all traces of correlation between the initial and the final velocities disappear. The *effective* mean free path ℓ_{eff} traversed by the electron until its velocity is randomized will now be larger than ℓ; to first order $\ell_{\text{eff}} = \nu\ell$. ℓ_{eff} is termed the *effective* or the *conduction mean free path*. The corresponding *effective scattering cross* section is

$$S_{\text{eff}} = \frac{1}{N_S \ell_{\text{eff}}} . \tag{2.10}$$

The expressions for mobility and conductivity become

$$\mu = \frac{e\nu\tau}{m_e^*} = \frac{e\nu\ell}{m_e^* u} = \frac{e\ell_{\text{eff}}}{m_e^* u} = \frac{e}{m_e^* u N_S S_{\text{eff}}} \tag{2.11}$$

and

$$\sigma = \frac{e^2\nu\tau}{m_e^*} = \frac{e^2 n\nu\ell}{m_e^* u} = \frac{e^2 n\ell_{\text{eff}}}{m_e^* u} = \frac{e^2 n}{m_e^* u N_S S_{\text{eff}}} . \tag{2.12}$$

Suppose that in a collision the electron is scattered at an angle θ to its original direction of travel as shown in Fig. 2.2. It is convenient to introduce a quantity $S_\theta(\theta)$, called the *differential scattering cross section*, defined so that $2\pi \sin\theta S_\theta \mathrm{d}\theta$ represents the probability of scattering at an angle between θ and $\theta + \mathrm{d}\theta$ with respect to the original direction. If the magnitude of the velocity is not changed then the fractional change in component of the velocity along the original direction is $1 - \cos\theta$. The average number of collisions ν required to randomize the velocity is then

$$\nu = \frac{1}{\langle 1 - \cos\theta \rangle} , \tag{2.13}$$

where the average is given by

$$\langle 1 - \cos\theta \rangle = \frac{\int_0^\pi (1 - \cos\theta)\, S_\theta\,(\theta) \sin\theta \mathrm{d}\theta}{\int_0^\pi S_\theta\,(\theta) \sin\theta \mathrm{d}\theta} . \tag{2.14}$$

The effective cross sectional area S_{eff} is then

$$S_{\text{eff}} = 2\pi \int_{\theta_m}^{\pi} (1 - \cos\theta)\, S_\theta\,(\theta) \sin\theta \mathrm{d}\theta . \tag{2.15}$$

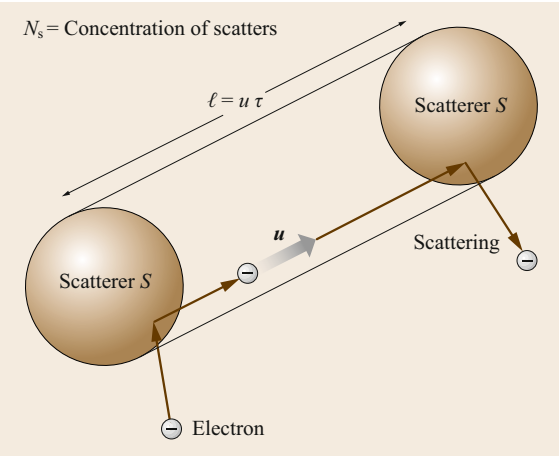


Fig. 2.1 Scattering of an electron from a scattering center. The electron travels a mean distance $\ell = u\tau$ between collisions

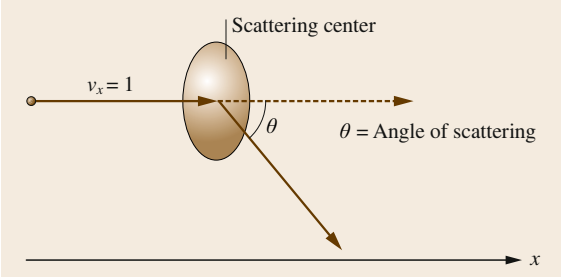


Fig. 2.2 An electron moving in the x-direction becomes scattered though an angle θ with respect to the original direction

As an example, consider conduction electrons scattering from charged impurities in a nondegenerate semiconductor where small angle deviations predominate. The differential cross section for coulombic scattering from a charged impurity center with charge $+Ze$ is

$$S_\theta(\theta) = \frac{16k^2}{u^4} \times \frac{1}{\theta^4}\,; \quad k = \frac{Ze^2}{4\pi\varepsilon_0\varepsilon_r m_e^*}\,, \tag{2.16}$$

where $k = Ze^2/4\pi\varepsilon_0\varepsilon_r m_e^*$ and ε_r is the relative permittivity of the semiconductor. Let r_m be the maximum effective radius of action for the impurity at which the minimum scattering angle θ_m occurs. Then the integral in (2.15) evaluates to

$$S_{\text{eff}} = \frac{2\pi k^2}{u^4} \ln\left(1 + \frac{r_m^2 u^4}{k^2}\right). \tag{2.17}$$

In a nondegenerate semiconductor, the equipartition theorem links velocity to temperature, $m_e^* u^2/2 = 3kT/2$, so that $u \propto T^{1/2}$. Thus,

$$S_{\text{eff}} = \frac{A}{T^2} \ln\left(1 + BT^2\right), \tag{2.18}$$

where A and B are constants. The drift mobility due to scattering from ionized impurities becomes

$$\mu_I = \frac{e}{m_e^* u S_{\text{eff}} N_I} \propto \frac{1}{T^{1/2}} \frac{T^2}{A\ln(1+BT^2)} \frac{1}{N_I} \approx \frac{CT^{3/2}}{N_I}\,, \tag{2.19}$$

where N_I is the density of ionized impurities (it represents N_S), and C is a new constant. At low temperatures where lattice scattering is insignificant, we expect $\mu_I \propto T^{3/2}/N_I$ for nondegenerate semiconductors.

The above semiquatitative description is sufficient to understand the basic principles of conduction. A more rigorous approach involves solving the Boltzmann charge transport equation and is addressed in Sect. 2.6.

2.2 Matthiessen's Rule

In general, the conduction electron whether in a metal or in a semiconductor can be scattered by a number of mechanisms, such as lattice vibrations, impurities, lattice defects such as dislocations, grain boundaries, vacancies, surfaces, or any other deviation from a perfectly periodic lattice. All these scattering processes increase the overall resistivity of the substance by reducing the mean scattering time. The relation between the types of scattering and the total scattering time can be obtained by considering scattering from lattice vibrations and impurities as shown in Fig. 2.3. We define two mean free times τ_L and τ_I:τ_L is the mean free time considering only scattering from lattice vibrations (phonons) and τ_I is the mean free time considering only collisions with impurities. In a small unit of time dt, the total probability of scattering (dt/τ) is simply the sum of the probability for phonon scattering (dt/τ_L) and the probability for impurity scattering (dt/τ_I), and thus

$$\frac{1}{\tau} = \frac{1}{\tau_L} + \frac{1}{\tau_I}\,. \tag{2.20}$$

We have assumed that neither τ_L nor τ_I is affected by the presence of the other scattering mechanism, that is each type of scattering is independent. The above expression can be generalized to include all types of independent scattering mechanisms yielding

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i}\,, \tag{2.21}$$

where τ_i is the mean scattering time considering the ith scattering process alone. Since the drift mobility μ_d

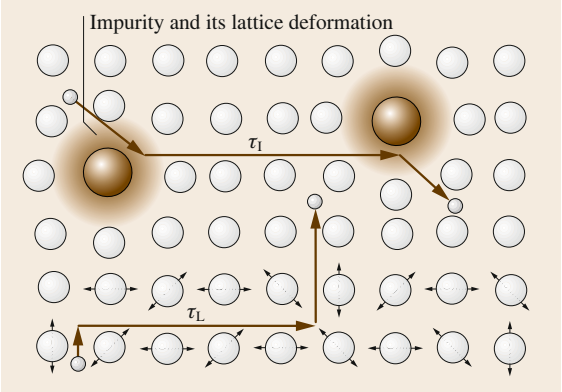


Fig. 2.3 Scattering from lattice vibrations alone with a mean scattering time τ_L, and from impurities alone with a mean scattering time τ_I

is proportional to τ, (2.21) can be written in terms of the drift mobilities determined by the various scattering mechanisms. In other words,

$$\frac{1}{\mu_d} = \sum_i \frac{1}{\mu_i} , \tag{2.22}$$

where μ_i is the drift mobility limited just by the ith scattering process. Finally, since the resistivity is inversely proportional to the drift mobility, the relation for resistivity is

$$\rho = \sum \rho_i , \tag{2.23}$$

where ρ_i is the resistivity of the material if only the ith scattering process were active. Equation (2.23) is known as Matthiessen's rule. For nearly perfect, pure crystals the resistivity is dominated by phonon scattering ρ_T. If impurities or defects are present, however, there are an additional resistivities ρ_I from the scattering off the impurities and ρ_D from defect scattering, and $\rho = \rho_T + \rho_I + \rho_D$.

Matthiessen's rule is indispensable for predicting the resistivities of many types of conductors. In some cases like thin films, the rule is obeyed only approximately, but it is nonetheless still useful for an initial (often quite good) estimate.

2.3 Resistivity of Metals

2.3.1 General Characteristics

The effective resistivity of a metal, by virtue of Matthiessen's rule, is normally written as

$$\rho = \rho_T + \rho_R , \tag{2.24}$$

where ρ_R is called the *residual resistivity* and is due to the scattering of electrons by impurities, dislocations, interstitial atoms, vacancies, grain boundaries and so on. The residual resistivity shows very little temperature dependence whereas ρ_T is nearly linear in absolute temperature. ρ_T will be the main resistivity term for many high-quality, nonmagnetic pure, crystalline metals. In classical terms, we can take the thermal vibrations of a lattice atom with mass M as having a mean kinetic energy KE of $(1/2)Ma^2\omega^2$, where a and ω are the amplitude and frequency of the vibrations. This mean KE must be of the order of kT so that the amplitude $a \propto T^{1/2}$. Thus the electron scattering cross section $S = \pi a^2 \propto T$. Since the mean speed of conduction electron in a metal is the Fermi speed and is temperature insensitive, $\mu \propto \tau \propto 1/S \propto T^{-1}$, and hence the resistivity $\rho \propto T$. Most nonmagnetic pure metals obey this relationship except at very low temperatures. Figure 2.4 shows the resistivity of Cu as a function of temperature where above ≈ 100 K, $\rho \propto T$.

The $\rho \propto T$ behavior can also be derived more rigorously by noting that at high temperatures, the phonon concentration n_{ph} increases as T. The mean scattering time τ is inversely proportional to n_{ph} so that $\rho \propto T$. Frequently, the resistivity versus temperature behavior of pure metals can be empirically represented by a power law of the form,

$$\rho = \rho_0 \left(\frac{T}{T_0}\right)^n , \tag{2.25}$$

where ρ_0 is the resistivity at the reference temperature, T_0, and n is a characteristic index that best fits the data. For the nonmagnetic pure metals, n is close to unity whereas it is close to 2 for the magnetic metals Fe and Ni [2.5]. Figure 2.5 shows ρ versus T for various metals. Table 2.1 summarizes the values ρ_0 and n for various metals.

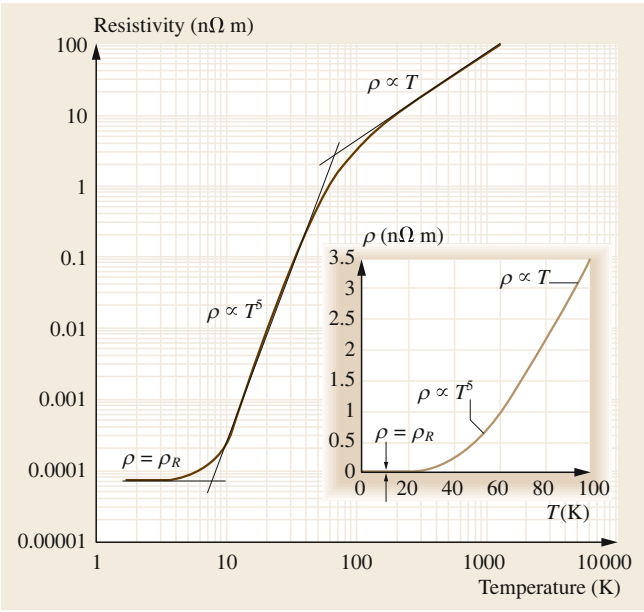


Fig. 2.4 The resistivity of copper from low to high temperatures (near its melting temperature, 1358 K) on a log–log plot. Above about 100 K, $\rho \propto T$, whereas at low temperatures, $\rho \propto T^5$, and at the lowest temperatures ρ approaches the residual resistivity ρ_R. The inset shows the ρ versus T behavior below 100 K on a linear plot. (ρ_R is too small to see on this scale.) (After [2.5])

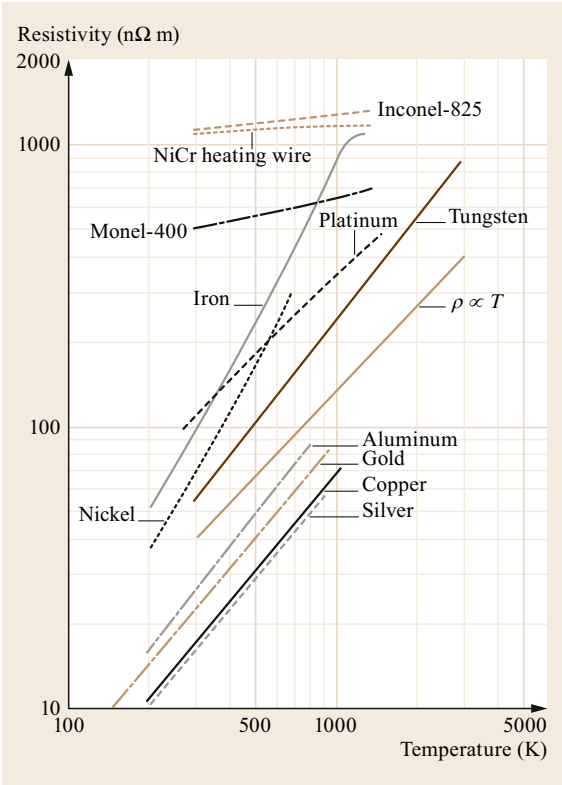


Fig. 2.5 The resistivities of various metals as a function of temperature above 0 °C. Tin melts at 505 K, whereas nickel and iron go through a magnetic-to-nonmagnetic (Curie) transformation at about 627 K and 1043 K, respectively. The theoretical behavior ($\rho \propto T$) is shown for reference. (After [2.5])

As apparent from Fig. 2.4, below ≈ 100 K, the $\rho \propto T$ behavior fails, and $\rho \propto T^5$. The reason is that, as the temperature is lowered, the scattering by phonons becomes less efficient, and it takes many more collisions to fully randomize the initial velocity of the electron. The mean number of collisions ν required to randomize the velocity scales with T^{-2} [2.5], and at low temperatures, the concentration n_{phonon} of phonons increases as T^3. Thus,

$$\sigma \propto \nu\tau \propto \frac{\nu}{n_{\text{phonon}}} \propto \frac{T^{-2}}{T^3} \propto T^{-5} \qquad (2.26)$$

which explains the low-temperature $\rho - T$ behavior in Fig. 2.4. The low temperature $\rho \propto T^5$ and high temperature $\rho \propto T$ regimes are roughly separated by the *Debye temperature* T_{D}. For $T > T_{\text{D}}$ we expect $\rho \propto T$, and for $T < T_{\text{D}}$ we expect $\rho \propto T^5$.

In the case of metals with impurities and for alloys, we need to include the ρ_{R} contribution to the overall resistivity. For $T > T_{\text{D}}$, the overall resistivity is

$$\rho \approx AT + \rho_{\text{R}} \,, \qquad (2.27)$$

where A is a constant, and the AT term in (2.27) arises from scattering from lattice vibrations. Normally, ρ_{R} has very little temperature dependence, and hence very roughly ρ versus T curves shift to higher values as ρ_{R} is increases due to the addition of impurities, alloying or cold working the sample (mechanical deformation that generates dislocations) as illustrated for Cu–Ni alloys in Fig. 2.6.

Table 2.1 Typical resistivities at 273 K (0 °C) ρ_0 and thermal coefficients of resistivity α_0 at 0 °C for various metals above 0 °C and below melting temperature. The resistivity index n in $\rho = \rho_0(T/T_0)^n$ is also shown. Note that n is fitted to resistivity data above 273 K but below the melting temperature of the metal. Data selectively combined from various sources, including [2.6, 7]

Metal	ρ_0 (nΩ m)	n	$\alpha_0 \times 10^{-3}$ (K^{-1})
Aluminium, Al	24.2	1.20	4
Antimony, Sb	302	1.27	4.7
Beryllium, Be	30.2	1.73	8
Bismuth, Bi	1070	–	4.5
Cadmium, Cd	68	–	4.3
Calcium, Ca	31.1	1.09	4.0
Cerium, Ce	730	–	0.9
Cesium, Cs	187	1.23	4.8
Cromium, Cr	118	1.01	2.9
Cobalt, Co	56	–	6.6
Copper, Cu	15.4	1.16	4.3
Gold, Au	20.5	1.13	4.0
Hafnium, Ha	304	1.21	4.4
Indium, In	80	1.31	5.0
Iridium, Ir	47	–	4.5
Iron, Fe	85.7	1.73	6.3
Lead, Pb	192	1.14	4.2
Magnesium, Mg	40.5	1.07	4.2
Molybdenum, Mo	48.5	1.21	5.0
Nickel, Ni	61.6	1.76	6.5
Niobium, Nb	152	–	2.3
Palladium, Pd	97.8	0.94	0.39
Platinum, Pt	98.4	1.01	3.9
Rhodium, Rh	43	–	4.4
Ruthenium, Ru	71	–	4.1
Silver, Ag	14.7	1.13	4.1
Strontium, Sr	123	0.99	3.6
Tantalum, Ta	122	0.93	3.6
Tin, Sn	115	1.1	4.0
Titanium, Ti	390	1.01	4.8
Tungsten, W	48.2	1.24	4.8
Vanadium, V	181	1.02	3.9
Zinc, Zn	54.6	1.14	4.2
Zirconium, Zr	388	1.00	4.2

Resistivity versus temperature behavior of nearly all metals is characterized by the *temperature coefficient of resistivity* (TCR) α_0 which is defined as the fractional change in the resistivity per unit temperature increase at the reference temperature T_0, i. e.,

$$\alpha_0 = \frac{1}{\rho_0}\left(\frac{\mathrm{d}\rho}{\mathrm{d}T}\right)_{T=T_0}, \tag{2.28}$$

where ρ_0 is the resistivity at the reference temperature T_0, usually at 273 K (0 °C) or 293 K (20 °C), $\mathrm{d}\rho = \rho - \rho_0$ is the change in the resistivity due to a small increase, $\mathrm{d}T = T - T_0$, in temperature. Assuming that α_0 is temperature independent over a small range from T_0 to T, we can integrate (2.28), which leads to the well known equation,

$$\rho = \rho_0[1 + \alpha_0(T - T_0)]\,. \tag{2.29}$$

Equation (2.29) is actually only valid when α_0 is *constant* over the temperature range of interest which requires (2.27) to hold. Over a limited temperature range this will usually be the case. Although it is not obvious from (2.28), we should, nonetheless, note that α_0 depends on the reference temperature, T_0 by virtue of ρ_0 depending on T_0.

It is instructive to mention that if $\rho \approx AT$ as we expect for an ideal pure metal, then $\alpha_0 = T_0^{-1}$. If we take the reference temperature T_0 as 273 K (0 °C), then α_0 should ideally be 1/(273 K) or $3.66 \times 10^{-3}\,\mathrm{K}^{-1}$. Examination of a_0 for various metals shows that $\rho \propto T$ is not a bad approximation for some of the familiar pure metals used as conductors, e.g., Cu, Al, Au, but fails badly for others, such as indium, antimony and, in particular, the magnetic metals, e.g., iron and nickel.

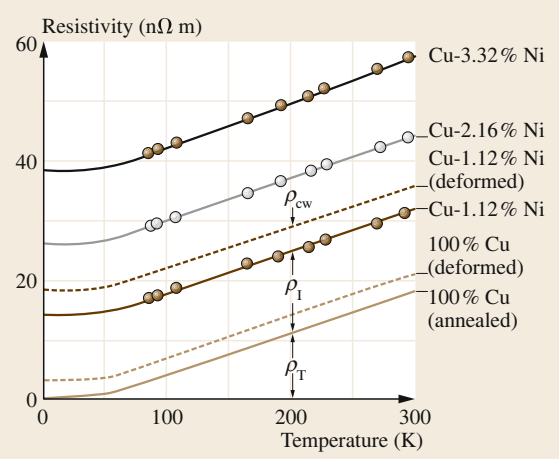


Fig. 2.6 Resistivities of annealed and cold-worked (deformed) copper containing various amounts of Ni (given in atomic percentages) versus temperature

Frequently we are given α_0 at a temperature T_0, and we wish to use some other reference temperature, say T_0', that is, we wish to use ρ_0' and α_0' for ρ_0 and α_0 respectively in (2.29) by changing the reference from T_0 to T_0'. Then we can find α_1 from α_0,

$$\alpha_0' = \frac{\alpha_0}{1 + \alpha_0(T_0' - T_0)}$$
$$\text{and} \quad \rho = \rho_0'\left[1 + \alpha_0'\left(T - T_0'\right)\right]. \tag{2.30}$$

For example, for Cu $\alpha_0 = 4.31 \times 10^{-3}\,\mathrm{K}^{-1}$ at $T_0 = 0\,°\mathrm{C}$, but it is $\alpha_0 = 3.96 \times 10^{-3}\,\mathrm{K}^{-1}$ at $T_0 = 20\,°\mathrm{C}$. Table 2.1 summarizes α_0 for various metals. Note that α_0 in Table 2.1 is at 0 °C and can be converted to α_0' at 20 °C using (2.30).

2.3.2 Fermi Electrons

The electrical properties of metals depend on the behavior of the electrons at the Fermi surface. The electron states at energies more than a few kT below E_F are almost fully occupied. The Pauli exclusion principle requires that an electron can only be scattered into an empty state, and thus scattering of deep electrons is highly suppressed by the scarcity of empty states (scattering where the energy changes by more than a few kT is unlikely). Only the electrons near E_F undergo scattering. Likewise, under the action of an external field, only the electron occupation near E_F is altered. As a result, the density of states (DOS) near the Fermi level is most important for the metal electrical properties, and only those electrons in a small range ΔE around E_F actually contribute to electrical conduction. The density of these electrons is approximately $\mathrm{g}(E_\mathrm{F})\Delta E$ where $\mathrm{g}(E_\mathrm{F})$ is the DOS at the Fermi energy. From simple arguments, the overall conductivity can be shown to be [2.5]

$$\sigma = \frac{1}{3}e^2 v_\mathrm{F}^2 \tau\, \mathrm{g}(E_\mathrm{F})\,, \tag{2.31}$$

where v_F is the Fermi speed and τ is the scattering time of these Fermi electrons. According to (2.31), what is important is the density of states at the Fermi energy, $\mathrm{g}(E_\mathrm{F})$. For example, Cu and Mg are metals with valencies I and II. Classically, Cu and Mg atoms each contribute 1 and 2 conduction electrons respectively into the crystal. Thus, we would expect Mg to have higher conductivity. However, the Fermi level in Mg is where the top tail of the 3p-band overlaps the bottom tail of the 3s band where the density of states is small.

In Cu, on the other hand, E_F is nearly in the middle of the 4 s band where the density of states is high. Thus, Mg has a lower conductivity than Cu.

The scattering time τ in (2.31) assumes that the scattered electrons at E_F remain in the same energy band as, for example, in Cu, whose simplified energy band diagram around E_F is shown in Fig. 2.7a. In certain metals, there are two different energy bands that overlap at E_F. For example, in Ni, 3d and 4s bands overlap at E_F as shown in Fig. 2.7b. An electron can be scattered from the 4s to the 3d band and vice versa. Electrons in the 3d band have very low drift mobilities and effectively do not contribute to conduction so that only $g(E_F)$ of the 4s band operates in (2.31). Since 4s to 3d band is an additional scattering mechanism, by virtue of Matthiessen's rule, the effective scattering time τ for the 4s band electrons is shortened and hence σ from (2.31) is smaller. Thus, Ni has poorer conductivity than Cu.

Equation (2.31) does not assume a particular density of states model. If we now apply the *free electron model* for $g(E_F)$, and also relate E_F to the total number of conduction electrons per unit volume n, we would find that the conductivity is the same as that in the Drude model,

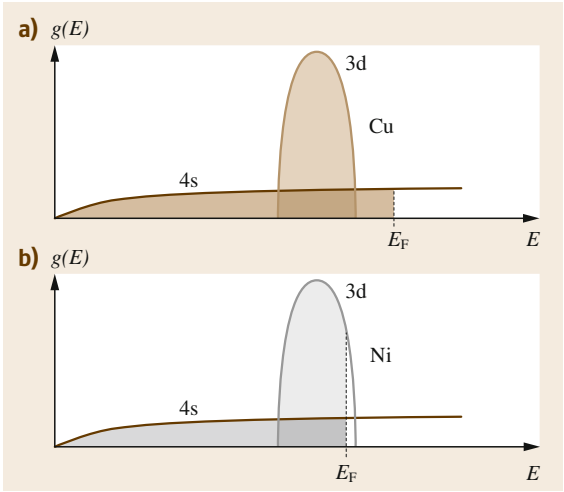


Fig. 2.7 Simplified energy band diagrams around E_F for **(a)** copper and **(b)** nickel

that is

$$\sigma = \frac{e^2 n \tau}{m_e} \,. \tag{2.32}$$

2.4 Solid Solutions and Nordheim's Rule

In an *isomorphous alloy* of two metals, i. e., a binary alloy which forms a binary *substitutional solid solution*, an additional mechanism of scattering appears, the *scattering off solute phase atoms*. This scattering contributes to lattice scattering, and therefore increases the overall resistivity. An important semi-empirical equation which can be used to predict the resistivity of an alloy is Nordheim's rule. It relates the impurity part of the resistivity ρ_I to the atomic fraction X of solute atoms in a solid solution via

$$\rho_I = CX(1-X) \,, \tag{2.33}$$

where the constant C is termed the Nordheim coefficient and represents the effectiveness of the solute atom in increasing the resistivity. Nordheim's rule was originally derived for crystals. Combining Nordheim's rule with Matthiessen's rule (2.23), the resistivity of an alloy of composition X should follow

$$\rho = \rho_{matrix} + CX(1-X) \,, \tag{2.34}$$

where ρ_{matrix} is the resistivity of the matrix due to scattering from thermal vibrations and from other defects, in the absence of alloying elements.

Nordheim's rule assumes that the solid solution has the solute atoms randomly distributed in the lattice. For sufficiently small amounts of impurity, experiments show that the increase in the resistivity ρ_I is nearly always simply proportional to the impurity concentration X, that is, $\rho_I \propto X$. For dilute solutions, Nordheim's rule predicts the same linear behavior, that is, $\rho_I = CX$ for $X \ll 1$.

Originally the theoretical model for ρ_I was developed by *Nordheim* [2.8] by assuming that the solute atoms simply perturb the periodic potential and thereby increase the probability of scattering. Quantum mechanical calculations for electron scattering within a single band, such as the s-band, at E_F show that

$$\rho_I \propto g(E_F) V^2_{scatter} X(1-X) \,, \tag{2.35}$$

where $g(E_F)$ is the DOS at E_F, and $V_{scatter}$ is matrix element for scattering from one wavefunction to another at the Fermi surface in the same band, which for an s-band is

$$V_{scatter} = \left\langle \psi_s^* \left| \Delta V \right| \psi_s \right\rangle , \tag{2.36}$$

where ΔV is the difference between the potentials associated with solvent and solute atoms, and ψ_s is the

Table 2.2 Nordheim coefficients (at 20 °C) for dilute alloys obtained from $\rho_i = CX$ and $X < 1$ at.%. Note: For many isomorphous alloys, C may be different at higher concentrations; that is, it may depend on the composition of the alloy [2.7, 9]. Maximum solubility data from [2.10]

Solute in solvent (element in matrix)	Nordheim coefficient (nΩ m)	Maximum solubility at 25 °C (at.%)
Au in Cu matrix	5500	100
Mn in Cu matrix	2900	24
Ni in Cu matrix	1250	100
Sn in Cu matrix	2900	0.6
Zn in Cu matrix	300	30
Cu in Au matrix	450	100
Mn in Au matrix	2410	25
Ni in Au matrix	790	100
Sn in Au matrix	3360	5
Zn in Au matrix	950	15

wavefunction of an electron in the s-band at E_F. It is clear that C is only independent of X if $g(E_F)$ and $V_{scatter}$ remain the same for various X which may not be true. For example, if the effective number of free electrons increases with X, E_F will be shifted higher, and C will not be constant.

Table 2.2 lists some typical Nordheim coefficients for various additions to copper and gold. The value of the Nordheim coefficient depends on the type of solute and the solvent. A solute atom that is drastically different in size to the solvent atom will result in a bigger increase in ρ_I and will therefore lead to a larger C. An important assumption in Nordheim's rule in (2.33) is that the alloying does not significantly vary the number of conduction electrons per atom in the alloy. Although this will be true for alloys with the same valency, that is, from the same column in the Periodic Table (e.g., Cu–Au, Ag–Au), it will not be true for alloys of different valency, such as Cu and Zn. In pure copper, there is just one conduction electron per atom, whereas each Zn atom can donate two conduction electrons. As the Zn content in brass is increased, more conduction electrons become available per atom. Consequently, the resistivity predicted by (2.34) at high Zn contents is greater than the actual value because C refers to dilute alloys. To get the correct resistivity from (2.34) we have to lower C, which is equivalent to using an *effective* Nordheim coefficient C_{eff} that decreases as the Zn content increases. In other cases, for example, in Cu–Ni alloys, we have to increase C at high Ni concentrations to account for additional electron scattering mechanisms that develop with Ni addition. Nonetheless, the Nordheim rule is still useful for predicting the resistivities of dilute alloys, particularly in the low-concentration region.

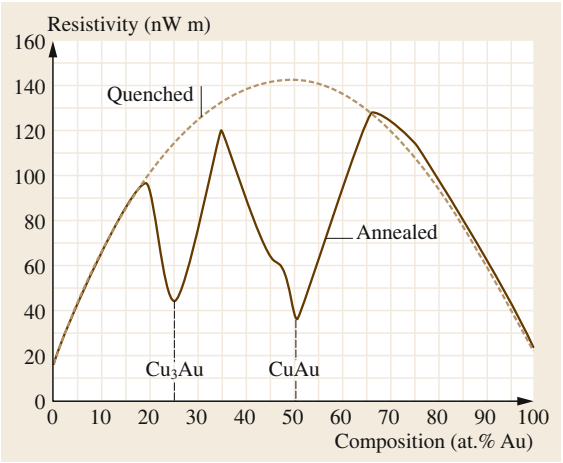


Fig. 2.8 Electrical resistivity versus composition at room temperature in Cu–Au alloys. The quenched sample (*dashed curve*) is obtained by quenching the liquid, and it has the Cu and Au atoms randomly mixed. The resistivity obeys the Nordheim rule. On the other hand, when the quenched sample is annealed or the liquid slowly cooled (*solid curve*), certain compositions (Cu_3Au and CuAu) result in an ordered crystalline structure in which Cu and Au atoms are positioned in an ordered fashion in the crystal and the scattering effect is reduced

In some solid solutions, at some concentrations of certain binary alloys, such as 75% Cu–25% Au and 50% Cu–50% Au, the annealed solid has an orderly structure; that is, the Cu and Au atoms are not randomly mixed, but occupy regular sites. In fact, these compositions can be viewed as a pure compound – like the solids Cu_3Au and CuAu. The resistivities of Cu_3Au and CuAu will therefore be less than the same composition random alloy that has been quenched from the melt. As a consequence, the resistivity ρ versus composition X curve does not follow the dashed parabolic curve throughout; rather, it exhibits sharp falls at these special compositions, as illustrated in Fig. 2.8. The *effective media approximation* may be used as an effective tool to estimate the resistivities of inhomogeneous media.

2.5 Carrier Scattering in Semiconductors

At low electric fields, ionized impurity scattering and phonon scattering predominate. Other types of scattering include carrier-carrier scattering, inter-valley scattering, and neutral impurity scattering; these may generally be ignored to a first approximation.

For phonon scattering, both *polar* and *non-polar phonon scattering* should be considered. In polar scattering, short wavelength oscillations of atoms on different sub-lattices vibrating out of phase produce an effective dipole moment proportional to the bond polarity. Since such vibrational modes are optically active (since this dipole moment can interact with an incident electromagnetic field), this type of lattice scattering is usually referred to as *polar optical phonon scattering*. Since a sub-lattice is necessary for optical modes, this scattering mechanism is not present in elemental semiconductors such as Si, Ge, or diamond.

Non-polar phonon scattering comes from long wavelength oscillations in the crystal, involving small displacements of tens to thousands of atoms. The wavelength depends on the material and its elastic properties. Such modes are very similar to sound vibrations and are thus referred to as acoustic modes. The associated atomic displacements correspond to an effective built-in strain, with local change in the lattice potential, causing carrier scattering known as *deformation potential acoustic phonon scattering*. Since the change in potential is relatively small, the scattering efficiency is relatively low as compared with polar optical phonon scattering.

Each scattering process contributes to the drift mobility according to *Matthiessen's rule*

$$\frac{1}{\mu_e} = \sum_i \frac{1}{\mu_i} \qquad (2.37)$$

as discussed in Sect. 2.3 above. Figure 2.9 shows the contributions of each scattering process for n-type ZnSe – a material that has applications in optoelectronics. At room temperature, polar optical phonon scattering and ionized impurity scattering dominate. These processes depend on the carrier concentration. The curve for ionized impurity scattering decreases markedly with increasing carrier concentration owing to the increasing concentration of ionized donors that supply these carriers. The ionized impurity scattering mobility is roughly inversely proportional to the concentration of ionized impurities. However, as the carrier concentration increases, as in a degenerate semiconductor, the average energy per carrier also increases (i. e., carriers move faster on average) and thus carriers are less susceptible to being scattered from the ionized impurity centers. In contrast, the polar phonon scattering rate is determined by the number of participating phonons which depends on the thermal energy available to create a given quantized vibrational mode.

The temperature dependence of the mobility may be estimated by considering the effect of temperature on ionized impurity and phonon scattering and combining these mechanisms using Matthiessen's rule. Phonon

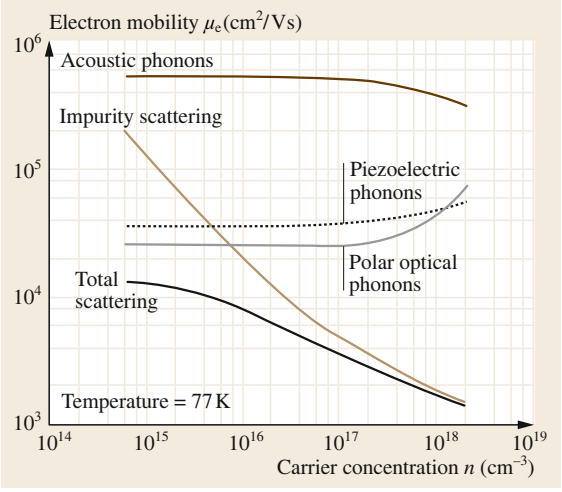


Fig. 2.9 Dependence of electron mobility on carrier concentration for ZnSe at 77 K. (After [2.11])

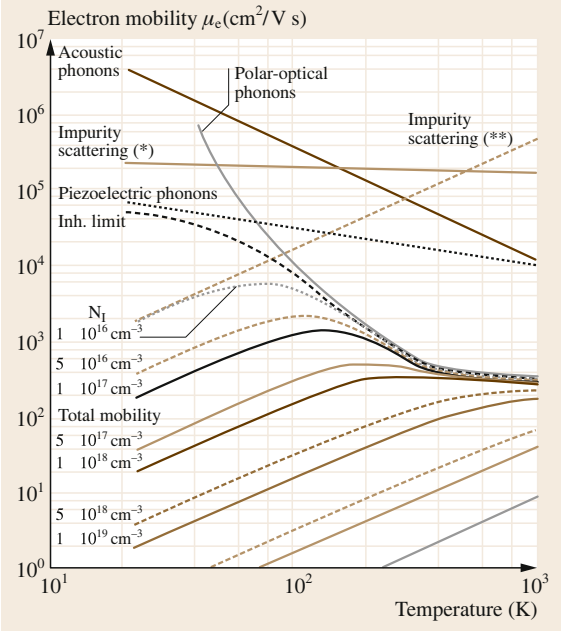


Fig. 2.10 Dependence of electron mobility on temperature and doping for ZnSe. (After [2.11])

scattering increases strongly with increasing temperature T due to the increase in the number of phonons resulting in a $T^{-3/2}$ dependence for the polar phonon mobility. For ionized impurity scattering, increasing the temperature increases the average carrier velocity and hence increases the carrier mobility for a set concentration of ionized impurities. Once a temperature is reached such that impurity ionization is complete, the ionized impurity based carrier mobility can be shown to increase with temperature T as approximately, $T^{+3/2}$. At low temperatures, the mobility is basically determined by ionized impurity scattering and at high temperatures by phonon scattering leading to a peaked curve. Invoking the previous discussions for the dependence of the total mobility on carrier concentration, it is clear that the peak mobility will depend on the doping level, and the peak location will shift to higher temperatures with increased doping as shown in Fig. 2.10.

2.6 The Boltzmann Transport Equation

A more rigorous treatment of charge transport requires a discussion of the *Boltzmann Transport Equation*. The electronic system is described by a distribution function $f(\boldsymbol{k}, \boldsymbol{r}, t)$ defined in such a way that the number of electrons in a six-dimensional volume element $\mathrm{d}^3\boldsymbol{k}\mathrm{d}^3\boldsymbol{r}$ at time t is given by $\frac{1}{4}\pi^{-3}f(\boldsymbol{k}, \boldsymbol{r}, t)\mathrm{d}^3\boldsymbol{k}\mathrm{d}^3\boldsymbol{r}$. In equilibrium, $f(\boldsymbol{k}, \boldsymbol{r}, t)$ depends only on energy and reduces to the Fermi distribution f_0 where the probability of occupation of states with momenta $+\boldsymbol{k}$ equals that for states with $-\boldsymbol{k}$, and $f_0(\boldsymbol{k})$ is symmetrical about $\boldsymbol{k} = 0$, giving no net charge transport. If an external field acts on the system (i. e., non-equilibrium), the occupation function $f(\boldsymbol{k})$ will become asymmetric in $\boldsymbol{k}$-space. If this non-equilibrium distribution function $f(\boldsymbol{k})$ is completely specified and appropriate boundary conditions supplied, the electronic transport properties can be completely determined by solving the *steady state Boltzmann transport equation* [2.12]

$$\boldsymbol{v}\cdot\nabla_{\boldsymbol{r}}f + \dot{\boldsymbol{k}}\cdot\nabla_{\boldsymbol{k}}f = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{c}} , \tag{2.38}$$

where:

1. $\boldsymbol{v}\cdot\nabla_{\boldsymbol{r}}f$ represents diffusion through a volume element $\mathrm{d}^3\boldsymbol{r}$ about the point $\boldsymbol{r}$ in phase space due to a gradient $\nabla_{\boldsymbol{r}}f$
2. $\dot{\boldsymbol{k}}\cdot\nabla_{\boldsymbol{k}}f$ represents drift through a volume element $\mathrm{d}^3\boldsymbol{k}$ about the point $\boldsymbol{k}$ in phase space due to a gradient $\nabla_{\boldsymbol{k}}f$ (for example, $\hbar\dot{\boldsymbol{k}} = e\left(\boldsymbol{E} + \frac{1}{c}\boldsymbol{v}\times\boldsymbol{B}\right)$ in the presence of electric and magnetic fields)
3. $(\partial f/\partial t)_{\mathrm{c}}$ is the collision term and accounts for the scattering of electrons from a point $\boldsymbol{k}$ (for example, this may be due to lattice or ionized impurity scattering).

Equation (2.38) may be simplified by using the *relaxation time approximation*

$$\left(\frac{\partial f}{\partial t}\right)_{\mathrm{c}} = \frac{\Delta f}{\tau} = -\frac{f - f_0}{\tau} \tag{2.39}$$

which is based on the assumption that for small changes in f carriers return to equilibrium in a characteristic time τ, dependent on the dominant scattering mechanisms. Further simplifications of (2.38) using (2.39) applicable for low electric fields lead to a simple equation connecting the mobility μ to the *average scattering time* $\langle\tau\rangle$

$$\mu \cong \frac{e\langle\tau\rangle}{m^*} . \tag{2.40}$$

The details of calculations may be found in various advanced textbooks, for example *Bube* [2.13], *Blatt* [2.4]. The average scattering time may be calculated assuming a *Maxwell–Boltzmann* distribution function and a parabolic band

$$\langle\tau\rangle = \frac{2}{3k_{\mathrm{B}}T}\frac{\int_0^\infty \tau(E)E^{3/2}e^{-E/k_{\mathrm{B}}T}\mathrm{d}E}{\int_0^\infty E^{3/2}e^{-E/k_{\mathrm{B}}T}\mathrm{d}E} . \tag{2.41}$$

Quantum mechanical perturbation theory can be used to calculate the carrier scattering rate for different processes i, giving,

$$\tau_i(E) = aE^{-\alpha} , \tag{2.42}$$

where a and α are constants and E is the electron energy. Substituting (2.42) into (2.41) gives

$$< \tau_i >= \frac{4a\Gamma\,(5/2-\alpha)}{3\pi^{1/2}(k_{\mathrm{B}}T)^{\alpha}} \tag{2.43}$$

in terms of the *gamma function* Γ. Using this approach, the expressions for the mobility for the case of lattice and impurity scattering may be easily found

$$\mu_{\mathrm{L}} \propto \frac{4e}{m^*\sqrt{9\pi k_{\mathrm{B}}}}T^{-\frac{3}{2}} , \tag{2.44}$$

$$\mu_{\mathrm{I}} \propto \frac{8ek_{\mathrm{B}}^{3/2}}{N_{\mathrm{I}}m^*\sqrt{\pi}}T^{+\frac{3}{2}} , \tag{2.45}$$

where N_{I} is the concentration of ionized impurities.

2.7 Resistivity of Thin Polycrystalline Films

Two new dominant scattering mechanisms must be considered in polycrystalline thin films – *scattering by grain boundaries* and *scattering at the surface*. The scattering by grain boundaries is schematically shown in Fig. 2.11. As a first approximation, the conduction electron may be considered free within a grain, but becomes scattered at the grain boundary. Its mean free path ℓ_{grains} is therefore roughly equal to the average grain size d. If $\lambda = \ell_{\text{crystal}}$ is the mean free path of the conduction electrons in the single crystal, then according to Matthiessen's rule

$$\frac{1}{\ell} = \frac{1}{\ell_{\text{crystal}}} + \frac{1}{\ell_{\text{grains}}} = \frac{1}{\lambda} + \frac{1}{d} \,. \tag{2.46}$$

The resistivity is inversely proportional to the mean free path which means the resistivity of the single crystal $\rho_{\text{crystal}} \propto 1/\lambda$ and resistivity of the polycrystalline sample $\rho \propto 1/\ell$. Thus,

$$\frac{\rho}{\rho_{\text{crystal}}} = 1 + \left(\frac{\lambda}{d}\right). \tag{2.47}$$

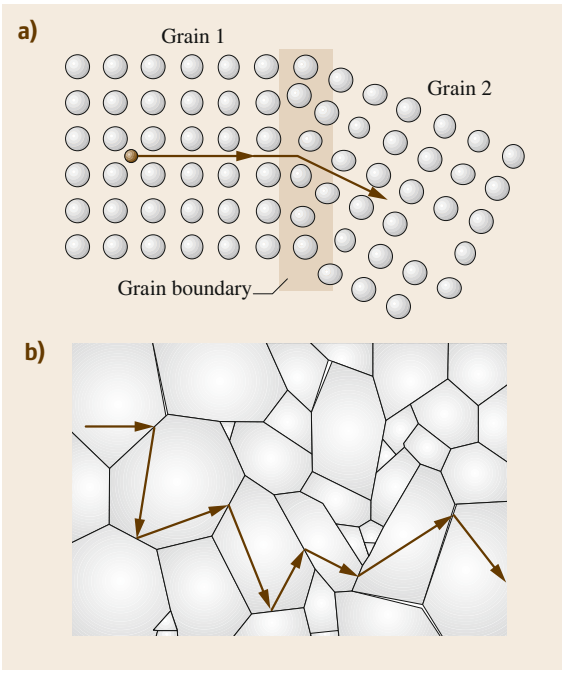


Fig. 2.11 **(a)** Grain boundaries cause electron scattering and therefore add to the resistivity according to Matthiessen's rule. **(b)** For a very grainy solid, the electron is scattered from grain boundary to grain boundary, and the mean free path is approximately equal to the mean grain diameter

Figure 2.12 clearly demonstrates that even simple considerations agree well with experimental data. However, in a more rigorous theory we have to consider a number of effects. It may take more than one scattering at a grain boundary to totally randomize the velocity so that we need to calculate the effective mean free path that accounts for how many collisions are needed to randomize the velocity. There is a possibility that the electron may be totally reflected back at a grain boundary (bounce back). Let σ be the conductivity of the polycrystalline (grainy) material and σ_{crystal} be the conductivity of the bulk single crystal. Suppose that the probability of reflection at a grain boundary is R. If d is the average grain size (diameter) then the two conductivities may be related using the *Mayadas* and *Shatzkes* formula [2.15]

$$\frac{\sigma}{\sigma_{\text{crystal}}} = 1 - \frac{3}{2}\beta + 3\beta^2 - 3\beta^3 \ln\left(1 + \frac{1}{\beta}\right)$$
$$\text{where} \quad \beta = \frac{\lambda}{d}\left(\frac{R}{1-R}\right), \tag{2.48}$$

The β in (2.48) represents the λ/d ratio adjusted for the reflection to transmission ratio of the electron at the grain boundary. When the grain size is large, β is small

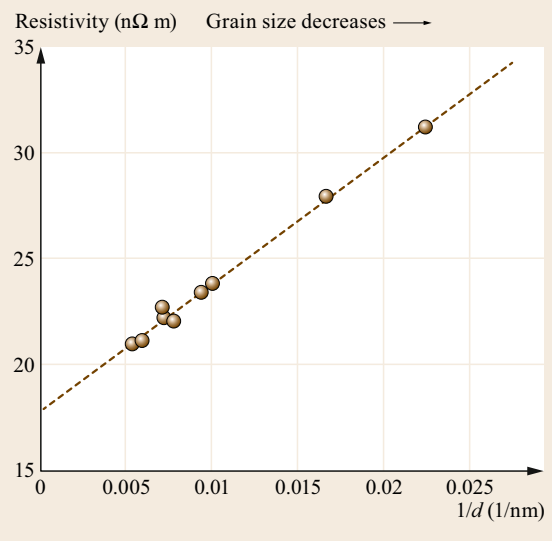


Fig. 2.12 Resistivity of Cu polycrystalline film versus reciprocal mean grain size (diameter), $1/d$. Film thickness $D = 250–900\,\text{nm}$ does not affect the resistivity. The *straight line* $\rho = 17.8\,\text{n}\Omega\,\text{m} + (595\,\text{n}\Omega\,\text{nm})(1/d)$. (After [2.14])

and (2.48) simples to

$$\frac{\rho}{\rho_{\text{crystal}}} \approx 1 + \frac{3}{2}\beta \,. \tag{2.49}$$

For highly polycrystalline films, the grain size would be small and $\beta \gg 1$

$$\frac{\rho}{\rho_{\text{crystal}}} \approx \frac{4}{3}\beta \,. \tag{2.50}$$

Equation (2.49) implies that the Matthiessen's rule is reasonably well obeyed when the grains are larger than the mean free path. For copper, typically R values are 0.24–0.40, and R is somewhat smaller for Al. Equation (2.48) for a Cu film with $R \approx 0.3$ predicts $\rho/\rho_{\text{crystal}} \approx 1.21$ for roughly $d \approx 3\lambda$ or a grain size $d \approx 120\,\text{nm}$ since in the bulk crystal $\lambda \approx 40\,\text{nm}$. *Tellier* et al. [2.16] have given extensive discussions of grain boundary scattering limited resistivity of thin films.

Scattering from the film surfaces must also be included in any resistivity calculation. It is generally assumed that the scattering from a surface is partially *inelastic*, that is the electron loses some of the velocity gained from the field. The inelastic scattering is also called *nonspecular*. (If the electron is elastically reflected from the surface just like a rubber ball bouncing from a wall, then there is no increase in the resistivity.) If the parameter p is the fraction of surface collisions which are specular (elastic) and if the thickness of film D is greater than λ, and $\sigma_{\text{bulk}} = 1/\rho_{\text{bulk}}$, than in accordance with Fuchs-Sondheimer equation [2.17, 18] the conductivity σ of the film is

$$\frac{\sigma}{\sigma_{\text{bulk}}} = 1 - \frac{3\lambda}{8D}(1-p) \,, \quad (\lambda/D > 1) \,. \tag{2.51}$$

If D is much shorter than ℓ,

$$\frac{\sigma}{\sigma_{\text{bulk}}} = \frac{3D}{4\lambda}\left[\ln\left(\frac{\lambda}{D}\right) + 0.423\right] \times (1+2p) \,, \quad (\lambda/D \ll 1) \,. \tag{2.52}$$

For purely nonspecular (inelastic) scattering, an approximate estimate can be obtained from

$$\frac{\rho}{\rho_{\text{bulk}}} \approx 1 + \frac{3}{8}\left(\frac{\lambda}{D}\right) . \tag{2.53}$$

Figure 2.13 shows the resistivity of Cu films as a function of film thickness D. The Cu thin films in this case are single crystal layers grown on the (001) surface of a single crystal of MgO (which is the substrate). As

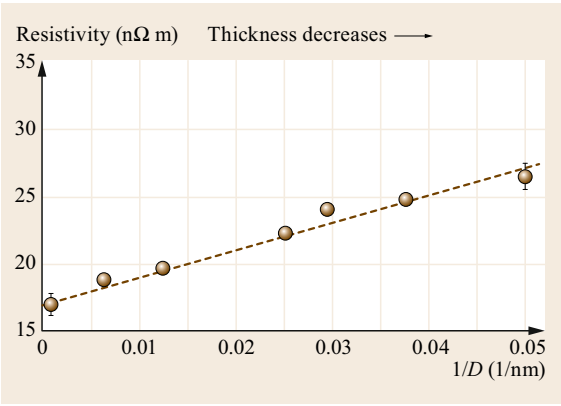


Fig. 2.13 The resistivity ρ_{film} of single crystal thin films of Cu against reciprocal film thickness $1/D$ at 25 °C. The films are grown on the surface of a single crystal of MgO and the best straight line is $\rho_{\text{film}} = 17.0\,\text{n}\Omega\,\text{m} + (200\,\text{n}\Omega\,\text{m}\,\text{nm})(1/D)$. (Data extracted from [2.19])

the film is a single crystal, there is no grain boundary scattering, and the observed increase in the resistivity with decreasing film thickness is due to the scattering of the electrons from the film surfaces. The experiments in Fig. 2.13 can be explained by the simplified Fuchs–Sondheimer equation with an average $p = 0.20$.

For elastic or specular scattering $p = 1$ and there is no change in the conductivity. The value of p depends on the film preparation method (e.g., sputtering, epitaxial growth etc.) and the substrate on which the film has been deposited. Table 2.3 summarizes the resistivities of thin Cu and Au gold films deposited by various preparation techniques. Notice the large difference between the Au films deposited on a noncrystalline glass substrate and on a crystalline mica substrate. Such differences between films are typically attributed to different values of p. The p-value can also change (increase) when the film is annealed. Obviously, the polycrystallinity of the film will also affect the resistivity as discussed above. Typically, most epitaxial thin films, unless very thin ($D \ll \ell$), deposited onto heated crystalline substrates exhibit highly specular scattering with $p = 0.8 - 1$.

It is generally very difficult to separate the effects of surface and grain boundary scattering in thin polycrystalline films; the contribution from grain boundary scattering is likely to exceed that from the surfaces. In any event, both contributions, by Matthiessen's general rule, increase the overall resistivity. Figure 2.12 shows an example in which the resistivity ρ_{film} of thin Cu polycrystalline films is due to grain boundary scattering, and thickness has no effect (D was 250–900 nm and much greater than λ). The resistivity ρ_{film} is plotted against the reciprocal mean grain size $1/d$, which then follows the

Table 2.3 Resistivities of some thin Cu and Au films at room temperature. PC: Polycrystalline film; RT is room temperature; D = film thickness; d = average grain size. FS and MS refer to Fuchs–Sondheimer and Mayadas–Shatzkes descriptions of thin film resistivity. At RT for Cu, $\lambda = 38-40$ nm, and for Au $\lambda = 36-38$ nm. Data selectively combined from various sources [2.14, 19–22, 24–27]

Film	D (nm)	d (nm)	ρ (nΩ m)	Comment
Cu films				
Cu encapsulated in SiO_2/Cu/SiO_2 [2.20]	45.3 31.7	101 41	28.0 35.5	DC sputtering. Cu film sandwiched in SiO_2/Cu/SiO_2, annealed at 150 °C. MS with $R = 0.50$
Cu encapsulated in SiO_2/Ta/Cu/Ta/SiO_2 [2.21]	34.2	39.4	37.3	DC sputtering. Cu film sandwiched in SiO_2/Ta/Cu/Ta/SiO_2,annealed at 600 °C. MS with $R = 0.47$
Cu on Ta/SiO_2/Si(001) [2.21]	35	40	31	Sputtering and then annealing at 350 °C
Cu (single crystal) on TiN/MgO (001) [2.22]	40 13	∞ ∞	21 29.7	Cu single crystal epitaxial layer on TiN(100) on MgO surface (001). Ultra-high vacuum, DC sputtering. FS with $p \approx 0.6$ in vacuum, $p = 0$ in air.
Cu on TiN, W and TiW [2.14]	> 250	186 44	21 31	CVD (chemical vapor deposition). Substrate temperature 200 °C, ρ depends on d not $D = 250-900$ nm. MS.
Cu on crystalline Si (100) surface [2.23]	51.2 17.2 8.6	$D/2.3$	32.2 70.5 126	Ion beam deposition with negative substrate bias. Resistivity follows FS and MS equations combined; surface and grain boundary scattering. FS and MS, $p = 0$, $R = 0.24$, $d = D/2.3$
Au films				
Au epitaxial film on mica	30		25	Single crystal on mica. $p \approx 0.8$, highly specular scattering
Au PC film on mica	30		54	PC. Sputtered on mica. p is small
Au film on glass	30		70	PC. Evaporated onto glass. p is small, nonspecular scattering
Au on glass [2.24]	40 40	8.5 3.8	92 189	PC. Sputtered films. $R = 0.27-0.33$

expected linear behavior in (2.49). On the other hand, Fig. 2.13 shows the resistivity of Cu films as a function of film thickness D. In this case, the thin films are grown epitaxially (as a single crystal) on MgO single crystal surfaces and the grain boundary scattering is absent, which is a clear cut case for surface scattering; the transport is described by (2.51). Chapter 28 provides a more advanced treatment of conduction in thin films.

2.8 Inhomogeneous Media: Effective Media Approximation

The *effective media approximation* (EMA) attempts to estimate the properties of inhomogeneous mixture of two or more components using the known physical properties of each component. The general idea of any EMA is to substitute for the original inhomogeneous mixture some imaginary homogeneous substance – the *effective medium* (EM) – whose response to an external excitation is the same as that of the original mixture. The EMA is widely used for investigations of non-uniform objects in a variety of applications such as composite materials [2.28, 29], microcrystalline and amorphous semiconductors [2.30–33], light scattering [2.34], conductivity of dispersed ionic semiconductors [2.35] and many others.

Calculations of the conductivity and dielectric constant of two component mixtures are reviewed by *Reynolds* and *Hough* [2.36] and summarized by *Rossiter* [2.1]. For such a mixture we assume that the two components α and β are randomly distributed in space with volume fractions of χ_α and $\chi_\beta = 1 - \chi_\alpha$. The dielectric properties are described by an *effective permittivity* ε_{eff} given by the ratio

$$\varepsilon_{\text{eff}} = \frac{\langle D \rangle}{\langle E \rangle} , \tag{2.54}$$

where $\langle E \rangle$ is the average electric field and $\langle D \rangle$ is the average displacement field. The displacement field averaged over a large volume may be calculated from

$$\begin{aligned} \langle D \rangle &= \frac{1}{V} \int_V D \mathrm{d}v = \frac{1}{V} \left(\int_{V_\alpha} D \mathrm{d}v + \int_{V_\beta} D \mathrm{d}v \right) \\ &= \chi_\alpha \langle D_\alpha \rangle + \chi_\beta \langle D_\beta \rangle , \end{aligned} \tag{2.55}$$

where $\langle D_\alpha \rangle$ and $\langle D_\beta \rangle$ are the average displacement fields inside regions of the respective components and V_α and V_β are their volumes. Likewise the electric field

is given by

$$\langle E\rangle = \chi_\alpha \langle E_\alpha\rangle + \chi_\beta \langle E_\beta\rangle . \tag{2.56}$$

From (2.54) one gets

$$\varepsilon_{\text{eff}} = \varepsilon_\beta + \left(\varepsilon_\alpha - \varepsilon_\beta\right) \chi_\alpha f_\alpha \tag{2.57}$$

or

$$\left(\varepsilon_{\text{eff}} - \varepsilon_\alpha\right) \chi_\alpha f_\alpha + \left(\varepsilon_{\text{eff}} - \varepsilon_\beta\right) \chi_\beta f_\beta = 0 \tag{2.58}$$

where ε_α and ε_β are the permittivities of the components and $f_\alpha = \langle E_\alpha\rangle / \langle E\rangle$ and $f_\beta = \langle E_\beta\rangle / \langle E\rangle$ are so-called *field factors*. The choice between (2.57) and (2.58) depends on particle geometry. Equation (2.57) is better when the particles of component are dispersed in a continuous medium β. Equation (2.58) is preferred when the particle size of the two components is of the same order of magnitude.

The field factors can be calculated analytically only for phase regions with special specific geometries. The field factor for ellipsoids is given by (*Stratton* [2.48])

$$f_\alpha = \sum_{i=1}^{3} \frac{\cos^2 \alpha_i}{1 + A_i \left(\frac{\varepsilon_\alpha}{\varepsilon^*} - 1\right)} , \tag{2.59}$$

where α_i are the angles between the ellipsoid axes and the applied field and the constants A_i depend on the axial ratios of the ellipsoids; they must satisfy

$$\sum_{i=1}^{3} A_i = 1 .$$

For a spheroid, $A_2 = A_3 = A$ and $A_1 = 1 - 2A$. For a random orientation of spheroids, $\cos^2\alpha_1 = \cos^2\alpha_2 = \cos^2\alpha_3 = \frac{1}{3}$. In the case of long particles with aligned axes, $\cos^2\alpha_1 = \cos^2\alpha_2 = \frac{1}{2}$ and $\cos^2\alpha_3 = 0$. The values of parameters entering (2.59) can be found in Table 2.4 which shows a set of *mixture rules*, i. e., a set of formulae allowing one to calculate ε_{eff} for some specific cases (such as spheres, rods, lamellae, etc.). The presence of some degree of orientation somewhat simplifies the calculations as shown in the Table 2.5.

The same formulae can be used to calculate the conductivity of mixtures by substituting the appropriate conductivity σ for ε. For some special cases, the mixture rules of Table 2.5 lead to very simple formulae which allows one to calculate the conductivity of inhomogeneous alloys with those specific geometries. These *mixture rules* are summarized in Table 2.6.

The most general approach to calculating the effective dielectric permittivity comes from

$$\varepsilon_{\text{eff}} = \varepsilon_\beta \left(1 - \chi_\alpha \int_0^1 \frac{G(L)}{t - L} \mathrm{d}L\right) , \tag{2.60}$$

Table 2.4 Mixture rules for randomly oriented particles. (Compiled from *Reynolds* and *Hough* [2.36])

Particle shape	Mixture rule	Factors in (2.59)		References
		A	ε^*	
Spheres	$\frac{\varepsilon_{\text{eff}} - \varepsilon_\beta}{\varepsilon_{\text{eff}} + 2\varepsilon_\beta} = \chi_\alpha \frac{\varepsilon_\alpha - \varepsilon_\beta}{\varepsilon_\alpha + 2\varepsilon_\beta}$	$\frac{1}{3}$	ε_β	[2.37–41]
Spheres	$\frac{\varepsilon_{\text{eff}} - \varepsilon_\beta}{3\varepsilon_\beta} = \chi_\alpha \frac{\varepsilon_\alpha - \varepsilon_\beta}{\varepsilon_\alpha + 2\varepsilon_\beta}$	$\frac{1}{3}$	ε_β	[2.42]
Spheres	$\chi_\alpha \frac{\varepsilon_\alpha - \varepsilon_{\text{eff}}}{\varepsilon_\alpha + 2\varepsilon_{\text{eff}}} + \chi_\beta \frac{\varepsilon_\beta - \varepsilon_{\text{eff}}}{\varepsilon_\beta + 2\varepsilon_{\text{eff}}} = 0$	$\frac{1}{3}$	ε_{eff}	[2.43]
Spheres	$\frac{\varepsilon_{\text{eff}} - \varepsilon_\beta}{3\varepsilon_{\text{eff}}} = \chi_\alpha \frac{\varepsilon_\alpha - \varepsilon_\beta}{\varepsilon_\alpha + 2\varepsilon_\beta}$	$\frac{1}{3}$	ε_{eff}	[2.44]
Spheroids	$\varepsilon_{\text{eff}} = \varepsilon_\beta + \frac{\chi_\alpha}{3(1-\chi_\alpha)} \sum_{i=1}^{3} \frac{\varepsilon_\alpha - \varepsilon_{\text{eff}}}{1 + A_i \left(\frac{\varepsilon_\alpha}{\varepsilon_\beta} - 1\right)}$	A	ε_β	[2.45]
Spheroids	$\varepsilon_{\text{eff}} = \varepsilon_\beta + \frac{\chi_\alpha}{3} \sum_{i=1}^{3} \frac{\varepsilon_\alpha - \varepsilon_\beta}{1 + A_i \left(\frac{\varepsilon_\alpha}{\varepsilon_{\text{eff}}} - 1\right)}$	A	ε_{eff}	[2.46]
Lamellae	$\varepsilon_{\text{eff}}^2 = \frac{2\left(\varepsilon_\alpha \chi_\alpha + \varepsilon_\beta \chi_\beta\right) - \varepsilon_{\text{eff}}}{\frac{\varepsilon_\alpha}{\chi_\alpha} + \frac{\varepsilon_\beta}{\chi_\beta}}$	0	ε_{eff}	[2.43]
Rods	$5\varepsilon_{\text{eff}}^3 + \left(5\varepsilon_p' - 4\varepsilon_p\right)\varepsilon_{\text{eff}}^2 - \left(\chi_\alpha \varepsilon_\alpha^2 + 4\varepsilon_\alpha \varepsilon_\beta + \chi_\beta \varepsilon_\beta^2\right) - \varepsilon_\alpha \varepsilon_\beta \varepsilon_p = 0$ where $\frac{1}{\varepsilon_p'} = \frac{\chi_\alpha}{\varepsilon_\beta} + \frac{\chi_\beta}{\varepsilon_\alpha}$ and $\frac{1}{\varepsilon_p} = \frac{\chi_\alpha}{\varepsilon_\alpha} + \frac{\chi_\beta}{\varepsilon_\beta}$	$\frac{1}{2}$		[2.47]

Table 2.5 Mixture rules for partially oriented particles. (Compiled from *Reynolds* and *Hough* [2.36])

Particle shape	Formula	Factors in (2.59)				References
		A	ε^*	$\cos\alpha_1 = \cos\alpha_2$	$\cos\alpha_3$	
Parallel cylinders	$\frac{\varepsilon_{\text{eff}}-\varepsilon_\beta}{\varepsilon_{\text{eff}}+\varepsilon_\beta} = \chi_\alpha \frac{\varepsilon_\alpha-\varepsilon_\beta}{\varepsilon_\alpha+\varepsilon_\beta}$	$\frac{1}{2}$	ε_β	$\frac{1}{2}$	0	[2.40, 41]
Parallel cylinders	$\chi_\alpha \frac{\varepsilon_\alpha-\varepsilon_{\text{eff}}}{\varepsilon_\alpha+\varepsilon_{\text{eff}}} + \chi_\beta \frac{\varepsilon_\beta-\varepsilon_{\text{eff}}}{\varepsilon_\beta+\varepsilon_{\text{eff}}}$	$\frac{1}{2}$	ε_{eff}	$\frac{1}{2}$	0	[2.43]
Parallel lamellae (with two axes randomly oriented)	$\varepsilon_{\text{eff}}^2 = \frac{\varepsilon_\alpha\chi_\alpha+\varepsilon_\beta\chi_\beta}{\frac{\varepsilon_\alpha}{\chi_\alpha}+\frac{\varepsilon_\beta}{\chi_\beta}}$	0	ε_{eff}	$\frac{1}{2}$	0	[2.43]
Lamellae with all axes aligned (current lines are perpendicular to lamellae planes)	$\frac{1}{\varepsilon_{\text{eff}}} = \frac{\chi_\alpha}{\varepsilon_\alpha}+\frac{\chi_\beta}{\varepsilon_\beta}$	0	ε_{eff}	0	1	[2.49]
Lamellae with all axes aligned (current lines are parallel to lamellae planes)	$\varepsilon_{\text{eff}} = \varepsilon_\alpha\chi_\alpha+\varepsilon_\beta\chi_\beta$	0	ε_{eff}	1	0	[2.49, 50]
Spheroids with all axes aligned (current lines are parallel to one of the axes)	$\varepsilon_{\text{eff}} = \varepsilon_\beta + \frac{\chi_\alpha(\varepsilon_\alpha-\varepsilon_\beta)}{1+A\left(\frac{\varepsilon_\alpha}{\varepsilon_\beta}-1\right)}$	A	ε_β	0	1	[2.51]
Spheroids with all axes aligned (current lines are parallel to one of the axes)	$\frac{\varepsilon_{\text{eff}}}{\varepsilon_\beta} = 1 + \frac{\chi_\alpha}{\left(\frac{\varepsilon_\alpha}{\varepsilon_\beta}-1\right)^{-1}+A\chi_\alpha}$	A	ε_β	0	1	[2.52]

Table 2.6 Conductivity/resistivity mixture rules. (After [2.5])

Particle shape	Formula	Commentary
Lamellae with all axes aligned (current lines are perpendicular to lamellae planes)	$\rho_{\text{eff}} = \chi_\alpha\rho_\alpha+\chi_\beta\rho_\beta$	*Resistivity mixture rule*: ρ_α and ρ_β are the resistivities of two phases and ρ_{eff} is the effective resistivity of mixture
Lamellae with all axes aligned (current lines are parallel to lamellae planes)	$\sigma_{\text{eff}} = \chi_\alpha\sigma_\alpha+\chi_\beta\sigma_\beta$	*Conductivity mixture rule*: σ_α and σ_β are the conductivities of two phases and σ_{eff} is the effective conductivity of mixture
Small spheroids (α-phase) in medium (β-phase)	$\rho_{\text{eff}} = \rho_\beta \frac{(1+\frac{1}{2}\chi_\alpha)}{(1-\chi_\alpha)}$	$\rho_\alpha > 10\rho_\beta$
Small spheroids (α-phase) in medium (β-phase)	$\rho_{\text{eff}} = \rho_\beta \frac{(1-\chi_\alpha)}{(1+2\chi_\alpha)}$	$\rho_\alpha < 0.1\rho_\beta$

Table 2.7 Mixture rules and corresponding spectral functions $G(L)$

Mixture rule by *Bruggemann* [2.43]: $\chi_\alpha \frac{\varepsilon_\alpha-\varepsilon_{\text{eff}}}{\varepsilon_\alpha+2\varepsilon_{\text{eff}}} + \chi_\beta \frac{\varepsilon_\beta-\varepsilon_{\text{eff}}}{\varepsilon_\beta+2\varepsilon_{\text{eff}}} = 0$

$$G(L) = \frac{3\chi_\alpha-1}{2\chi_\alpha}\delta(L)\Theta(3\chi_\alpha-1) + \frac{3}{4\pi\chi_\alpha L}\sqrt{(L-L^-)(L^+-L)}\Theta(L-L^-)\Theta(L^+-L)$$

where $L^{+/-} = \frac{1}{3}\left(1+\chi_\alpha \pm 2\sqrt{2\chi_\alpha-2\chi_\alpha^2}\right)$

Mixture rule by *Maxwell-Garnett* [2.53]: $\frac{\varepsilon_{\text{eff}}-\varepsilon_\beta}{\varepsilon_{\text{eff}}+2\varepsilon_\beta} = \chi_\alpha \frac{\varepsilon_\alpha-\varepsilon_\beta}{\varepsilon_\alpha+2\varepsilon_\beta}$

$$G(L) = \delta\left(L-\frac{1-\chi_\alpha}{3}\right)$$

Table 2.7 (continued)

Mixture rule by *Looyenga* [2.54]: $\varepsilon_{\text{eff}}^{1/3} = \chi_\alpha \varepsilon_\alpha^{1/3} + \chi_\beta \varepsilon_\beta^{1/3}$

$$G(L) = \chi_\alpha^2 \delta(L) + \frac{3\sqrt{3}}{2\pi}\left(\chi_\beta^2 \left|\frac{L-1}{L}\right|^{1/3} + \chi_\alpha \chi_\beta \left|\frac{L-1}{L}\right|^{2/3}\right)$$

Mixture rule by *Monecke* [2.55]: $\varepsilon_{\text{eff}} = \dfrac{2\left(\chi_\alpha \varepsilon_\alpha + \chi_\beta \varepsilon_\beta\right)^2 + \varepsilon_\alpha \varepsilon_\beta}{(1+\chi_\alpha)\varepsilon_\alpha + (2-\chi_\alpha)\varepsilon_\beta}$

$$G(L) = \frac{2\chi_\alpha}{1+\chi_\alpha}\delta(L) + \frac{1-\chi_\alpha}{1+\chi_\alpha}\delta\left(L - \frac{1+\chi_\alpha}{3}\right)$$

Mixture rule for hollow sphere equivalent by *Bohren* and *Huffman* [2.56]:

$$\varepsilon_{\text{eff}} = \varepsilon_\alpha \frac{(3-2f)\varepsilon_\beta + 2f\varepsilon_\alpha}{f\varepsilon_\beta + (3-f)\varepsilon_\alpha}$$

$$G(L) = \frac{2}{3-f}\delta(L) + \frac{1-f}{3-f}\delta\left(L - \frac{3-f}{3}\right)$$

where $f = 1 - \dfrac{r_{\text{i}}^3}{r_{\text{o}}^3}$ and $r_{\text{i/o}}$ is the inner/outer radius of the sphere

where $t = \varepsilon_\beta/(\varepsilon_\beta - \varepsilon_\alpha)$ and $G(L)$ is the *spectral function* which describes the geometry of particles in terms of the *depolarization factor* L. Equation (2.60) is commonly known as the *Bergman* representation [2.57] and represents a general approach to modeling a heterogeneous mixture in terms of effective properties [2.58]. The advantage of the spectral representation is that it distinguishes between the influence of geometrical quantities and that of the dielectric properties of the components on the effective behavior of the system. Although the spectral function $G(L)$ is generally unknown for an arbitrary two-phase composite, it's analytically known or can be numerically derived for any existing mixture rule. Examples of spectral functions and corresponding solutions are shown in Table 2.7.

2.9 The Hall Effect

The Hall effect is closely related to the phenomenon of conductivity and is observed as the occurrence of a voltage appearing across a conductor carrying an electric current in a magnetic field. The schematic of the experiment is shown in the Fig. 2.14. The effect is often characterized by the *Hall coefficient*

$$R_{\text{H}} = \frac{E_y}{J_x B_z}\,, \tag{2.61}$$

where E_y is the Hall effect electric field in the y-direction, J_x is the current density in the x-direction and B_z is magnetic field in the z-direction.

The *Hall effect for ambipolar conduction* in a sample where there are both negative and positive charge carriers, e.g., electrons and holes in a semiconductor, involves not only the concentrations of electrons and holes, n and p respectively, but also the electron and hole drift mobilities, μ_{e} and μ_{h}. In the first approximation, the Hall coefficient can be calculated in the following way. Both charge carriers experience a Lorentz force in the same direction since they are drifting in the opposite directions but of course have opposite signs as illustrated in Fig. 2.14. Thus, both

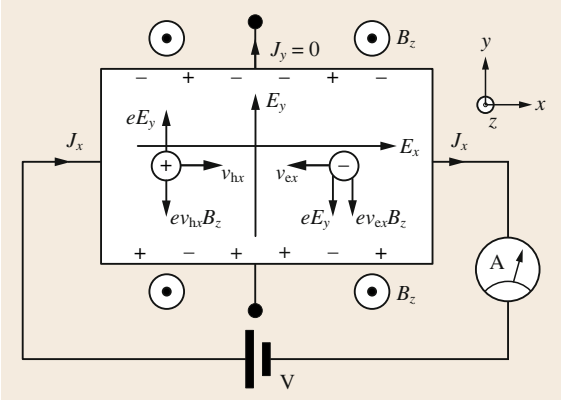


Fig. 2.14 Hall effect for ambipolar conduction. The magnetic field B_z is out of the plane of the paper. Both electrons and holes are deflected toward the bottom surface of the conductor and so the Hall voltage depends on the relative mobilities and concentrations of electrons and holes

holes and electrons accumulate near the bottom surface. The magnitude of the Lorentz force, however, will be different since the drift mobilities and hence drift velocities will be different. Once equilibrium is reached, there should be no current flowing in the y-direction as we have an open circuit. The latter physical arguments lead to the following Hall coefficient [2.5],

$$R_{\mathrm{H}} = \frac{p\mu_{\mathrm{h}}^2 - n\mu_{\mathrm{e}}^2}{e\,(p\mu_{\mathrm{h}} + n\mu_{\mathrm{e}})^2}$$

$$\text{or} \quad R_{\mathrm{H}} = \frac{p - nb^2}{e\,(p + nb)^2}, \tag{2.62}$$

where $b = \mu_{\mathrm{e}}/\mu_{\mathrm{h}}$.

It is clear that the Hall coefficient depends on both the drift mobility ratio and the concentrations of holes and electrons. For $p > nb^2$, R_{H} will be positive and for $p < nb^2$, it will be negative. Note that the carrier concentration is not zero when the Hall coefficient is zero but rather $n/p = (\mu_{\mathrm{h}}/\mu_{\mathrm{e}})^2$. As an example, Fig. 2.15 shows the dependence of Hall coefficient versus electron concentration for a single crystal silicon. The calculations are based on (2.62) and the law of mass action

$$np = n_{\mathrm{i}}^2\,, \tag{2.63}$$

where n_{i} is the electron concentration in the intrinsic semiconductor.

In the case of *monopolar conduction*, e.g., conduction in metals or in doped semiconductors, (2.62) reduces to

$$R_{\mathrm{H}} = -\frac{1}{en}\,, \quad (\text{for } n \gg p) \tag{2.64}$$

or

$$R_{\mathrm{H}} = \frac{1}{ep}\,, \quad (\text{for } p \gg n)\,. \tag{2.65}$$

Therefore, (considering as an example a n-type semiconductor where $\sigma = ne\mu_{\mathrm{n}}$) one can write

$$\mu_{\mathrm{H}} = \frac{\sigma}{ne} = -\sigma R_{\mathrm{H}} \tag{2.66}$$

which provides a simple expression for determining the electron mobility known as the *Hall mobility*. Note, however, that the Hall mobility may differ from the drift mobility discussed in the previous sections. The difference arises from the carriers in a semiconductor having

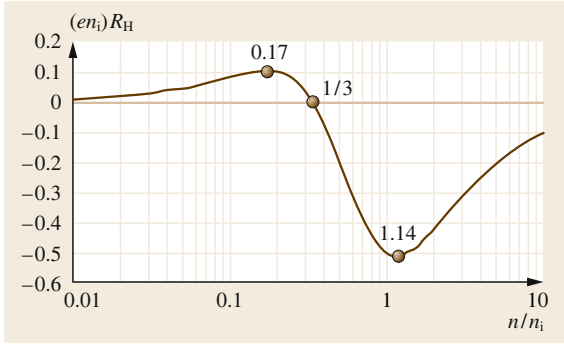


Fig. 2.15 Normalized Hall coefficient versus normalized electron concentration plot for single-crystal silicon. The values 0.17, 1.14 and 0.33 shown are the n/n_{i} values when the magnitude R reaches maxima and zero respectively. In single-crystal silicon, $n_{\mathrm{i}} = 1.5\times10^{10}\,\mathrm{cm}^{-3}$, $\mu_{\mathrm{e}} = 1350\,\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$ and $\mu_{\mathrm{h}} = 450\,\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$

a distribution of energies. An average is used to describe the effect of carriers occupying different allowed states. (This is distinct from the earlier discussions of mobility where it was assumed that all the carriers have the same mean free time between collisions.) To include this, it is necessary to use a formal analysis based on the Boltzmann transport equation, as discussed in Sect. 2.6. If we express the averaging of the carriers with energy E and distribution function $f(E)$, we can write the energy-averaged $\tau(E)$ (i. e., $\langle\tau\rangle$) as

$$\langle\tau\rangle = \frac{\int \tau(E) f(E)\mathrm{d}E}{\int f(E)\mathrm{d}E} \tag{2.67}$$

and the energy-averaged τ^2 (i. e., $\langle\tau^2\rangle$) as

$$\langle\tau\rangle = \frac{\int \tau^2(E) f(E)\mathrm{d}E}{\int f(E)\mathrm{d}E}\,. \tag{2.68}$$

The rigorous analysis [2.12] shows that the Hall mobility μ_{H} in terms of the drift mobility μ_{d} is

$$\mu_{\mathrm{H}} = r_{\mathrm{H}} \times \mu_{\mathrm{d}}\,, \tag{2.69}$$

where r_{H} is the *Hall factor* given by the ratio

$$r_{\mathrm{H}} = \frac{\langle\tau^2\rangle}{\langle\tau\rangle^2}\,. \tag{2.70}$$

The magnitude of the *Hall factor* will depend on the magnitude of the scattering mechanisms that contribute to τ, but usually at low magnetic fields, $r_{\mathrm{H}} \approx 1$ and the two mobilities are identical.

2.10 High Electric Field Transport

The previous sections focused on carrier transport in weak electric fields where the energy gained by carriers from the field is lost to the lattice through collisions with phonons or ionized impurities. However at higher fields, the efficiency of collision mechanisms diminishes and the carrier system contains more energy than the lattice. The carriers are then called *hot* with effective temperatures T_e for electrons and T_h for holes. In this case, the drift velocity no longer obeys Ohm's law, and becomes non-linear in the applied field with a clear tendency to saturation due to the appearance of a new dissipation mechanism involving optical phonon generation. Figure 2.16 shows the drift velocity saturation for both electrons and holes in Si and electrons in GaAs; GaAs shows a region of electron velocity overshoot and then negative differential resistivity due to inter-valley scattering, i. e., the transfer of electrons from the Γ minimum to the L conduction band minimum.

Solving the Boltzmann transport equation by analogy with (2.6) and (2.7) the electron (or hole) mobility may be defined as

$$\mu = \frac{e\,\langle\tau(T_e)\rangle}{m^*}\,, \tag{2.71}$$

where e is elementary charge and m^* is the electron (or hole) effective mass and $\langle\tau(T_e)\rangle$ is the mean free time

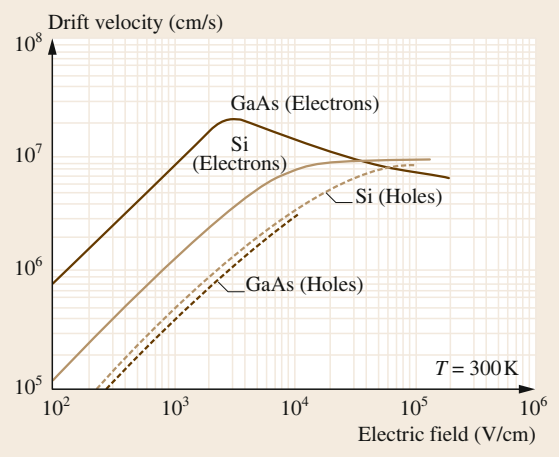


Fig. 2.16 Dependence of carrier drift velocity on electric field for GaAs and Si. (After [2.59])

which now strongly depends on T_e. Therefore, the high-field mobility is related to the low field mobility μ_0 by

$$\mu = \mu_0\left(\frac{T_e}{T}\right)^{\beta}\,, \tag{2.72}$$

where β depends on electric field and scattering mechanisms. Thus, in order to determine the dependence of the mobility on electric field, the dependence of the effective carrier temperature on field is required. This may be found by using the time-dependent Boltzmann transport equation. Suppose that F is the field, then for non-degenerate conditions with $T_e \gg T$

$$\frac{T_e}{T} \propto F^{-\frac{2}{2\beta-1}} \tag{2.73}$$

and hence

$$\mu(E) \propto \mu_0 F^{-\frac{2}{2\beta-1}}\,. \tag{2.74}$$

For acoustic phonon scattering, $\beta = -3/2$ and the drift mobility shows no saturation, increasing with field as $F^{1/4}$. Saturation in the drift velocity may be achieved only when $\beta \to \infty$ due to optical phonon scattering where large energy changes are involved. The saturation velocity v_s (related to the saturation mobility by $v_s = F\mu_s$) may be calculated using the energy and momentum rate equations for optical phonon scattering

$$\frac{\mathrm{d}\,\langle E\rangle}{\mathrm{d}t} = eFv_s - \frac{E_{op}}{\tau_e} \tag{2.75}$$

$$\frac{\mathrm{d}\,\langle m^* v_s\rangle}{\mathrm{d}t} = eF - \frac{m^* v_s}{\tau_m} \tag{2.76}$$

where E_{op} is the optical phonon energy, τ_e and τ_m are the energy and momentum relaxation times, respectively. At steady state and for not extremely high fields, one may assume that $\tau_e \approx \tau_m$. (It is worth noting that at the highest electric fields $\tau_e > \tau_m$ and may lead to the appearance of avalanche, as discussed in Sect. 2.11). Therefore, the solution of (2.75) and (2.76) is

$$v_s = \left(\frac{E_{op}}{m^*}\right)^{1/2}\,. \tag{2.77}$$

in agreement with the experimental values shown in Fig. 2.16.

2.11 Impact Ionization

At very high electric field (in the range 2×10^5 V/cm or larger) a new possibility appears: a carrier may have kinetic energy in excess of the binding energy of a valence electron to its parent atom. In colliding with an atom, such a carrier can rupture the covalent bond and produce an electron-hole pair. This process is called *impact ionization* and is characterized by the *impact ionization coefficient* α (α_e for electrons and α_h for holes). The released electrons and holes may, in turn, impact ionize more atoms producing new electron-hole pairs. This process is called *avalanche* and is characterized by the *multiplication factor*, M, which is the ratio of number of collected carriers to the number of initially injected carriers.

The field dependence of the impact ionization coefficient, at least over the limited fields where avalanche is observed, is usually modeled by experimentalists by using

$$\alpha = A\exp\left[-\left(\frac{B}{F}\right)^n\right], \tag{2.78}$$

where F is the field A, B, n are constants that depend on the semiconductor material properties such as the E-$\boldsymbol{k}$ electronic structure, phonon energies and spectra, scattering mechanisms and so on. The constant B has been semiquantitatively argued to depend on E_g/τ where E_g is the bandgap and $1/\tau$ is the phonon scattering rate; higher bandgap semiconductors tend to have steeper slopes on log α versus $1/F$ plots and the log α versus $1/F$ curve tends to shift to higher fields. Typically, it has been found that $n\approx1$ at low fields and $n\approx2$ for high fields. Figure 2.17 shows experimental data for a variety of materials over a wide range of electric fields.

Avalanche multiplication in crystalline semiconductors has been well studied over several decades and is now well developed and used in various commercial applications, such as avalanche photodiodes in optical communications, high sensitivity radiation detectors, and x-ray and nuclear detectors [2.69, 70]. In the case of amorphous semiconductors, the present work points to only a-Se exhibiting avalanche multiplication in the bulk for reasons explained in reference [2.71], with reports dating back to the 1980s [2.61, 62, 72]. Avalanche in a-Se has also been commercialized in ultrasensitive video tubes [2.73], with potential application in medical imaging [2.74].

The origin of (2.78) in its simple $n=1$ form lies in *Shockley*'s [2.75] *lucky electron* model. When a carrier moves a distance z down-stream (along the field) without being scattered, it gains an energy eFz. An unlucky carrier is scattered so frequently that its eFz never reaches the threshold ionization energy E_I for impact ionization. On the other hand, a *lucky electron* is a ballistic electron that avoids scattering for a substantial distance, and hence is able to build its eFz to reach E_I and thereby cause impact ionization. If λ is the mean free path of collisions, then Shockley's model gives

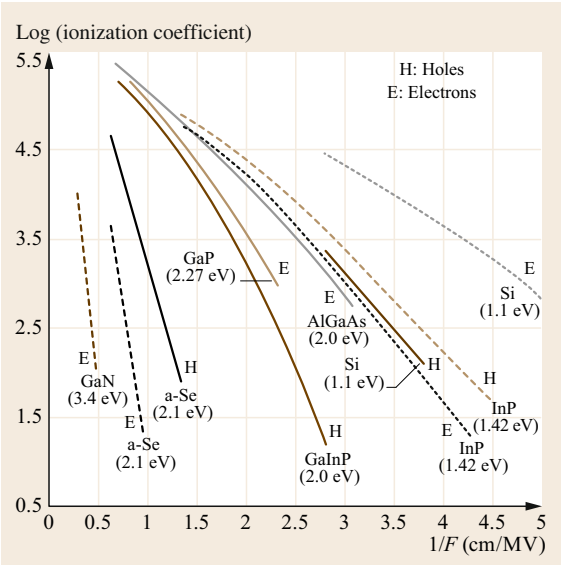


Fig. 2.17 Semilogarithmic plot of the dependence of the impact ionization coefficient on the reciprocal field for various crystalline and amorphous semiconductors for comparison; H indicates holes and E electrons. The y-axis is a base-10 logarithm of α in which α is in 1/cm. Data for a-Se electrons and holes from [2.60–62]; Data for crystalline semiconductors are for GaP [2.63]; GaInP ($Ga_{0.52}In_{0.48}P$) [2.64]; $Al_{0.60}Ga_{0.40}As$ [2.65]; InP [2.66]; Si, [2.67]; GaN (calculation only) [2.68]

$$\alpha = \frac{eF}{E_I}\exp\left(-\frac{E_I}{e\lambda F}\right). \tag{2.79}$$

The main problem with the Shockley model is that there are just not enough ballistic electrons to cause sufficient impact ionizations to explain the experiments. A better model was developed by *Baraff* [2.76] who numerically solved the Boltzmann transport equation for a simple parabolic band and energy independent mean free path to provide a relationship between α and F in terms of four parameters, that is, threshold energy for impact ionization, mean free path associated with ionizations, optical phonon energy, and mean free path for optical phonon scattering. Baraff's theory served experimental-

ists quite well in terms of comparing their results even though the model was not intuitive and was limited in terms of its assumptions and applicability to real semiconductors.

Impact ionization theory in crystalline solids only reached an acceptable level of confidence and understanding in the 1980s and 1990s with the development of the lucky-drift model by *Ridley* [2.77] and its extension by *Burt* [2.78], and *Mackenzie* and *Burt* [2.79]. The latter major advancement in the theory appeared as the *lucky drift* (LD) *model*, and it was based on the realization that at high fields, hot electrons do not relax momentum and energy at the same rates. Momentum relaxation rate is much faster than the energy relaxation rate. An electron can drift, being scattered by phonons, and have its momentum relaxed, which controls the drift velocity, but it can still gain energy during this drift. Stated differently, the mean free path λ_E for energy relaxation is much longer than the mean free path λ for momentum relaxation.

In the *Mackenzie* and *Burt* [2.79] version of the LD model, the probability $P(E)$ that a carrier attains an energy E is given by,

$$P(E) = \exp\left(-\int_0^E \frac{dE'}{eF\lambda(E')}\right) + \int_0^E \frac{dE_1}{eF\lambda(E_1)} \times \exp\left(-\int_0^{E_1} \frac{dE'}{eF\lambda(E')}\right) \times \exp\left(-\int_{E_1}^{E} \frac{dE'}{eF\lambda_E(E')}\right), \tag{2.80}$$

where as mentioned above λ is the mean free path associated with momentum relaxing collisions and λ_E is the mean energy relaxation length associated with the energy relaxing collisions. The first term is the Shockley lucky electron probability, i. e., the electron moves ballistically to energy E. The second term is the lucky drift probability term which is composed of the following: the electron first moves ballistically to some intermediate energy E_1 $(0 < E_1 < E)$ from where it begins its lucky drift to energy E_I; hence the integration over all possible E_1. The impact ionization coefficient can then readily be evaluated from

$$\alpha = \frac{eFP(E_I)}{\int_0^{E_I} P(E)dE}\,. \tag{2.81}$$

The model above is based on a hard threshold ionization energy E_I, that is, when a carrier attains the energy E_I, ionization ensues. The model has been further refined by the inclusion of soft threshold energies which represent the fact the ionization does not occur immediately when the carrier attains the energy E_I, and the carrier drifts further to gain more energy than E_I before impact ionization [2.80–84].

Assuming λ and λ_E are energy independent, which would be the case for a single parabolic band in the crystalline state, (2.80) and (2.81) can be solved analytically to obtain

$$\alpha = \frac{1}{\lambda} \times \frac{\frac{\lambda}{\lambda_E}\exp\left(\frac{-E_I}{eF\lambda_E}\right) + \left(\frac{\lambda}{\lambda_E}\right)^2 \exp\left(\frac{-E_I}{eF\lambda}\right)}{1 - \exp\left(\frac{-E_I}{eF\lambda_E}\right) - \left(\frac{\lambda}{\lambda_E}\right)^2 \left[1 - \exp\left(\frac{-E_I}{eF\lambda}\right)\right]}\,. \tag{2.82}$$

For $\lambda_E > \lambda$, and in the *low field region*, where typically $(\alpha\lambda) < 10^{-1}$, or $x = E_I/eF\lambda > 10$, (2.82) leads to a simple expression for α,

$$\alpha = \left(\frac{1}{\lambda_E}\right)\exp\left(-\frac{E_I}{eF\lambda_E}\right). \tag{2.83}$$

For crystalline semiconductors, one typically also assumes that λ_E depends on the field F, λ and the optical phonon energy $\hbar\omega$ as

$$\lambda_E = \frac{eF\lambda^2}{2\hbar\omega}\coth\left(\frac{\hbar\omega}{2kT}\right). \tag{2.84}$$

As the field increases, λ_E eventually exceeds λ, and allows lucky drift to operate and the LD carriers to reach the ionization energy.

It is worth noting that the model of lucky drift is successfully used not only for crystalline semiconductors but to amorphous semiconductors [2.83].

2.12 Two-Dimensional Electron Gas

Heterostructures offer the ability to spatially engineer the potential in which carriers move. In such structures having layers deposited in the z-direction, when the width of a region with confining potential $t_z < \lambda_{dB}$, the de Broglie electron wavelength, electron states become stationary states in that direction, retaining Bloch wave character in the other two directions (i. e., x- and y-directions), and is hence termed a *2-D electron gas* (2DEG). These structures are notable for their extremely high carrier mobility.

High mobility structures are formed by selectively doping the wide bandgap material behind an initially undoped spacer region of width d as shown in Fig. 2.18a. Ionization and charge transfer leads to carrier build-up in the low potential region of narrow bandgap material adjacent to the hetero-interface. The equilibrium band bending (i. e., through Poisson's equation) in the well region, as shown in Fig. 2.18b, results from the equalization of ionized donor concentration in the wide bandgap material and 2DEG concentration adjacent to the heterointerface. When the associated interfacial field is sufficiently strong, carriers are confined within λ_{dB} and electron states are quantized into subbands (i. e., E_0 and E_1), as shown in Fig. 2.18b.

Figure 2.19 shows the contributions of component scattering mechanisms to the low temperature mobility of a 2DEG formed at a $Ga_{0.70}Al_{0.3}As$-GaAs heterointerface, as a function of the electron gas density. As for bulk samples, the most important mechanism limiting the low temperature mobility is ionized impurity scat-

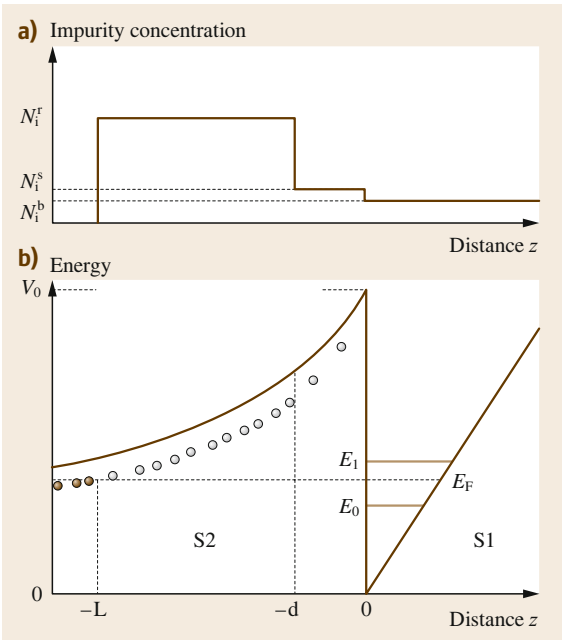


Fig. 2.18 **(a)** Doping profile for a selectively doped 2DEG heterostructure. The S1 is a narrow bandgap material adjacent to the heterointerface. The S2 is selectively doped wide-bandgap material including an initially undoped spacer region of width d. **(b)** Conduction energy band structure for a selectively doped 2DEG heterostructure. The equilibrium band bending (through Poisson's equation) in the well region results from the equalization of the ionized donor concentration in the wide-bandgap material and the 2DEG concentration adjacent to the heterointerface. When the associated interfacial field is sufficiently strong, carriers are confined within λ_{dB} and electron states are quantized into the sub-bands E_0 and E_1. (After [2.84])

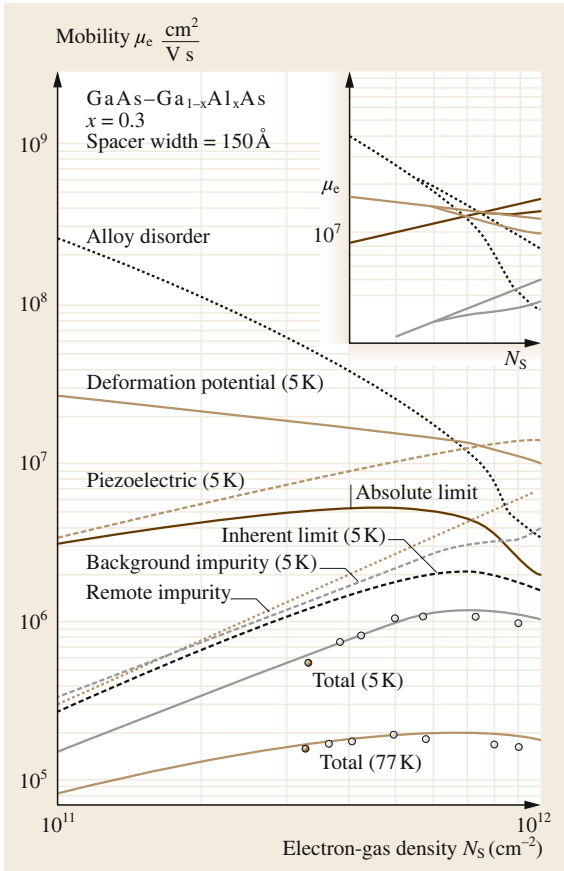


Fig. 2.19 Dependence of 2DEG electron mobility on carrier concentration for a $Ga_{0.70}Al_{0.3}As$-GaAs heterostructure. (After [2.84])

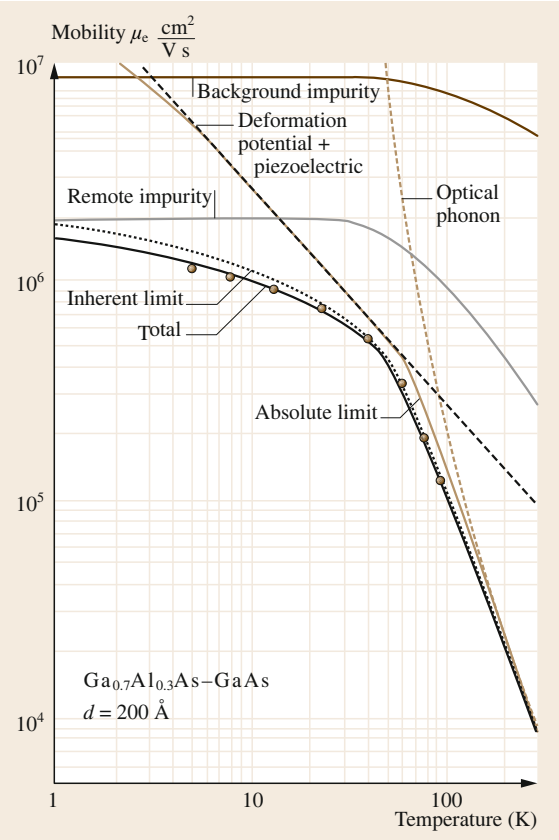


Fig. 2.20 Dependence of 2DEG electron mobility on temperature for a $Ga_{0.70}Al_{0.3}As$-GaAs heterostructure. (After [2.84])

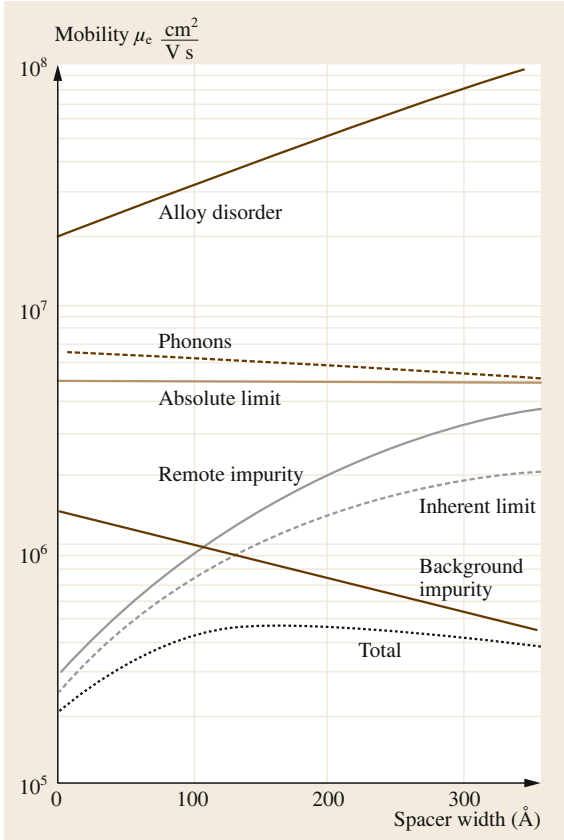


Fig. 2.21 Dependence of 2DEG electron mobility on spacer width for a $Ga_{0.70}Al_{0.3}As$-GaAs heterostructure. (After [2.84])

tering, except at high electron densities, where so-called *alloy disorder scattering* is significant. Ionized impurity scattering may be further broken down into scattering from ionized impurities that are with the GaAs quantum well, known as *background impurities*, those beyond the spacer region, termed *remote impurity scattering*. For high purity growth, the unintentional background impurity concentration can be kept to very low limits and impurity scattering based mobility values are then dictated by remote impurity scattering. Since carriers in the well are only weakly scattered by the tail field of these remote Coulomb centers, the mobility of such 2DEG systems can be orders of magnitude higher than bulk samples. The temperature dependence of the electron mobility for such a system is shown in Fig. 2.20. Notice how, similar to bulk sample, increasing temperature increases the phonon population such that for the example shown, above about 100 K, polar optical phonon scattering controls the mobility.

Figure 2.21 shows the dependence of the 2DEG mobility on the spacer width. Two competing factors are active – for narrow spacer widths, the transfer efficiency of carriers to the GaAs well is high and so a lower remote doping concentration is sufficient to provide for a given constant 2DEG concentration, but since the Coulomb scatters are so close to the 2DEG, they scatter very efficiently and limit the mobility. On the other hand, for large spacer widths, carrier transfer efficiency is quite poor requiring higher remote doping to supply the given 2DEG concentration; however, being relatively far away each of the scattering centers are less effective at lowering the mobility, but given their high concentration, the net effect is still a decrease in the mobility at large spacer widths as seen in the figure.

2.13 One-Dimensional Conductance

In the case where carriers are confined within regions of width L_x, $L_y < \lambda_{dB}$ in two directions x and y, respectively, the electron energy in those two directions become quantized with quantum numbers n_x and n_y, respectively. In the third direction z, electrons travel as Bloch waves with energy that may be approximated by $\hbar^2 k_z^2/2m^*$ giving an expression for the total energy E of the so-called *1-D electron system* or *quantum wire* as

$$E(n_x, n_y, n_z) = \frac{\hbar^2 \pi^2}{2m^*}\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2}\right) + \frac{\hbar^2 k_z^2}{2m^*} . \quad (2.85)$$

The associated electron wavefunctions are

$$\Psi(x, y, z) = \frac{1}{2\sqrt{L_x L_y L_z}} \sin\left(\frac{n_x \pi x}{L_x}\right) \times \sin\left(\frac{n_y \pi y}{L_y}\right) e^{ik_z z} . \quad (2.86)$$

Using these equations, one can readily derive an expression for the density of states per unit energy range

$$\mathrm{DOS} = 2 \times 2\left(\frac{L_z}{2\pi}\right)\left(\nabla_{k_z} E\right)^{-1} = \frac{2L_z}{h}\sqrt{\frac{m^*}{2(E - E_{n_x,n_y})}} . \quad (2.87)$$

In order to evaluate the conductance of this quantum wire, consider the influence of a weak applied potential V. Similar to the case for bulk transport the applied field displaces the *Fermi surface* and results in a change in the electron wave-vector from $\boldsymbol{k}_0$ (i. e., with no applied potential) to $\boldsymbol{k}_V$ (i. e., when the potential is applied). When V is small compared with the electron energy

$$k_0 = \sqrt{\frac{2m^*(E - E_{n_x,n_y})}{\hbar^2}} , \quad (2.88)$$

$$k_V = k_0\sqrt{1 + \frac{eV}{E - E_{n_x,n_y}}} \approx k_0\left(1 + \frac{1}{2}\frac{eV}{E - E_{n_x,n_y}}\right) . \quad (2.89)$$

This leads to establishing a current density J in the wire

$$J = \frac{2e^2\,(\mathrm{DOS})\sqrt{(E_F - E_{n_x,n_y})}}{\sqrt{2m^*}} . \quad (2.90)$$

Which may be simplified to the following expressions for J and the current flowing in the wire for a given quantum state $E\{n_x, n_y\}$, I

$$J_{n_x,n_y} = \frac{2e^2 V L_z}{h} \quad \text{and} \quad I_{n_x,n_y} = \frac{2e^2 V}{h} . \quad (2.91)$$

The expression for the conductance through one channel corresponding to a given quantum state $\{n_x, n_y\}$ is then given by

$$G_{n_x,n_y} = \frac{I_{n_x,n_y}}{V} = \frac{2e^2}{h} . \quad (2.92)$$

Notice how the conductance is quantized in units of e^2/h with each populated channel contributing equally to the conductance – moreover, this is a fundamental result, being independent of the material considered. In practice, deviations from this equation can occur (although generally less than 1%) owing to the finiteness of real nanowires and impurities in or near the channel, influencing the conductivity and even resulting in weak localization. Generally, unlike both bulk and 2DEG systems, ionized impurity scattering is suppressed in nanowires. The main reason for this is that an incident electron in a quantum state $\{n_x, n_y\}$ traveling along the wire with wave-vector $\boldsymbol{k}_z\{n_x, n_y\}$, can not be elastically scattered into any states except those in a small region of $\boldsymbol{k}$-space in the vicinity of $-\boldsymbol{k}_z\{n_x, n_y\}$. Such a scattering event involves a large change in momentum of $\approx 2\boldsymbol{k}_z\{n_x, n_y\}$ and thus, the probability of such events is very small. As a result, the mean free path and mobility of carriers in such quantum wires are substantially increased.

The nature of carrier transport in quantum wires depends on the wire dimensions (i. e., length L_{Wire} and diameter d_{Wire}) as compared with the carrier mean free path, l_{Carrier}. When $l_{\mathrm{Carrier}} \gg L_{\mathrm{Wire}}$, d_{Wire} the only potential seen by the carriers is that associated with the wire walls, and carriers exhibit wavelike behavior, being guided through the wire as if it were a waveguide without any internal scattering. Conversely, if $d_{\mathrm{Wire}} \ll \lambda_{\mathrm{DeBroglie}}$, only a few energy states in the wire are active, and in the limit of an extremely small waveguide, only one state or channel is active, analogous to a single mode waveguide cavity – this case is termed *quantum ballistic transport*. In the limit, $l_{\mathrm{Carrier}} \ll L_{\mathrm{Wire}}$, d_{Wire}, scattering dominates transport throughout the wire – with numerous scattering events occurring before a carrier can traverse the wire or move far along its length. In such a case the transport is said to be *diffusive*.

As discussed previously, ionized impurity and lattice scattering contribute to l_{Carrier}, with l_{Carrier} decreasing with increasing temperature due to phonon scattering. For strong impurity scattering, this may not occur until relatively high temperatures. In the intermediate case of $L_{\text{Wire}} \gg l_{\text{Carrier}} \gg d_{\text{Wire}}$ and where $d_{\text{Wire}} \ll \lambda_{\text{DeBroglie}}$ scattering is termed *mixed mode* and is often called *quasi-ballistic*.

2.14 The Quantum Hall Effect

The observation of, and first explanation for the Hall Effect in a 2DEG by von *Klitzing* et al. [2.85], won them a Nobel Prize. As shown in Fig. 2.22 the Hall resistivity exhibits plateaus for integer values of h/e^2, independent of any material dependent parameters. This discovery was later shown to be correct to a precision of at least one part in 10^7, enabling extremely accurate determinations of the fine structure constant (i. e., $\alpha = (\mu_0 c e^2/2h) \approx 1/137$) and a fundamental resistance standard to be established.

For the six point Hall geometry used (as shown in the insert to Fig. 2.18), one can define the *Hall resistivity* $\rho_{xy} = g_1 V_{\text{H},y}/I_x$ and *longitudinal resistivity* $\rho_{xx} = g_2 V_{\text{L},x}/I_x$ where g_1 and g_2 are geometric constants related to the sample geometry. These resistivities are related to corresponding conductivities through the conductivity and resistivity tensors

$$\begin{aligned}
\rho_{xy} &= \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2} \,, \\
\rho_{xx} &= \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2} \,, \\
\rho_{yx} &= -\rho_{xy} \,, \\
\rho_{yy} &= \rho_{xx} \,.
\end{aligned} \tag{2.93}$$

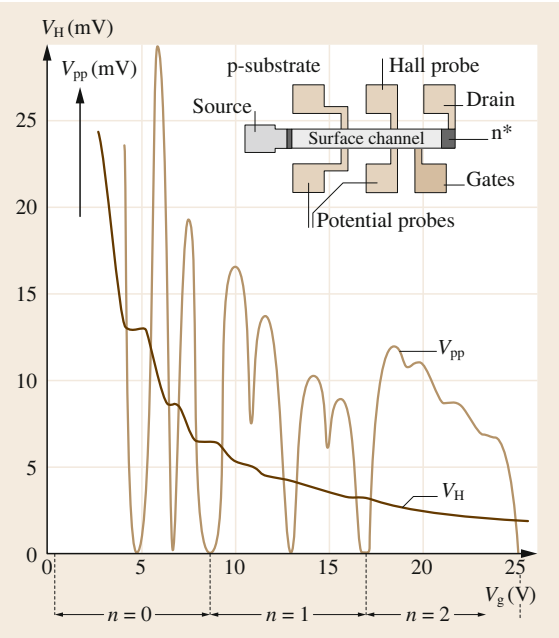


Fig. 2.22 Hall voltage V_{H} and voltage drop across electrodes V_{PP} as a function of gate voltage V_{g} at 1.5 K, when $B = 18$ T. Source-drain current is 1 μA. *Insert* shows plan view of a device with length 400 μm, width 50 μm and an interprobe separation of 130 μm. (After [2.85])

Starting from a classical equation of motion for electrons in an electric field E_x, magnetic field B_z, and defining the *cyclotron frequency* $\omega_{\text{c}} = eB_z/m^*$, the velocities perpendicular to the applied magnetic field can be deduced as

$$v_x = \frac{E_x}{B_x} \sin \omega_{\text{c}} t \quad \text{and} \quad v_y = \frac{E_x}{B_x} (\cos \omega_{\text{c}} t - 1) \tag{2.94}$$

with the time averaged velocities

$$\langle v_x \rangle = \frac{E_x}{B_x} \quad \text{and} \quad \langle v_y \rangle = \frac{E_x}{B_x} \tag{2.95}$$

leading to Hall and longitudinal resistivities of

$$\sigma_{xx} = \sigma_{yy} = 0 \quad \text{and} \quad \sigma_{xy} = -\sigma_{yx} = \frac{n_{\text{s}} e}{B_z} \,, \tag{2.96}$$

where n_{s} is the areal electron concentration. Below, a quantum approach is used to establish a relationship for the electron concentration n_{s}. Note that the motion of the electrons in the crossed fields are quantized with allowed levels termed *Landau levels* E_n

$$E_n = \left(n + \frac{1}{2}\right) \hbar \omega_{\text{c}} + g^* \mu_{\text{B}} B_z + \varepsilon_z \,, \tag{2.97}$$

where n is the quantum number describing the particular Landau Level, g^* is the *Landé factor*, μ_{B} is the Bohr magneton and $g^* \mu_{\text{B}} B_z$ is *the spin magnetic energy*, and E_z is the energy associated with the z-motion

of the carriers. xy-plane carrier motion is characterized by the cyclotron energy term E_{xy},

$$E_{xy} = \frac{\hbar^2 k_{xy}^2}{2m_e^*} = \left(n + \frac{1}{2}\right)\hbar\omega_c \,. \tag{2.98}$$

Following this description and noting that the motion of electrons in the xy-plane may be expressed in terms of wavefunctions of the *harmonic oscillator* using the *Landau gauge* of the *vector potential* $[0, xB_z, 0]$, we may write the density of states per unit area, $\mathrm{DOS_A}$ as

$$\mathrm{DOS_A} = \frac{m_e^* \omega_c L_x L_y}{2\pi\hbar} \,. \tag{2.99}$$

Since the degeneracy of each Landau level is one (i. e., since they are single spin states), this enables one to find n_s assuming Landau state filling up to the p^{th} level (for integer p)

$$n_s = \frac{m_e^* \omega_c p}{2\pi\hbar} \,. \tag{2.100}$$

Using the definition of the cyclotron frequency gives the final form

$$n_s = \frac{peB_z}{\hbar} \tag{2.101}$$

which may be used to rewrite the previous expressions for Hall and longitudinal resistivity

$$\rho_{xx} = \rho_{yy} = 0 \quad \text{and} \quad \rho_{xy} = -\rho_{yx} = \frac{h}{pe^2} \,. \tag{2.102}$$

Equation (2.102) shows that Hall resistivity is quantized in units of h/pe^2 whenever the Fermi energy lies between filled Landau levels. Consistent with observation, the result is independent of the semiconductor being studied. Although this model provides an excellent basis for understanding experiments, understanding the details of the results (i. e., in particular the existence of a finite width for the Hall effect plateaus and zero longitudinal resistance dips) requires a more complete treatment involving so-called localized states.

Acknowledgments. The authors thank Dr. Robert Johanson for helpful discussions on the subject.

References

2.1 P.L. Rossiter: *The Electrical Resisitivity of Metals and Alloys* (Cambridge Univ. Press, Cambridge 1987)
2.2 J.S. Dugdale: *The Electrical Properties of Metals and Alloys* (Arnold, London 1977)
2.3 B.R. Nag: *Theory of Electrical Transport in Semiconductors* (Pergamon, Oxford 1972)
2.4 F.J. Blatt: *Physics of Electronic Conduction in Solids* (McGraw-Hill, New York 1968)
2.5 S.O. Kasap: *Principles of Electronic Materials and Devices*, 4th edn. (McGraw-Hill, Dubuque 2017)
2.6 W.M. Haynes (Ed.): *CRC Handbook of Chemistry and Physics*, 97th edn. (CRC Press, Boca Raton 2016), Chap. 12
2.7 Kaye & Laby Tables of Physical and Chemical Constants (Section 2.6) (National Physical Laboratory, Teddington 2016) provided by the website of the National Physical Laboratory
2.8 L. Nordheim: Ann. Phys. **9**, 664 (1931)
2.9 J.K. Stanley: *Electrical and Magnetic Properties of Metals* (American Society for Metals, Metals Park 1963)
2.10 M. Hansen, K. Anderko: *Constitution of Binary Alloys*, 2nd edn. (McGraw-Hill, New York 1985)
2.11 H.E. Ruda: J. Appl. Phys. **59**, 1220 (1986)
2.12 M. Lundstrom: *Fundamentals of Carrier Transport* (Cambridge Univ. Press, Cambridge 2000)
2.13 R.H. Bube: *Electronic Properties of Crystalline Solids* (Academic, New York 1974) p. 211
2.14 S. Riedel, J. Röber, T. Geßner: Microelectron. Eng. **33**, 165 (1997)
2.15 A.F. Mayadas, M. Shatzkes: Phys. Rev. B **1**, 1382 (1970)
2.16 C.R. Tellier, C.R. Pichard, A.J. Tosser: J. Phys. F **9**, 2377 (1979)
2.17 K. Fuchs: Proc. Camb. Philos. Soc. **34**, 100 (1938)
2.18 E.H. Sondheimer: Adv. Phys. **1**, 1 (1952)
2.19 J.S. Chawla, D. Gall: Appl. Phys. Letts. **94**, 252101 (2009)
2.20 T. Sun, B. Yao, A.P. Warren, V. Kumar, S. Roberts, K. Barmak, K.R. Coffey: J. Vac. Sci. Technol. A **26**, 605 (2008)
2.21 T. Sun, B. Yao, A.P. Warren, K. Barmak, M.F. Toney, R.E. Peale, K.R. Coffey: Phys. Rev. B **81**, 155454 (2010)
2.22 J.S. Chawla, X.Y. Zhang, D. Gall: J. Appl. Phys. **110**, 043714 (2011)
2.23 J.W. Lim, M. Isshiki: J. Appl. Phys. **99**, 094909 (2006)
2.24 H.-D. Liu, Y.-P. Zhao, G. Ramanath, S.P. Murarka, G.-C. Wang: Thin Solid Films **384**, 151 (2001)
2.25 J.-W. Lim, K. Mimura, M. Isshiki: Appl. Surf. Sci. **217**, 95 (2003)
2.26 R. Suri, A.P. Thakoor, K.L. Chopra: J. Appl. Phys. **46**, 2574 (1975)
2.27 R.H. Cornely, T.A. Ali: J. Appl. Phys. **49**, 4094 (1978)
2.28 J.S. Ahn, K.H. Kim, T.W. Noh, D.H. Riu, K.H. Boo, H.E. Kim: Phys. Rev. B **52**, 15244 (1995)
2.29 R.J. Gehr, G.L. Fisher, R.W. Boyd: J. Opt. Soc. Am. B **14**, 2310 (1997)
2.30 D.E. Aspnes, J.B. Theeten, F. Hottier: Phys. Rev. B **20**, 3292 (1979)
2.31 Z. Yin, F.W. Smith: Phys. Rev. B **42**, 3666 (1990)

2.32 M.F. MacMillan, R.P. Devaty, W.J. Choyke, D.R. Goldstein, J.E. Spanier, A.D. Kurtz: J. Appl. Phys. **80**, 2412 (1996)
2.33 C. Ganter, W. Schirmacher: Phys. Status Solidi B **218**, 71 (2000)
2.34 R. Stognienko, Th. Henning, V. Ossenkopf: Astron. Astrophys. **296**, 797 (1995)
2.35 A.G. Rojo, H.E. Roman: Phys. Rev. B **37**, 3696 (1988)
2.36 J.A. Reynolds, J.M. Hough: Proc. Phys. Soc. **70**, 769 (1957)
2.37 R. Clausius: *Die Mechanische Wärmetheorie*, Vol. 2 (Vieweg, Braunschweig 1879)
2.38 L. Lorenz: Ann. Phys. Lpz. **11**, 70 (1880)
2.39 O.F. Mosotti: Mem. Math. Fisica Modena II **24**, 49 (1850)
2.40 V.I. Odelevskii: Zh. Tekh. Fiz. **6**, 667 (1950)
2.41 Lord Rayleigh: Philos. Mag. **34**, 481 (1892)
2.42 K.W. Wagner: Arch. Electrochem. **2**, 371 (1914)
2.43 D.A.G. Bruggeman: Ann. Phys. Lpz. **24**, 636 (1935)
2.44 C.J.F. Bottcher: Rec. Trav. Chim. Pays-Bas **64**, 47 (1945)
2.45 H. Fricke: Phys. Rev. **24**, 575 (1924)
2.46 D. Polder, J.M. Van Santen: Physica **12**, 257 (1946)
2.47 W. Niesel: Ann. Phys. Lpz. **10**, 336 (1952)
2.48 J.A. Stratton: *Electromagnetic Theory* (McGraw-Hill, New York 1941)
2.49 O. Wiener: Abh. Sachs. Ges. Akad. Wiss. Math. Phys. **32**, 509 (1912)
2.50 L. Silberstein: Ann. Phys. Lpz. **56**, 661 (1895)
2.51 R.W. Sillars: J. Inst. Elect. Eng. **80**, 378 (1937)
2.52 F. Ollendorf: Arch. Electrochem. **25**, 436 (1931)
2.53 J.C.M. Maxwell-Garnett: Phil. Trans. R. Soc. Lond. **203**, 385 (1904)
2.54 H. Looyenga: Physica **31**, 401 (1965)
2.55 J. Monecke: J. Phys. Condens. Mat. **6**, 907 (1994)
2.56 C.F. Bohren, D.R. Huffman: *Absorption and Scattering of Light by Small Particles* (Wiley, New York 1983)
2.57 D.J. Bergman: Phys. Rep. **43**, 377 (1978)
2.58 W. Theiss: Adv. Solid State Phys. **33**, 149 (2007)
2.59 P.Y. Yu, M. Cardona: *Fundamentals of Semiconductors* (Springer, Berlin, Heidelberg 1996)
2.60 K. Tsuji, Y. Takasaki, T. Hirai, K. Taketoshi: J. Non-Cryst. Solids **14**, 94 (1989)
2.61 G. Juska, K. Arlauskas: Phys. Status Solidi **77**, 387 (1983)
2.62 G. Juska, K. Arlauskas: Phys. Status Solidi **59**, 389 (1980)
2.63 R.A. Logan, H.G. White: J. Appl. Phys. **36**, 3945 (1965)
2.64 R. Ghin, J.P.R. David, S.A. Plimmer, M. Hopkinson, G.J. Rees, D.C. Herbert, D.R. Wight: IEEE Trans. Electron Dev. **ED45**, 2096 (1998)
2.65 S.A. Plimmer, J.P.R. David, R. Grey, G.J. Rees: IEEE Trans. Electron Dev. **ED47**, 21089 (2000)
2.66 L.W. Cook, G.E. Bulman, G.E. Stillma: Appl. Phys. Lett. **40**, 589 (1982)
2.67 C.A. Lee, R.A. Logan, R.L. Batdorf, J.J. Kleimack, W. Wiegmann: Phys. Rev. **134**, B766 (1964)
2.68 C. Bulutay: Semicond. Sci. Technol. **17**, L59 (2002)
2.69 B. Nabet: *Photodetectors: Materials, Devices and Applications* (Elsevier, Amsterdam 2015)
2.70 S.N. Ahmed: *Physics and Engineering of Radiation Detection*, 2nd edn. (Elsevier, Amsterdam 2014)
2.71 A. Reznik, S.D. Baranovskii, O. Rubel, G. Juska, S.O. Kasap, Y. Ohkawa, K. Tanioka, J.A. Rowlands: J. Appl. Phys. **102**, 53711 (2007)
2.72 M. Akiyama, M. Hanada, H. Takao, K. Sawada, M. Ishida: Jpn. J. Appl. Phys **41**, 2552 (2002)
2.73 M. Kubota, T. Kato, S. Suzuki, H. Maruyama, K. Shidara, K. Tanioka, K. Sameshima, T. Makishima, K. Tsuji, T. Hirai, T. Yoshida: IEEE Trans. Broadcast. **42**, 251 (1996)
2.74 O. Bubon, G. i DeCrescenzo, W. Zhao, Y. Ohkawa, K. Miyakawa, T. Matsubara, K. Kikuchi, K. Tanioka, M. Kubota, J.A. Rowlands, A. Reznik: Curr. Appl. Phys. **12**, 983 (2012)
2.75 W. Shockley: Solid State Electron. **2**, 35 (1961)
2.76 G.A. Baraff: Phys. Rev. **128**, 2507 (1962)
2.77 B.K. Ridley: J. Phys. C **16**, 4733 (1983)
2.78 M.G. Burt: J. Phys. C **18**, L477 (1985)
2.79 S. MacKenzie, M.G. Burt: Semicond. Sci. Technol. **2**, 275 (1987)
2.80 B.K. Ridley: Semicond. Sci. Technol. **2**, 116 (1987)
2.81 J.S. Marsland: Solid State Electron. **30**, 125 (1987)
2.82 J.S. Marsland: Semicond. Sci. Technol. **5**, 177 (1990)
2.83 S.O. Kasap, J.A. Rowlands, S.D. Baranovskii, K. Tanioka: J. Appl. Phys. **96**, 2037 (2004)
2.84 W. Walukiewicz, H.E. Ruda, J. Lagowski, H.C. Gatos: Phys. Rev. B **30**, 4571 (1984)
2.85 K.V. Klitzing, G. Dorda, M. Pepper: Phys. Rev. Lett. **45**, 494 (1980)

3. Optical Properties of Electronic Materials: Fundamentals and Characterization

Jan Mistrik, Safa Kasap, Harry E. Ruda, Cyril Koughia, Jai Singh

Light interacts with materials in a variety of ways; this chapter focuses on refraction and absorption. Refraction is characterized by a material's refractive index. We discuss some of the most useful models for the frequency dependence of the refractive index, such as those due to Cauchy, Sellmeier, Gladstone–Dale, and Wemple–DiDominico. Examples are given of the applicability of the models to actual materials. We present various mechanisms of light absorption, including absorption by free carriers, phonons, excitons and impurities. Special attention is paid to fundamental and excitonic absorption in disordered semiconductors and to absorption by rare earth, trivalent ions due to their importance in modern photonics. We also discuss the effect of an external electric field on absorption, and the Faraday effect. Practical techniques for determining the optical parameters of thin films are outlined. Finally, we present a short technical classification of optical glasses and materials.

3.1 **Optical Constants** 47
3.1.1 Refractive Index and Extinction Coefficient 48
3.1.2 Kramers–Kronig Relations 49
3.2 **Refractive Index** 50
3.2.1 Cauchy Dispersion Equation 50
3.2.2 Sellmeier Dispersion Equation 50
3.2.3 Gladstone–Dale Formula 51
3.2.4 Wemple–DiDominico Dispersion Relation 52
3.2.5 Group Index (N_g) 53
3.3 **Optical Absorption** 53
3.3.1 Lattice or Reststrahlen Absorption and Infrared Reflection 53
3.3.2 Free-Carrier Absorption (FCA) 54
3.3.3 Band-to-Band or Fundamental Absorption 56
3.3.4 Exciton Absorption 62
3.3.5 Impurity Absorption 65
3.3.6 Effects of External Fields 68
3.4 **Optical Characterization** 69
3.4.1 Bulk Samples 71
3.4.2 Thin Film Optics 75
3.5 **Optical Materials** 79
3.5.1 Abbe Number or Constringence 79
3.5.2 Optical Materials 79
3.5.3 Optical Glasses 80
References 82

3.1 Optical Constants

The changes that light undergoes upon interacting with a particular substance are known as the optical properties of that substance. These optical properties are influenced by the macroscopic and microscopic properties of the substance, such as the nature of its surface and its electronic structure. Since it is usually far easier to detect the way a substance modifies light than to investigate its macroscopic and microscopic properties directly, the optical properties of a substance are often used to probe other properties of the material. There are many optical properties, including the most well known: reflection, refraction, transmission and absorption. Many of these optical properties are associated with important optical constants, such as the refractive index and the extinction coefficient. In this section we review these optical constants, such as the refractive index and the extinction coefficient. Books by *Adachi* [3.1], *Fox* [3.2] and *Simmons* and *Potter* [3.3] are highly recommended. In addition, *Adachi* also discusses the optical properties of III–V compounds in this handbook in Chap. 30.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_3

3.1.1 Refractive Index and Extinction Coefficient

The refractive index n of an optical or dielectric medium is the ratio of the velocity of light c in a vacuum to its velocity v in the medium; $n = c/v$. Using this and Maxwell's equations, one obtains the well-known formula for the refractive index of a substance as $n = \sqrt{\varepsilon_r \mu_r}$, where ε_r is the relative permittivity (dielectric constant) and μ_r is the relative magnetic permeability. As $\mu_r = 1$ for nonmagnetic substances, one gets $n = \sqrt{\varepsilon_r}$, which is very useful for relating the dielectric properties of a material to its optical properties at any particular frequency of interest. Since ε_r depends on the wavelength of the light, the refractive index depends on it too, which is called dispersion. In addition to dispersion, an electromagnetic wave propagating through a lossy medium (one that absorbs or scatters radiation passing through it) experiences attenuation, which means that it loses its energy due to various loss mechanisms such as the generation of phonons (lattice waves), photogeneration, free-carrier absorption and scattering. In such materials, the refractive index is a complex function of the frequency of the light wave.

The complex refractive index N, with real part n and imaginary part K (called the extinction coefficient), is related to the complex relative permittivity, $\varepsilon_r = \varepsilon_r' - i\varepsilon_r''$ by

$$N = n - iK = \sqrt{\varepsilon_r} = \sqrt{\varepsilon_r' - i\varepsilon_r''} \tag{3.1}$$

where ε_r' and ε_r'' are, respectively, the real and imaginary parts of ε_r. Equation (3.1) gives

$$n^2 - K^2 = \varepsilon_r' \quad \text{and} \quad 2nK = \varepsilon_r'' \,. \tag{3.2}$$

The optical constants n and K can be determined by measuring the reflectance from the surface of a material as a function of polarization and the angle of incidence. For normal incidence, the reflection coefficient r is obtained as

$$r = \frac{1-N}{1+N} = \frac{1-n+iK}{1+n-iK} \,. \tag{3.3}$$

The reflectance R is then defined by

$$R = |r|^2 = \left|\frac{1-n+iK}{1+n-iK}\right|^2 = \frac{(1-n)^2+K^2}{(1+n)^2+K^2} \,. \tag{3.4}$$

Notice that whenever K is large, for example over a range of wavelengths, the absorption is strong, and the reflectance is almost unity. The light is then reflected, and any light in the medium is highly attenuated.

Optical properties of materials are typically presented by showing either the frequency dependence (dispersion relation) of n and K or ε_r' and ε_r''. An intuitive guide to explaining dispersion in insulators is based on a single oscillator model in which the electric field in the light induces forced dipole oscillations in the material (displaces the electron shells to oscillate about the nucleus) with a single resonant frequency ω_0 [3.1]. The frequency dependences of ε_r' and ε_r'' are then obtained as

$$\varepsilon_r' = 1 + \frac{N_{at}}{\varepsilon_0}\alpha_e' \quad \text{and} \quad \varepsilon_r'' = 1 + \frac{N_{at}}{\varepsilon_0}\alpha_e'' \,, \tag{3.5}$$

where N_{at} is the number of atoms per unit volume, ε_0 is the vacuum permittivity, and α_e' and α_e'' are the real and imaginary parts of the electronic polarizability, given respectively by

$$\alpha_e' = \alpha_{e0}\frac{1-(\omega/\omega_0)^2}{[1-(\omega/\omega_0)^2]^2 + (\gamma/\omega_0)^2(\omega/\omega_0)^2} \,, \tag{3.6a}$$

and

$$\alpha_e'' = \alpha_{e0}\frac{(\gamma/\omega_0)(\omega/\omega_0)}{[1-(\omega/\omega_0)^2]^2 + (\gamma/\omega_0)^2(\omega/\omega_0)^2} \,, \tag{3.6b}$$

where α_{e0} is the direct current (DC) polarizability corresponding to $\omega = 0$ and γ is the loss coefficient. Using (3.1), (3.2) and (3.5) with (3.6), it is then possible to study the frequency dependence of n and K. Figure 3.1a shows the dependence of n and K on the normalized frequency ω/ω_0 for a simple single electronic dipole oscillator of resonance frequency ω_0. It is seen that n and K peak close to $\omega = \omega_0$. If a material has a $\varepsilon_r'' \gg \varepsilon_r'$, then $\varepsilon_r \approx -i\varepsilon_r''$ and $n = K \approx \sqrt{\varepsilon_r''/2}$ is obtained from (3.2). Figure 3.1b shows the dependence of the reflectance R on the frequency. It is observed that R reaches its maximum value at a frequency slightly above $\omega = \omega_0$, and then remains high until ω reaches nearly $3\omega_0$; so the reflectance is substantial while absorption is strong. The normal dispersion region is the frequency range below ω_0 where n falls as the frequency decreases; that is, n decreases as the wavelength λ increases. The anomalous dispersion region is the frequency range above ω_0, where n decreases as ω increases. Below ω_0, K is small and if ε_{DC} is $\varepsilon_r(0)$, the refractive index becomes (approximately)

$$n^2 \approx 1 + (\varepsilon_{DC} - 1)\frac{\omega_0^2}{\omega_0^2 - \omega^2} \,; \quad \omega < \omega_0 \,. \tag{3.7}$$

Since $\lambda = 2\pi c/\omega$, defining $\lambda_0 = 2\pi c/\omega_0$ as the resonance wavelength, one gets

$$n^2 \approx 1 + (\varepsilon_{DC} - 1)\frac{\lambda^2}{\lambda^2 - \lambda_0^2} \,; \quad \lambda > \lambda_0 \,. \tag{3.8}$$

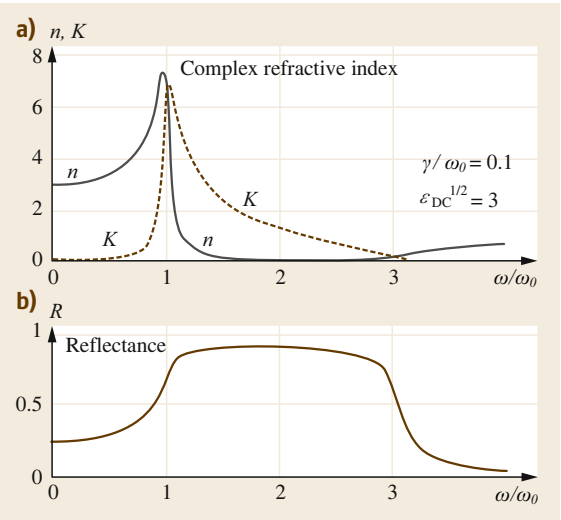


Fig. 3.1a,b The dipole oscillator model. **(a)** Refractive index and extinction coefficient versus normalized frequency. **(b)** Reflectance versus normalized frequency

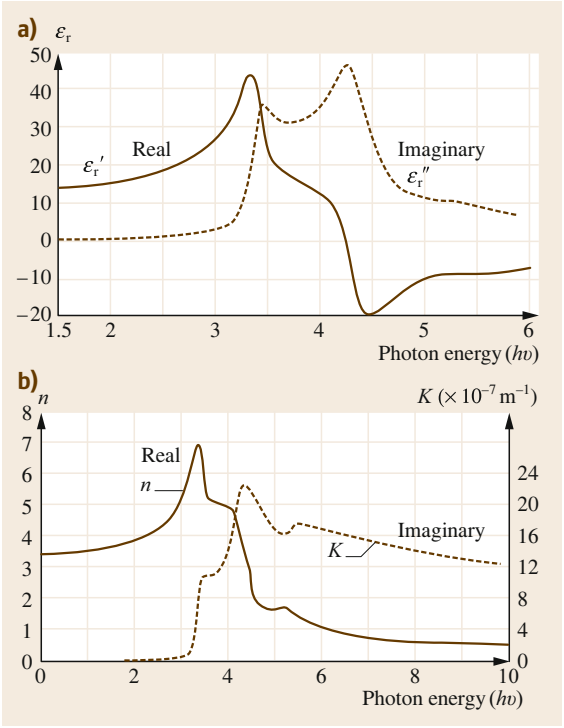


Fig. 3.2 **(a)** Complex relative permittivity of a silicon crystal as a function of photon energy plotted in terms of real (ε_r') and imaginary (ε_r'') parts. **(b)** Optical properties of a silicon crystal versus photon energy in terms of real (n) and imaginary (K) parts of the complex refractive index. (After [3.4])

While intuitively useful, the dispersion relation in (3.8) is far too simple. More rigorously, we have to consider the dipole oscillator quantum-mechanically, which means that a photon excites the oscillator to a higher energy level; see, for example, [3.2, 3]. The result is that we have a series of $\lambda^2/(\lambda^2-\lambda_i^2)$ terms with various weighting factors A_i that add to unity, where the λ_i represent different resonance wavelengths. The weighting factors A_i involve quantum-mechanical matrix elements.

Figure 3.2 shows the complex relative permittivity and the complex refractive index of crystalline silicon in terms of photon energy $h\nu$. For photon energies below the bandgap energy (1.1 eV), both ε_r'' and K are negligible and n is close to 3.7. Both ε_r'' and K increase and change strongly as the photon energy becomes greater than 3 eV, far beyond the bandgap energy. Notice that both ε_r' and n peak at $h\nu \approx 3.5$ eV, which corresponds to a direct photoexcitation process – the excitation of electrons from the valence band to the conduction band – as discussed later.

3.1.2 Kramers–Kronig Relations

If we know the frequency dependence of the real part, ε_r', of the relative permittivity of a material, then we can use the Kramers–Kronig relations to determine the frequency dependence of the imaginary part ε_r'', and vice versa. The transform requires that we know the frequency dependence of either the real or imaginary part over a frequency range that is as wide as possible, ideally from DC to infinity, and that the material is linear; in other words it has a relative permittivity that is independent of the applied field. The Kramers–Kronig relations for $\varepsilon_r = \varepsilon_r' + i\varepsilon_r''$ are given by [3.4]

$$\varepsilon_r'(\omega) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{\omega' \varepsilon_r''(\omega')}{\omega'^2 - \omega^2} \, d\omega' \,, \tag{3.9a}$$

and

$$\varepsilon_r''(\omega) = -\frac{2\omega}{\pi} P \int_0^\infty \frac{\varepsilon_r'(\omega')}{\omega'^2 - \omega^2} \, d\omega' \,, \tag{3.9b}$$

where ω' is the integration variable, P represents the Cauchy principal value of the integral and the singularity at $\omega = \omega'$ is avoided.

Similarly, one can relate the real and imaginary parts of the polarizability, $\alpha'(\omega)$ and $\alpha''(\omega)$, and those of the complex refractive index, $n(\omega)$ and $K(\omega)$ as well.

For $\alpha'(\omega)$ and $\alpha''(\omega)$, one can analogously write

$$\alpha'(\omega) = \frac{2}{\pi} P \int_0^\infty \frac{\omega' \alpha''(\omega')}{\omega'^2 - \omega^2} \, \mathrm{d}\omega' \tag{3.10a}$$

and

$$\alpha''(\omega') = -\frac{2\omega}{\pi} P \int_0^\infty \frac{\alpha'(\omega')}{\omega'^2 - \omega^2} \, \mathrm{d}\omega' \,. \tag{3.10b}$$

3.2 Refractive Index

There are several simplified models describing the spectral dependence of the refractive index n.

3.2.1 Cauchy Dispersion Equation

The dispersion relationship for the refractive index (n) versus the wavelength of light (λ) is stated in the following form

$$n = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} \,, \tag{3.11}$$

where A, B and C are material-dependent specific constants. The Cauchy equation (3.11) is typically used in the visible region of the spectrum for various optical glasses, and is applied to normal dispersion. The third term is sometimes dropped for a simpler representation of n versus λ behavior. The original expression was a series in terms of the wavelength, λ, or frequency, ω, of light as

$$n = a_0 + a_2\lambda^{-2} + a_4\lambda^{-4} + a_6\lambda^{-6} + \ldots \lambda > \lambda_\mathrm{h} \,, \tag{3.12a}$$

or

$$n = n_0 + n_2\omega^2 + n_4\omega^4 + n_6\omega^6 + \ldots \omega < \omega_\mathrm{h} \,, \tag{3.12b}$$

where $\hbar\omega$ is the photon energy, $\hbar\omega_\mathrm{h} = hc/\lambda_\mathrm{h}$ is the optical excitation threshold (the bandgap energy), while a_0, a_2 ... , and n_0, n_2, ... are constants. A Cauchy relation of the following form

$$n = n_{-2}(\hbar\omega)^{-2} + n_0 + n_2(\hbar\omega)^2 + n_4(\hbar\omega)^4 \,, \tag{3.13}$$

can be used satisfactorily over a wide range of photon energies. The dispersion parameters, calculated from (3.13), of a few materials are listed in Table 3.1.

Cauchy's dispersion relation, given in (3.13), was originally called the elastic-ether theory of the refractive index [3.5–7]. It has been widely used for many materials, although in recent years it has been largely replaced by the Sellmeier equation, which we consider next.

3.2.2 Sellmeier Dispersion Equation

The dispersion relationship can be quite complicated in practice. An example of this is the Sellmeier equation, which is an empirical relation between the refractive index n of a substance and the wavelength λ of light in the form of a series of $\lambda^2/(\lambda^2 - \lambda_i^2)$ terms, given by

$$n^2 = 1 + \frac{A_1\lambda^2}{\lambda^2 - \lambda_1^2} + \frac{A_2\lambda^2}{\lambda^2 - \lambda_2^2} + \frac{A_3\lambda^2}{\lambda^2 - \lambda_3^2} + \cdots \,, \tag{3.14}$$

where λ_i is a constant and A_1, A_2, A_3, λ_1, λ_2 and λ_3 are called Sellmeier coefficients, which are determined by fitting this expression to the experimental data. The full Sellmeier formula has more terms of similar form, such as $A_i\lambda^2/(\lambda^2 - \lambda_i^2)$, where $i = 4, 5, \ldots$ but these can generally be neglected when considering n versus λ behavior over typical wavelengths of interest and by ensuring that the three terms included in the Sellmeier equation correspond to the most important or relevant terms in the summation. Examples of Sellmeier coefficients for some materials, including pure silica (SiO_2) and 86.5 mol% SiO_2-13.5 mol% GeO_2, are given in Table 3.2. Two methods are used to find the refractive index of silica-germania glass $(SiO_2)_{1-x}(GeO_2)_x$: (a) a simple, but approximate, linear interpolation of the refractive index between known compositions, for example $n(x) - n(0.135) = (x - 0.135)[n(0.135) - n(0)]/0.135$ for $(SiO_2)_{1-x}(GeO_2)_x$, so $n(0.135)$ is used

Table 3.1 Cauchy's dispersion parameters (obtained from (3.11)) for a few materials. (Data compiled from [3.5])

Material	$\hbar\omega$ (eV)	n_{-2} (eV^2)	n_0	n_2 (eV^{-2})	n_4 (eV^{-4})
Diamond	0.05–5.47	-1.07×10^{-5}	2.378	8.01×10^{-3}	1.04×10^{-4}
Si	0.002–1.08	-2.04×10^{-8}	3.4189	8.15×10^{-2}	1.25×10^{-2}
Ge	0.002–0.75	-1.00×10^{-8}	4.0030	2.20×10^{-1}	1.40×10^{-1}

Table 3.2 Sellmeier coefficients of a few materials (λ_1, λ_2, λ_3 are in μm)

Material	A_1	A_2	A_3	λ_1	λ_2	λ_3
SiO_2 (fused silica)	0.696749	0.408218	0.890815	0.0690660	0.115662	9.900559
86.5%SiO_2-13.5%GeO_2	0.711040	0.451885	0.704048	0.0642700	0.129408	9.425478
GeO_2	0.80686642	0.71815848	0.85416831	0.068972606	0.15396605	11.841931
Barium fluoride	0.63356	0.506762	3.8261	0.057789	0.109681	46.38642
Sapphire	1.023798	1.058264	5.280792	0.0614482	0.110700	17.92656
Diamond	0.3306	4.3356		0.175	0.106	
Quartz, n_0	1.35400	0.010	0.9994	0.092612	10.700	9.8500
Quartz, n_e	1.38100	0.0100	0.9992	0.093505	11.310	9.5280
KDP, n_0	1.2540	0.0100	0.0992	0.09646	6.9777	5.9848
KDP, n_e	1.13000	0.0001	0.9999	0.09351	7.6710	12.170

for 86.5 mol% SiO_2-13.5 mol% GeO_2 and $n(0)$ is used for SiO_2; (b) an interpolation for the coefficients A_i and λ_i between SiO_2 and GeO_2

$$n^2 - 1 = \frac{\{A_1(S) + X[A_1(G) - A_1(S)]\}\lambda^2}{\lambda^2 - \{\lambda_1(S) + X[\lambda_1(G) - \lambda_1(S)]\}^2} + \cdots , \tag{3.15}$$

where X is the atomic fraction of germania, S and G in parentheses refer to silica and germania [3.8]. The theoretical basis of the Sellmeier equation is that the solid is represented as a sum of N lossless (frictionless) Lorentz oscillators such that each takes the form of $\lambda^2/(\lambda^2 - \lambda_i^2)$ with different λ_i, and each has different strengths, with weighting factors (A_i, $i = 1, 2, \ldots N$) [3.9, 10]. Knowledge of appropriate dispersion relationships is essential when designing photonic devices, such as waveguides.

There are other dispersion relationships that inherently take account of various contributions to optical properties, such as the electronic and ionic polarization and the interactions of photons with free electrons. For example, for many semiconductors and ionic crystals, two useful dispersion relations are

$$n^2 = A + \frac{B\lambda^2}{\lambda^2 - C} + \frac{D\lambda^2}{\lambda^2 - E} , \tag{3.16}$$

and

$$n^2 = A + \frac{B}{\lambda^2 - \lambda_0^2} + \frac{C}{\left(\lambda^2 - \lambda_0^2\right)^2} + D\lambda^2 + E\lambda^4 , \tag{3.17}$$

where A, B, C, D, E and λ_0 are constants particular to a given material. Table 3.3 provides a few examples.

The refractive index of a semiconductor material typically decreases with increasing bandgap energy E_g. There are various empirical and semi-empirical rules and expressions that relate n to E_g. In Moss' rule, n and E_g are related by $n^4 E_g = K =$ constant (≈ 100 eV). In the *Hervé–Vandamme* relationship [3.11],

$$n^2 = 1 + \left(\frac{A}{E_g + B}\right)^2 , \tag{3.18}$$

where A and B are constants ($A \approx 13.6$ eV and $B \approx 3.4$ eV and $dB/dT \approx 2.5 \times 10^{-5}$ eV/K). The refractive index typically increases with increasing temperature. The temperature coefficient of the refractive index (TCRI) of a semiconductor can be found from the *Hervé–Vandamme* relationship [3.11]

$$\mathrm{TCRI} = \frac{1}{n}\frac{dn}{dT} = \frac{(n^2 - 1)^{3/2}}{13.6n^2}\left(\frac{dE_g}{dT} + \frac{dB}{dT}\right) . \tag{3.19}$$

TCRI is typically in the range 10^{-6}–10^{-4}.

3.2.3 Gladstone–Dale Formula

The *Gladstone–Dale formula* is an empirical equation that allows the average refractive index n of an oxide glass to be calculated from its density ρ and its constituents as

$$\frac{n-1}{\rho} = p_1 k_1 + p_2 k_2 + \cdots = \sum_{i=1}^{N} p_i k_i = C_{GD} , \tag{3.20}$$

where the summation is for various oxide components (each a simple oxide), p_i is the weight fraction of the i-th oxide in the compound, and k_i is the refraction coefficient that represents the polarizability of the i-th oxide. The right-hand side of (3.20) is called the Gladstone–Dale coefficient C_{GD}. In more general terms, as a mixture rule for the overall refractive index, the Gladstone–Dale formula is frequently written as

$$\frac{n-1}{\rho} = \frac{n_1 - 1}{\rho_1} w_1 + \frac{n_2 - 1}{\rho_2} w_2 + \cdots , \tag{3.21}$$

Table 3.3 Parameters from (3.16) and (3.17) for some selected materials (Si data from [3.12]; others from [3.13])

Material	λ_0 (μm)	A	B (μm)^2	C (μm)$^{-4}$	D (μm)$^{-2}$	E (μm)$^{-4}$
Silicon	0.167	3.41983	0.159906	-0.123109	1.269×10^{-6}	-1.951×10^{-9}
MgO	0.11951	2.95636	0.021958	0	-1.0624×10^{-2}	-2.05×10^{-5}
LiF	0.16733	1.38761	0.001796	-4.1×10^{-3}	-2.3045×10^{-3}	-5.57×10^{-6}
AgCl	0.21413	4.00804	0.079009	0	-8.5111×10^{-4}	-1.976×10^{-7}

Table 3.4 Examples of parameters for the Wemple–DiDomenico dispersion relationship (3.23), for various materials

Material	N_c	Z_a	N_e	E_0 (eV)	E_d (eV)	β (eV)	β	Comment
NaCl	6	1	8	10.3	13.6	0.28	β_i	Halides, LiF, NaF, etc.
CsCl	8	1	8	10.6	17.1	0.27	β_i	CsBr, CsI, etc.
TlCl	8	1	10	5.8	20.6	0.26	β_i	TlBr
CaF_2	8	1	8	15.7	15.9	0.25	β_i	BaF_2, etc.
CaO	6	2	8	9.9	22.6	0.24	β_i	Oxides, MgO, TeO_2, etc.
Al_2O_3	6	2	8	13.4	27.5	0.29	β_i	
$LiNbO_3$	6	2	8	6.65	25.9	0.27	β_i	
TiO_2	6	2	8	5.24	25.7	0.27	β_i	
ZnO	4	2	8	6.4	17.1	0.27	β_i	
ZnSe	4	2	8	5.54	27	0.42	β_c	II–VI, Zinc blende, ZnS, ZnTe, CdTe
GaAs	4	3	8	3.55	33.5	0.35	β_c	III–V, Zinc blende, GaP, etc.
Si (crystal)	4	4	8	4.0	44.4	0.35	β_c	Diamond, covalent bonding; C (diamond), Ge, β-SiC etc.
SiO_2 (crystal)	4	2	8	13.33	18.10	0.28	β_i	Average crystalline form
SiO_2 (amorphous)	4	2	8	13.38	14.71	0.23	β_i	Fused silica
CdSe	4	2	8	4.0	20.6	0.32	$\beta_i-\beta_c$	Wurtzite

where n and ρ are the effective refractive index and effective density of the whole mixture respectively, n_1, n_2, ... are the refractive indices of the constituents, and ρ_1, ρ_2, ... represent the densities of each constituent. Gladstone–Dale equations for the polymorphs of SiO_2 and TiO_2 give the average n values as [3.14]

$$n(SiO_2) = 1+0.21\rho \quad \text{and} \quad n(TiO_2) = 1+0.40\rho\,. \tag{3.22}$$

3.2.4 Wemple–DiDominico Dispersion Relation

The *Wemple–DiDominico dispersion relation* is a semi-empirical relationship used to find the refractive indices of a variety of materials for photon energies below the interband absorption edge, given by [3.15]

$$n^2 = 1+\frac{E_0E_d}{E_0^2-(h\nu)^2}\,, \tag{3.23}$$

where ν is the frequency, h is the Planck constant, E_0 is the single oscillator energy, E_d is the dispersion energy, which is a measure of the average strength of interband optical transitions; $E_d = \beta N_c Z_a N_e$ (eV), where N_c is the effective coordination number of the cation nearest-neighbor to the anion ($N_c = 6$ in NaCl, $N_c = 4$ in Ge), Z_a is the formal chemical valency of the anion ($Z_a = 1$ in NaCl; 2 in Te; 3 in GaP), N_e is the effective number of valence electrons per anion excluding the cores ($N_e = 8$ in NaCl, Ge; 10 in TlCl; 12 in Te; $9\frac{1}{3}$ in As_2Se_3), and β is a constant that depends on whether the interatomic bond is ionic (β_i) or covalent (β_c): $\beta_i = 0.26\pm0.04$ eV (this applies to halides such as NaCl and ThBr and most oxides, including Al_2O_3, for example), while $\beta_c = 0.37\pm0.05$ eV (applies to tetrahedrally bonded A^NB^{8-N} zinc blende- and diamond-type structures such as GaP, ZnS, for instance). (Note that wurtzite crystals have a β that is intermediate between β_i and β_c.) Also, empirically,

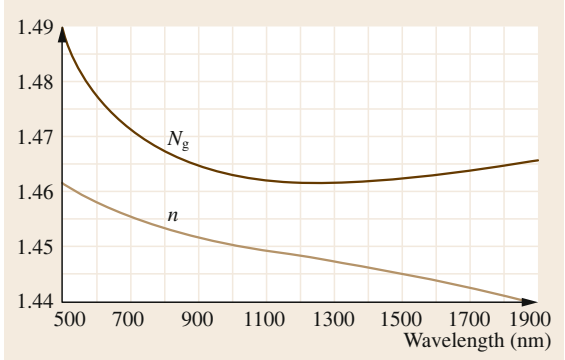


Fig. 3.3 Refractive index n and the group index N_g of pure SiO_2 (silica) glass as a function of wavelength

$E_0 = CE_g(D)$, where $E_g(D)$ is the lowest direct bandgap and C is a constant, typically ≈ 1.5. E_0 has been associated with the main peak in the $\varepsilon_r''(h\nu)$ versus $h\nu$ spectrum. The parameters required to calculate n from (3.23) are listed in Table 3.4 [3.15].

3.2.5 Group Index (N_g)

The group index represents the factor by which the group velocity of a group of waves in a dielectric medium is reduced with respect to propagation in free space, $N_g = c/v_g$, where v_g is the group velocity. The group index can be determined from the ordinary refractive index n via

$$N_g = n - \lambda \frac{dn}{d\lambda}, \tag{3.24}$$

where λ is the wavelength of light. The relation between N_g and n for SiO_2 is illustrated in Fig. 3.3.

3.3 Optical Absorption

The main optical properties of a semiconductor are typically its refractive index n and its extinction coefficient K or absorption coefficient α (or equivalently the real and imaginary parts of the relative permittivity), as well as their dispersion relations (their dependence on the electromagnetic radiation wavelength λ or photon energy $h\nu$) and the changes in the dispersion relations with temperature, pressure, alloying, impurities, and so on. A typical relationship between the absorption coefficient and the photon energy observed in a crystalline semiconductor is shown in Fig. 3.4, where various possible absorption processes are illustrated. The important features of the behavior of the α versus $h\nu$ can be summarized as follows:

1. Free-carrier absorption due to the presence of free electrons and holes, an effect that decreases with increasing photon energy
2. An impurity absorption band (usually narrow) due the various dopants
3. Reststrahlen or lattice absorption in which the radiation is absorbed by vibrations of the crystal ions
4. Exciton absorption peaks that are usually observed at low temperatures and are close the fundamental absorption edge; and
5. Band-to-band or fundamental absorption of photons, which excites an electron from the valence to the conduction band.

Type 5 absorption has a large absorption coefficient and occurs when the photon energy reaches the bandgap energy E_g. It is probably the most important absorption effect; its characteristics for $h\nu > E_g$ can be predicted using the results of Sect. 3.3.3. The values of E_g, and its temperature shift, dE_g/dT, are therefore important factors in semiconductor-based optoelectronic devices. In nearly all semiconductors E_g decreases with temperature, hence shifting the fundamental absorption to longer wavelengths. The refractive index n also changes with temperature. dn/dT depends on the wavelength, but for many semiconductors $(dn/dT)/n \approx 5 \times 10^{-5}\,K^{-1}$; for example, for GaAs, $(dn/dT)/n = 4 \times 10^{-4}\,K^{-1}$ at $\lambda = 2\,\mu m$. There is a good correlation between the refractive indices and the bandgaps of semiconductors in which, typically, n decreases as E_g increases; semiconductors with wider bandgaps have lower refractive indices. The refractive index n and the extinction coefficient K (or α) are related by the Kramers–Kronig relations. Thus, large increases in the absorption coefficient for $h\nu$ near and above the bandgap energy E_g also result in increases in the refractive index n versus $h\nu$ in this region. Optical (and some structural) properties of various semiconductors are listed in Table 3.5.

3.3.1 Lattice or Reststrahlen Absorption and Infrared Reflection

In the infrared wavelength region, ionic crystals reflect and absorb light strongly due to the resonance interaction of the electromagnetic (EM) wave field with the transverse optical (TO) phonons. The dipole oscillator

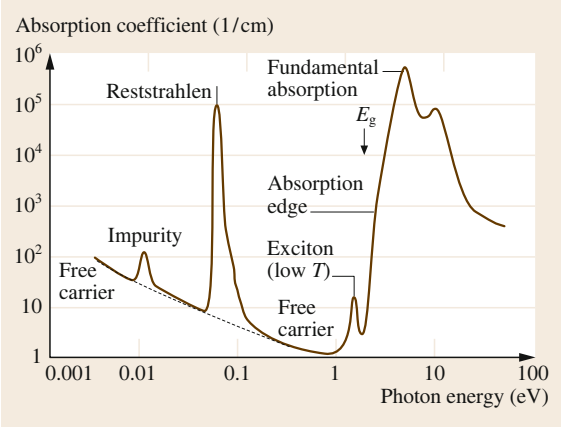


Fig. 3.4 Absorption coefficient plotted as a function of the photon energy in a typical semiconductor, illustrating various possible absorption processes

Table 3.5 Crystal structure, lattice parameter a, bandgap energy E_g at 300 K, type of bandgap (D=direct and I=indirect), change in E_g per unit temperature change (dE_g/dT) at 300 K, bandgap wavelength λ_g and refractive index n close to λ_g (A=amorphous, D=diamond, W=wurtzite, ZB=zinc blende). Approximate data from various sources

Semiconductors	Crystal	a (nm)	E_g (eV)	Type	dE_g/dT (meV/K)	λ_g (μm)	$n(\lambda_g)$	dn/dT (10^{-5} K^{-1})
Group IV								
Diamond	D	0.3567	5.48	I	−0.05	0.226	2.74	1.1
Ge	D	0.5658	0.66	I	−0.37	1.87	4	27.6 42.4 (4 μm)
Si	D	0.5431	1.12	I	−0.25	1.11	3.45	13.8 16 (5 μm)
a-Si:H	A		1.7–1.8			0.73		
SiC(α)	W	0.3081 a 1.5120 c	2.9	I	−0.33	0.42	2.7	9
III–V Compounds								
AlAs	ZB	0.5661	2.16	I	−0.50	0.57	3.2	15
AlP	ZB	0.5451	2.45	I	−0.35	0.52	3	11
AlSb	ZB	0.6135	1.58	I	−0.3	0.75	3.7	
GaAs	ZB	0.5653	1.42	D	−0.45	0.87	3.6	15
$GaAs_{0.88}Sb_{0.12}$	ZB		1.15	D		1.08		
GaN	W	0.3190 a 0.5190 c	3.44	D	−0.45	0.36	2.6	6.8
GaP	ZB	0.5451	2.26	I	−0.54	0.40	3.4	
GaSb	ZB	0.6096	0.73	D	−0.35	1.7	4	33
$In_{0.53}Ga_{0.47}As$ on InP	ZB	0.5869	0.75	D		1.65		
$In_{0.58}Ga_{0.42}As_{0.9}P_{0.1}$ on InP	ZB	0.5870	0.80	D		1.55		
$In_{0.72}Ga_{0.28}As_{0.62}P_{0.38}$ on InP	ZB	0.5870	0.95	D		1.3		
InP	ZB	0.5869	1.35	D	−0.40	0.91	3.4–3.5	9.5
InAs	ZB	0.6058	0.36	D	−0.28	3.5	3.8	2.7
InSb	ZB	0.6479	0.18	D	−0.3	7	4.2	29
II–VI Compounds								
ZnSe	ZB	0.5668	2.7	D	−0.50	0.46	2.3	6.3
ZnTe	ZB	0.6101	2.3	D	−0.45	0.55	2.7	

model based on ions driven by an EM wave results in

$$\varepsilon_r = \varepsilon_r' - i\varepsilon_r'' = \varepsilon_{r\infty} + \frac{\varepsilon_{r\infty} - \varepsilon_{r0}}{\left(\frac{\omega}{\omega_T}\right)^2 - 1 + i\frac{\gamma}{\omega_T}\left(\frac{\omega}{\omega_T}\right)}, \tag{3.25}$$

where ε_{r0} and $\varepsilon_{r\infty}$ are the relative permittivity at $\omega = 0$ (very low frequencies) and $\omega = \infty$ (very high frequencies) respectively, γ is the loss coefficient per unit reduced mass, representing the rate of energy transfer from the EM wave to optical phonons, and ω_0 is a resonance frequency that is related to the *spring* constant between the ions. By definition, the frequency ω_T is

$$\omega_T^2 = \omega_0^2\left(\frac{\varepsilon_{r\infty} + 2}{\varepsilon_{r0} + 2}\right). \tag{3.26}$$

The loss ε_r'' and the absorption are maxima when $\omega = \omega_T$, and the wave is attenuated by the transfer of energy to the transverse optical phonons, so the EM wave couples to the transverse optical phonons. At $\omega = \omega_L$, the wave couples to the longitudinal optical (LO) phonons. Figure 3.5 shows the optical properties of AlSb [3.16] in terms of n, K and R versus wavelength. The peaks in the extinction coefficient K and reflectance R occur over roughly the same wavelength region, corresponding to the coupling of the EM wave to the transverse optical phonons. At wavelengths close to $\lambda_T = 2\pi/\omega_T$, n and K peak, and there is strong absorption of light (which corresponds to the EM wave resonating with the TO lattice vibrations), and then R rises sharply.

3.3.2 Free-Carrier Absorption (FCA)

An electromagnetic wave with sufficiently low frequency oscillations can interact with free carriers in

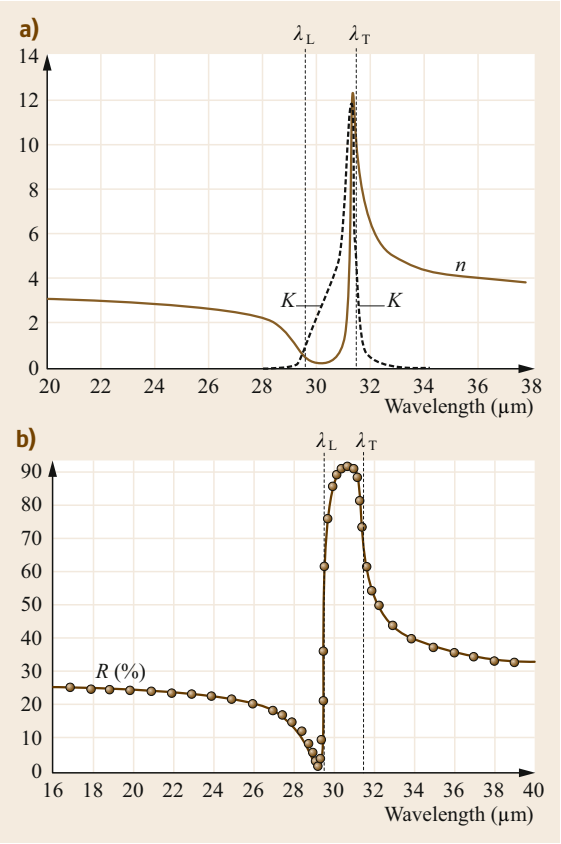


Fig. 3.5a,b Infrared refractive index n, extinction coefficient K **(a)**, and reflectance R **(b)** of AlSb. Note that the wavelength axes are not identical, and the wavelengths λ_T and λ_L, corresponding to ω_T and ω_L respectively, are shown as *dashed vertical lines*. (After [3.16])

a material and thereby drift the carriers. This interaction results in an energy loss from the EM wave to the lattice vibrations through the carrier scattering processes. Based on the Drude model, the relative permittivity $\varepsilon_r(\omega)$ due to N free electrons per unit volume is given by

$$\varepsilon_r = \varepsilon_r' - i\varepsilon_r'' = 1 - \frac{\omega_p^2}{\omega^2 - i\omega/\tau} ,$$
$$\omega_p^2 = \frac{Ne^2}{\varepsilon_0 m_e} , \qquad (3.27)$$

where ω_p is a plasma frequency that depends on the electron concentration, while τ is the relaxation time of the scattering process (the mean scattering time). For metals where the electron concentration is very large, ω_p is of the order of $\approx 10^{16}$ rad/s, at ultraviolet (UV) frequencies, and for $\omega > \omega_p$, $\varepsilon_r \approx 1$, and the reflectance becomes very small. Metals lose their free-electron reflectance in the UV range, thus becoming UV-transparent. The reflectance does not fall to zero because there are other absorption processes such as interband electron excitations or excitations from core levels to energy bands. Plasma edge transparency, where the reflectance almost vanishes, can also be observed in doped semiconductors. For example, the reflectance of doped InSb has a plasma edge wavelength that decreases with increasing free-carrier concentration [3.17]. Equation (3.27) can be written in terms of the conductivity σ_0 at low frequencies (DC) as

$$\varepsilon_r = \varepsilon_r' - i\varepsilon_r'' = 1 - \frac{\tau\sigma_0}{\varepsilon_0[(\omega\tau)^2 + 1]} - i\frac{\sigma_0}{\varepsilon_0\omega[(\omega\tau)^2 + 1]} . \qquad (3.28)$$

In metals, σ_0 is high. At frequencies where $\omega < 1/\tau$, the imaginary part $\varepsilon_r'' = \sigma_0/(\varepsilon_0\omega)$ is normally much more than 1, and $n = K \approx \sqrt{\varepsilon_r''/2}$, so that the free-carrier attenuation coefficient α is then given by

$$\alpha = 2k_0K \approx \frac{2\omega}{c}\left(\frac{\varepsilon_r''}{2}\right)^{1/2} \approx (2\sigma_0\mu_0)^{1/2}\,\omega^{1/2} . \qquad (3.29)$$

Furthermore, the reflectance can also be calculated using $n = K \approx \sqrt{\varepsilon_r''/2}$, which leads to the well-known *Hagen–Rubens* relationship [3.18]

$$R \approx 1 - 2\left(\frac{2\omega\varepsilon_0}{\sigma_0}\right)^{1/2} . \qquad (3.30)$$

In semiconductors, one typically encounters $\sigma_0/(\varepsilon_0\omega) < 1$, since the free-electron concentration is small, and we can treat n as constant due to various other polarization mechanisms such as electronic polarization. Since $2nK = \varepsilon_r''$, the absorption coefficient becomes [3.19]

$$\alpha = 2k_0K \approx \frac{2\omega}{c}\left(\frac{\varepsilon_r''}{2n}\right) = \frac{\sigma_0}{nc\varepsilon_0[(\omega\tau)^2 + 1]} . \qquad (3.31)$$

At low frequencies where $\omega < 1/\tau$, we have $\alpha(\lambda) \propto \sigma_0/n(\lambda)$ so that α should be controlled by the DC conductivity, and hence the amount of doping. Furthermore, α will exhibit the frequency dependence of the refractive index n ($\alpha(\lambda) \propto 1/n(\lambda)$), in which n will typically be determined by the electronic polarization of the crystal.

At high frequencies where $\omega > 1/\tau$,

$$\alpha \propto \frac{\sigma_0}{\omega^2} \propto N\lambda^2 , \qquad (3.32)$$

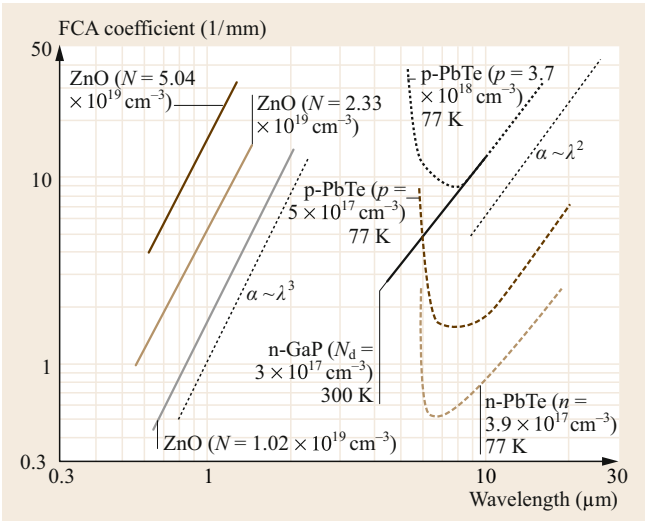


Fig. 3.6 Free-carrier absorption in n-GaP at 300 K [3.20], p- and n-type PbTe [3.21] at 77 K and In doped n-type ZnO at room temperature. (After [3.22])

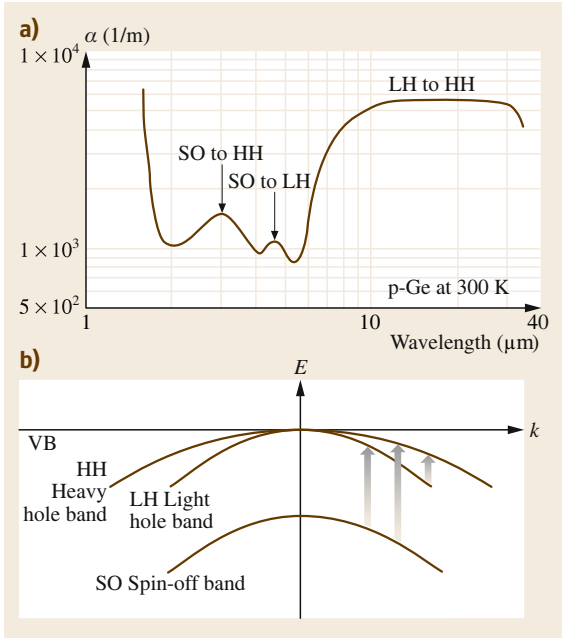


Fig. 3.7 **(a)** Free-carrier absorption due to holes in p-Ge [3.27]. **(b)** The valence band of Ge has three bands; heavy hole, light hole and spin-off bands

where α is proportional to N, the free-carrier concentration, and λ^2. Experimental observations on FCA in doped semiconductors are in general agreement with these predictions. For example, α increases with N, whether N is increased by doping or by carrier injection [3.23]. However, not all semiconductors show the simple $\alpha \propto \lambda^2$ behavior. A proper account of the field-driven electron motion and scattering must consider the fact that τ will depend on the electron energy. The correct approach is to use the Boltzmann transport equation [3.24] with the appropriate scattering mechanism. FCA can be calculated using a quantum-mechanical approach based on second-order time-dependent perturbation theory with Fermi–Dirac statistics [3.25].

Absorption due to free carriers is commonly written as $\alpha \propto \lambda^p$, where the index p depends primarily on the scattering mechanism, although it is also influenced by the compensation doping ratio, if the semiconductor has been doped by compensation, and the free-carrier concentration. In the case of lattice scattering, one must consider scattering from acoustic and optical phonons. For acoustic phonon scattering $p \approx 1.5$, for optical phonon scattering $p \approx 2.5$, and for impurity scattering $p \approx 3.5$. The observed free-carrier absorption coefficient will then have all three contributions

$$\alpha = A_{\text{acoustic}}\lambda^{1.5} + A_{\text{optical}}\lambda^{2.5} + A_{\text{impurity}}\lambda^{3.5} \,. \quad (3.33)$$

Inasmuch as α for FCA depends on the free-carrier concentration N, it is possible to evaluate the latter from the experimentally measured α, given its wavelength dependence and p as discussed by [3.26]. The free-carrier absorption coefficients α (mm^{-1}) for GaP, n-type PbTe and n-type ZnO are shown in Fig. 3.6.

Free-carrier absorption in p-type Ge demonstrates how the FCA coefficient α can be dramatically different than what is expected from (3.33). Figure 3.7a shows the wavelength dependence of the absorption coefficient for p-Ge over the wavelength range from about 2 to 30 μm [3.27]. The observed absorption is due to excitations of electrons from the spin-off band to the heavy hole band, from the spin-off band to the light hole band, and from the light hole band to the heavy hole band, as marked in the Fig. 3.7b.

3.3.3 Band-to-Band or Fundamental Absorption

Crystalline Solids

Band-to-band absorption or fundamental absorption of radiation is due to the photoexcitation of an electron from the valence band to the conduction band. There are two types of band-to-band absorptions, corresponding to direct and indirect transitions.

A direct transition is a photoexcitation process in which no phonons are involved. The photon momentum is negligible compared with the electron momentum, so that when the photon is absorbed, exciting an electron from the valence band (VB) to the conduction band (CB); the electron's k-vector does not change. A direct

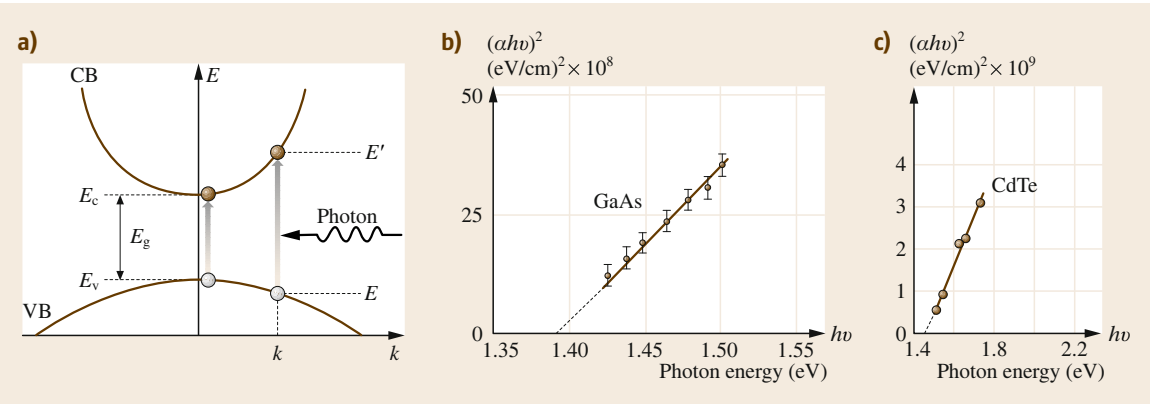


Fig. 3.8 **(a)** A direct transition from the valence band (VB) to the conduction band (CB) through the absorption of a photon. Absorption behavior represented as $(\alpha h\nu)^2$ versus photon energy $h\nu$ near the band edge for single crystals of **(b)** p-type GaAs, from [3.28] and **(c)** CdTe. (After [3.29])

transition on an $E-k$ diagram is a vertical transition from an initial energy E and wavevector k in the VB to a final energy E' and a wavevector k' in the CB where $k' = k$, as shown in Fig. 3.8a. The energy $(E' - E_c)$ is the kinetic energy $(\hbar k)^2/(2m_e^*)$ of the electron with an effective mass m_e^*, and $(E_v - E)$ is the kinetic energy $(\hbar k)^2/(2m_h^*)$ of the hole left behind in the VB. The ratio of the kinetic energies of the photogenerated electron and hole depends inversely on the ratio of their effective masses.

The absorption coefficient α is derived from the quantum-mechanical probability of transition from E to E', the occupied density of states at E in the VB from which electrons are excited, and the unoccupied density of states in the CB at $E + h\nu$. Thus, α depends on the joint density of states at E and $E + h\nu$, and we have to suitably integrate this joint density of states. Near the band edges, the density of states can be approximated by a parabolic band, and α rises with the photon energy following

$$\alpha h\nu = A(h\nu - E_g)^{1/2} , \tag{3.34}$$

where the constant $A \approx [e^2/(nch^2 m_e^*)](2\mu^*)^{3/2}$ in which μ^* is a reduced electron and hole effective mass, n is the refractive index, and E_g is the direct bandgap, with minimum $E_c - E_v$ at the same k value (see [3.3] for the derivation). Experiments indeed show this type of behavior for photon energies above E_g and close to E_g, as shown in Fig. 3.8b for a GaAs crystal [3.28] and in Fig. 3.8c for a CdTe crystal [3.29]. The extrapolation to zero photon energy gives the direct bandgap E_g, which is about 1.40 eV for GaAs and 1.46–1.49 eV for CdTe. For photon energies very close to the bandgap energy, the absorption is usually due to exciton absorption, especially at low temperatures, and is discussed later in this chapter.

In indirect bandgap semiconductors, such as Si and Ge, the photon absorption for photon energies near E_g requires the absorption and emission of phonons during the absorption process, as illustrated in Fig. 3.9a. The absorption onset corresponds to a photon energy of $(E_g - h\vartheta)$, which represents the absorption of a phonon with energy $h\vartheta$. For the latter, α is proportional to $[h\nu - (E_g - h\vartheta)]^2$. Once the photon energy reaches $(E_g + h\vartheta)$, then the photon absorption process can also occur by phonon emission, for which the absorption coefficient is larger than that for phonon absorption. The absorption coefficients for the phonon absorption and emission processes are given by [3.30]

$$\alpha_{\text{absorption}} = A[f_{BE}(h\vartheta)][h\nu - (E_g - h\vartheta)]^2 ; \quad h\nu > (E_g - h\vartheta) , \tag{3.35}$$

and

$$\alpha_{\text{emission}} = A[(1 - f_{BE}(h\vartheta)][h\nu - (E_g + h\vartheta)]^2 ; \quad h\nu > (E_g + h\vartheta) , \tag{3.36}$$

where A is a constant and $f_{BE}(h\vartheta)$ is the Bose–Einstein distribution function at the phonon energy $h\vartheta$, $f_{BE}(h\vartheta) = [\exp(h\vartheta/(k_B T)) - 1]^{-1}$, where k_B is the Boltzmann constant and T is the temperature. As we increase the photon energy in the range $(E_g - h\vartheta) < h\nu < (E_g + h\vartheta)$, the absorption is controlled by $\alpha_{\text{absorption}}$ and the plot of $\alpha^{1/2}$ versus $h\nu$ has an intercept of $(E_g - h\vartheta)$.

On the other hand, for $h\nu > (E_g + h\vartheta)$, the overall absorption coefficient is $\alpha_{\text{absorption}} + \alpha_{\text{emission}}$, but at slightly higher photon energies than $(E_g + h\vartheta)$, α_{emission} quickly dominates over $\alpha_{\text{absorption}}$ since $f_{BE}(h\vartheta) \gg 1 - f_{BE}(h\vartheta)$. Figure 3.9b shows the behavior of $\alpha^{1/2}$ versus

photon energy for Si at two temperatures for $h\nu$ near band edge absorption. At low temperatures, $f_{BE}(h\vartheta)$ is small and $\alpha_{absorption}$ decreases with decreasing temperature, as apparent from Fig. 3.9b. Equations (3.35) and (3.36) intersect the photon energy axis at $(E_g - h\vartheta)$ and $(E_g + h\vartheta)$, which can be used to obtain E_g.

An examination of the extinction coefficient K or ε_r'' versus photon energy for Si in Fig. 3.2 shows that absorption peaks at certain photon energies, $h\nu \approx 3.5$ and 4.3 eV. These peaks are due to the fact that the joint density of states function peaks at these energies. The absorption coefficient peaks whenever there is a direct transition in which the E versus $\boldsymbol{k}$ curve in the VB is parallel to the E versus $\boldsymbol{k}$ curve in the CB, as schematically illustrated in Fig. 3.9a, where a photon of energy $h\nu_{12}$ excites an electron from state 1 in the VB to state 2 in the CB in a direct transition, i. e., $\boldsymbol{k}_1 = \boldsymbol{k}_2$. Such transitions where E versus $\boldsymbol{k}$ curves are parallel at a photon energy $h\nu_{12}$ result in a peak in the absorption versus photon energy behavior, and can be represented by the condition that

$$(\nabla_{\boldsymbol{k}} E)_{CB} - (\nabla_{\boldsymbol{k}} E)_{VB} = 0\,. \tag{3.37}$$

The above condition is normally interpreted as the joint density of states reaching a peak value at certain points in the Brillouin zone called van Hove singularities. Identification of peaks in K versus $h\nu$ involves the examination of all E versus $\boldsymbol{k}$ curves of a given crystal that can participate in a direct transition. The silicon ε_r'' peaks at $h\nu \approx 3.5$ eV and 4.3 eV correspond to (3.37) being satisfied at points L, along $\langle 111\rangle$ in $\boldsymbol{k}$-space, and X along $\langle 100\rangle$ in $\boldsymbol{k}$-space, at the edges of the Brillouin zone.

In degenerate semiconductors, the Fermi level E_F is in a band; for example, in the CB for a degenerate n-type semiconductor. Electrons in the VB can only be excited to states above E_F in the CB rather than to the bottom of the CB. The absorption coefficient then depends on the free-carrier concentration since the latter determines E_F. Fundamental absorption is then said to depend on band filling, and there is an apparent shift in the absorption edge, called the *Burstein–Moss shift*. Furthermore, in degenerate indirect semiconductors, the indirect transition may involve a nonphonon scattering process, such as impurity or electron-electron scattering, which can change the electron's wavevector $\boldsymbol{k}$. Thus, in degenerate indirect bandgap semiconductors, absorption can occur without phonon assistance and the absorption coefficient becomes

$$\alpha \propto [h\nu - (E_g + \Delta E_F)]^2\,, \tag{3.38}$$

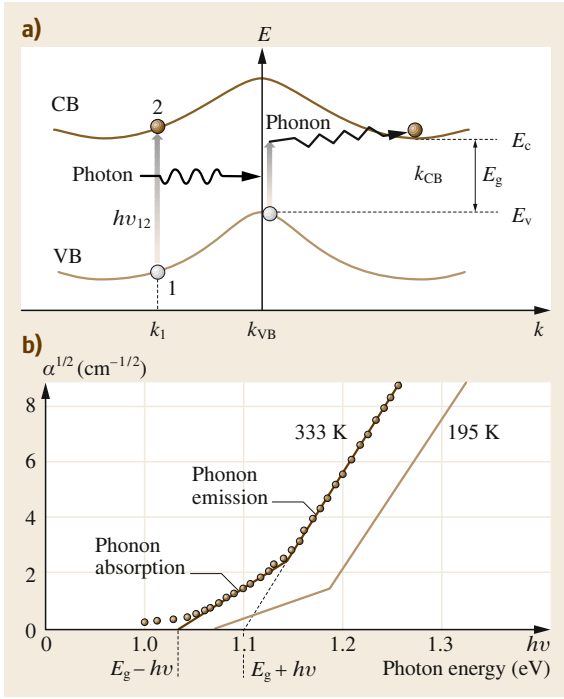


Fig. 3.9 **(a)** Indirect transitions across the bandgap involve phonons. Direct transitions in which dE/dk in the CB is parallel to dE/dk in the VB lead to peaks in the absorption coefficient. **(b)** Fundamental absorption in Si at two temperatures. The overall behavior is well described by (3.35) and (3.36)

where ΔE_F is the energy depth of E_F into the band measured from the band edge.

Heavy doping of degenerate semiconductors normally leads to a phenomenon called bandgap narrowing and bandtailing. Bandtailing means that the band edges at E_v and E_c are no longer well-defined cut-off energies, and there are electronic states above E_v and below E_c where the density of states falls sharply away with energy from the band edges. Consider a degenerate direct bandgap p-type semiconductor. One can excite electrons from states below E_F in the VB, where the band is nearly parabolic, to tail states below E_c, where the density of states decreases exponentially with energy into the bandgap, away from E_c. Such excitations lead to α depending exponentially on $h\nu$, a dependence that is called the *Urbach rule* [3.31, 32], given by

$$\alpha = \alpha_0 \exp\left(\frac{h\nu - E_0}{\Delta E}\right)\,, \tag{3.39}$$

where α_0 and E_0 are material-dependent constants, and ΔE, called the Urbach width, is also a material-dependent constant. The Urbach rule was originally

reported for alkali halides. It has been observed for many ionic crystals, degenerately doped crystalline semiconductors, and almost all amorphous semiconductors. While exponential bandtailing can explain the observed Urbach tail of the absorption coefficient versus photon energy, it is also possible to attribute the absorption tail behavior to strong internal fields arising, for example, from ionized dopants or defects.

Amorphous Solids

In a defect-free crystalline semiconductor, a well-defined energy gap exists between the valence and conduction bands. In contrast, in an amorphous semiconductor, the distributions of conduction band and valence band electronic states do not terminate abruptly at the band edges. Instead, some electronic states called the tail states encroach into the gap region [3.33]. In addition to tail states, there are also other localized states deep within the gap region [3.34]. These localized tail states in amorphous semiconductors are contributed by defects. The defects in amorphous semiconductors are considered to be all cases of departure from the normal nearest-neighbor coordination (or normal valence requirement). Examples of defects are: broken and dangling bonds (typical for amorphous silicon); over- and under-coordinated atoms (such as 'valence alternation pairs' in chalcogenide glasses); and voids, pores, cracks and other macroscopic defects. Mobility edges exist, which separate these localized states from their extended counterparts; tail and deep defect states are localized [3.35–37]. These localized tail and deep defect states are responsible for many of the unique properties exhibited by amorphous semiconductors.

Despite years of intensive investigation, the exact form of the distribution of electronic states associated with amorphous semiconductors remains a subject of some debate. While there are still some unresolved theoretical issues, there is general consensus that the tail states arise as a consequence of the disorder present within amorphous networks, and that the width of these tails reflects the amount of disorder present [3.38]. Experimental results (from, for example, [3.39, 40]) suggest exponential distributions for the valence and conduction band tail states in a-Si:H, although other possible functional forms [3.41] cannot be ruled out. *Singh* and *Shimakawa* [3.37] have derived separate effective masses of charge carriers in their extended and tail states. That means the density of states (DOS) of extended and tail states can be represented in two different parabolic forms. The relationship between the absorption coefficient and the distribution of electronic states for the case of a-Si:H may be found in [3.37, 42–44].

The existence of tail states in amorphous solids has a profound impact upon the band-to-band optical absorption. Unlike in a crystalline solid, the absorption of photons in an intrinsic amorphous solid can also occur for photon energies $\hbar\omega \leq E_0$ due to the presence of tail states in the forbidden gap. E_0 is the energy of the optical gap, which is usually close to the mobility gap – the energy difference between the conduction band and valence band mobility edges.

In a crystalline semiconductor, the energy and crystal momentum of an electron involved in an optical transition must be conserved. In an amorphous semiconductor, however, only the energy needs to be conserved [3.36, 37]. As a result, for optical transitions caused by photons of energy $\hbar\omega \geq E_0$ in amorphous semiconductors, the approach of a joint density of states is not applicable [3.37, 45]. One has to consider the product of the densities of both conduction and valence electronic states when calculating the absorption coefficient [3.37, 46]. Assuming that both the valence band and conduction band DOS functions have square-root dependencies on energy, one can derive the absorption coefficient α as [3.37]

$$(\alpha\hbar\omega)^{1/2} = B^{1/2}(\hbar\omega - E_0)\,, \tag{3.40}$$

where, if one assumes that the transition matrix element is constant,

$$B = \frac{1}{nc\varepsilon_0}\left(\frac{e}{m_e^*}\right)^2\left(\frac{L(m_e^* m_h^*)^{3/2}}{2^2\hbar^3}\right)\,, \tag{3.41}$$

where m_e^* and m_h^* are the effective masses of an electron and a hole respectively, L denotes the average interatomic separation in the sample, and n is the refractive index. Plotting $(\alpha\hbar\omega)^{1/2}$ as a function of $\hbar\omega$ yields a straight line that is usually referred to as Tauc's plot, from which one can determine the optical gap energy E_0 experimentally. Equation (3.40) is also known as Tauc's relation and B as *Tauc*'s constant [3.47]. Experimental data from many, but not all, amorphous semiconductors and chalcogenides fit to (3.40) very well. Deviations from Tauc's relation have been observed. For instance, some experimental data in a-Si:H fit much better to a cubic root relation given by *Mott* and *Davis* [3.36]

$$(\alpha\hbar\omega)^{1/3} = C(\hbar\omega - E_0) \tag{3.42}$$

and therefore the cubic root has been used to determine the optical gap E_0. Here C is another constant.

If one considers that the optical transition matrix element is independent of photon energy [3.37, 48], one finds that the cubic root dependence on photon energy can be obtained only when the valence band and conduction band DOS depend linearly on energy. Using such DOS functions, the cubic root dependence has

been explained by *Mott* and *Davis* [3.36]. Another way to obtain the cubic root dependence has been suggested in [3.49]. Using (3.40), *Sokolov* et al. [3.49] have modeled the cubic root dependence on photon energy by considering the fluctuations in the optical bandgap due to structural disorder. Although their approach gives a way of getting the cubic root dependence, Sokolov et al.'s model is hardly different from the linear density of states model suggested by *Mott* and *Davis* [3.36]. *Cody* [3.50] has shown an alternative approach using a photon energy-dependent transition matrix element. Thus, the absorption coefficient is obtained as [3.37]

$$(\alpha\hbar\omega) = B'(\hbar\omega)^2(\hbar\omega - E_0)^2 \,, \tag{3.43}$$

where

$$B' = \frac{e^2}{nc\varepsilon_0}\left(\frac{\left(m_e^* m_h^*\right)^{3/2}}{2\pi^2\hbar^7 \nu\rho_A}\right) Q_a^2 \,, \tag{3.44}$$

and Q_a is the average separation between the excited electron and hole pair in an amorphous semiconductor, ν denotes the number of valence electrons per atom and ρ_A represents the atomic density per unit volume.

Equation (3.43) suggests that $(\alpha\hbar\omega)$ depends on the photon energy in the form of a polynomial of order 4. Then, depending on which term of the polynomial may be more significant in which material, one can get square, cubic or fourth root dependences of $(\alpha\hbar\omega)$ on the photon energy. In this case, (3.43) may be expressed as

$$(\alpha\hbar\omega)^x \propto (\hbar\omega - E_0) \,, \tag{3.45}$$

where $x \leq 1/2$. Thus, in a way, any deviation from the square root or Tauc's plot may be attributed to the energy-dependent matrix element [3.37, 46]. Another possible explanation has been recently discussed by *Shimakawa* and coworkers on the basis of fractal theory [3.51].

Another problem is how to determine the constants B (3.41) and B' (3.44), which involve the effective masses of an electron and a hole. In other words, how do we determine the effective masses of charge carriers in amorphous solids? Recently, a simple approach [3.37, 46] has been developed to calculate the effective masses of charge carriers in amorphous solids. Different effective masses of charge carriers are obtained in the extended and tail states. The approach applies the concepts of tunneling and effective medium, and one obtains the effective mass of an electron in the conduction extended states, denoted by m_{ex}^*, and in the tail states, denoted by m_{et}^*, as

$$m_{ex}^* \approx \frac{E_L}{2(E_2 - E_c)a^{1/3}} m_e \,, \tag{3.46}$$

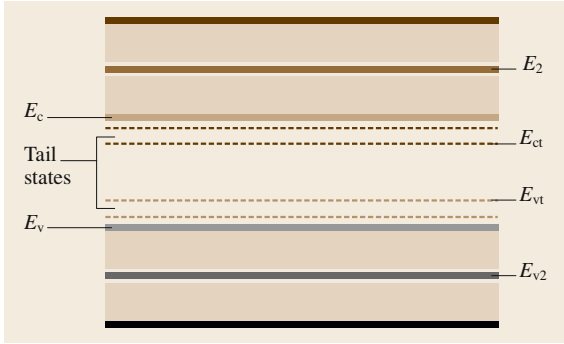


Fig. 3.10 Schematic illustration of the electronic energy states E_2, E_c, E_{ct}, E_{vt}, E_v and E_{v2} in amorphous semiconductors. The *shaded region* represents the extended states. Energies E_2 and E_{v2} correspond to the centers of the conduction and valence extended states and E_{ct} and E_{vt} represent the ends of the conduction and valence tail states respectively

and

$$m_{et}^* \approx \frac{E_L}{(E_c - E_{ct})b^{1/3}} m_e \,, \tag{3.47}$$

where

$$E_L = \frac{\hbar^2}{m_e L^2} \,. \tag{3.48}$$

Here $a = N_1/N < 1$, N_1 is the number of atoms contributing to the extended states, $b = N_2/N < 1$, N_2 is the number of atoms contributing to the tail states such that $a + b = 1$ $(N = N_1 + N_2)$, and m_e is the free-electron mass. The energy E_2 in (3.46) corresponds to the energy of the middle of the extended conduction states, at which the imaginary part of the dielectric constant becomes maximum (Fig. 3.10; see also 3.2).

Likewise, the effective masses of the hole m_{hx}^* and m_{ht}^* in the valence extended and tail states are obtained, respectively, as

$$m_{hx}^* \approx \frac{E_L}{2(E_v - E_{v2})a^{1/3}} m_e \,, \tag{3.49}$$

and

$$m_{ht}^* \approx \frac{E_L}{(E_{vt} - E_v)b^{1/3}} m_e \,, \tag{3.50}$$

where E_{v2} and E_{vt} are energies corresponding to the half-width of the valence extended states and the end of the valence tail states respectively; see Fig. 3.10.

Using (3.46) and (3.47) and the values of the parameters involved, different effective masses of an electron

Table 3.6 Effective mass of electrons in the extended and tail states of a-Si:H and a-Ge:H calculated using (3.46) and (3.47) for $a = 0.99$, $b = 0.01$ and $E_{ct} = E_{vt} = E_c/2$. E_L is calculated from (3.48). All energies are given in eV. Note that since the absorption coefficient is measured in cm^{-1}, the value used for the speed of light is in cm/s ([a] [3.52]; [b] [3.53]; [c] [3.33]; [d] [3.54])

	L (nm)	E_2	E_c	E_L	$E_c - E_{ct}$	m^*_{ex}	m^*_{et}
a-Si:H	0.235[a]	3.6[b]	1.80[c]	1.23	0.9	0.34 m_e	6.3 m_e
a-Ge:H	0.245[a]	3.6	1.05[d]	1.14	0.53	0.22 m_e	10.0 m_e

are obtained in the extended and tail states. Taking, for example, the density of weak bonds contributing to the tail states as 1 at.%, so $b = 0.01$ and $a = 0.99$, the effective masses and energies E_L calculated for hydrogenated amorphous silicon (a-Si:H) and germanium (a-Ge:H) are given in Table 3.6.

According to (3.46), (3.47), (3.49) and (3.50), for sp^3 hybrid amorphous semiconductors such as a-Si:H and a-Ge:H, effective masses of the electron and hole are expected to be the same. In these semiconductors, since the conduction and valence bands are two equal halves of the same electronic band, their widths are the same and that gives equal effective masses for the electron and the hole [3.37, 55]. This is one of the reasons for using $E_{ct} = E_{vt} = E_c/2$, which gives equal effective masses for electrons and holes in the tail states as well. This is different from crystalline solids where m^*_e and m^*_h are usually not the same. This difference between amorphous and crystalline solids is similar to, for example, having direct and indirect crystalline semiconductors but only direct amorphous semiconductors.

Using the effective masses from Table 3.6 and (3.41), B can be calculated for a-Si:H and a-Ge:H. The values thus obtained with the refractive index $n = 4$ for a-Si:H and a-Ge:H are $B = 6.0 \times 10^6\,cm^{-1}\,eV^{-1}$ for a-Si:H and $B = 4.1 \times 10^6\,cm^{-1}\,eV^{-1}$ for a-Ge:H, which are an order of magnitude higher than those estimated from experiments [3.36]. However, considering the quantities involved in B (3.41), this can be regarded as a reasonable agreement.

In a recent paper, *Malik* and *O'Leary* [3.56] studied the distributions of conduction and valence band electronic states associated with a-Si:H. They noted that the effective masses associated with a-Si:H are material parameters that are yet to be experimentally determined. In order to remedy this deficiency, they fitted square-root DOS functions to experimental DOS data and found that $m^*_h = 2.34 m_e$ and $m^*_e = 2.78 m_e$.

The value of the constant B' in (3.44) can also be calculated theoretically, provided that Q_a is known. Using the atomic density of crystalline silicon and four valence electrons per atom, *Cody* [3.50] estimated $Q_a^2 = 0.9\,Å^2$, which gives $Q_a \approx 0.095\,nm$, less than half the interatomic separation of 0.235 nm in a-Si:H, but of the same order of magnitude. Using $\nu = 4$, $\rho_A = 5 \times 10^{28}\,m^{-3}$, $Q_a^2 = 0.9\,Å^2$, and extended state effective masses, we get $B' = 4.6 \times 10^3\,cm^{-1}\,eV^{-3}$ for a-Si:H and $1.3 \times 10^3\,cm^{-1}\,eV^{-3}$ for a-Ge:H. Cody estimated an optical gap, $E_0 = 1.64\,eV$, for a-Si:H, which, using (3.43), gives $\alpha = 1.2 \times 10^3\,cm^{-1}$ at a photon energy of $\hbar\omega = 2\,eV$. This agrees reasonably well with the $\alpha = 6.0 \times 10^2\,cm^{-1}$ used by Cody. If we use interatomic spacing in place of Q_a in (3.44), we get $B' = 2.8 \times 10^4\,cm^{-1}\,eV^{-3}$, and then the corresponding absorption coefficient becomes $3.3 \times 10^3\,cm^{-1}$. This suggests that, in order to get an estimate, one may use the interatomic spacing in place of Q_a if the latter is unknown. Thus, both B and B' can be determined theoretically, a task not possible before due to a lack of knowledge of the effective masses in amorphous semiconductors.

The absorption of photons of energy less than the bandgap energy $\hbar\omega < E_0$ in amorphous solids involves the localized tail states and hence follows neither (3.40) nor (3.42). Instead, the absorption coefficient depends on the photon energy exponentially, as given in (3.39), giving rise to Urbach's tail. *Abe* and *Toyozawa* [3.57] have calculated the interband absorption spectra in crystalline solids, introducing the Gaussian site diagonal disorder and applying the coherent potential approximation. They have shown that Urbach's tail occurs due to static disorder (structural disorder). However, the current stage of understanding is that Urbach's tail in amorphous solids occurs due to both thermal effects and static disorder [3.50]. More recent issues in this area have been addressed by *Orapunt* and *O'Leary* [3.48].

Keeping the above discussion in mind, the optical absorption in amorphous semiconductors near the absorption edge is usually characterized by three types of optical transitions corresponding to transitions between tail and tail states, tail and extended states, and extended and extended states. The first two types correspond to $\hbar\omega \leq E_0$, and the third one corresponds to $\hbar\omega \geq E_0$. Thus, the plot of absorption coefficient versus photon energy (α versus $\hbar\omega$) has three different regions, A, B and C respectively, that correspond to these three characteristic optical transitions shown in Fig. 3.11.

In the small absorption coefficient range A (also called the weak absorption tail (WAT), where $\alpha < 10^{-1}\,cm^{-1}$, the absorption is controlled by optical transitions from tail-to-tail states. As stated above, the

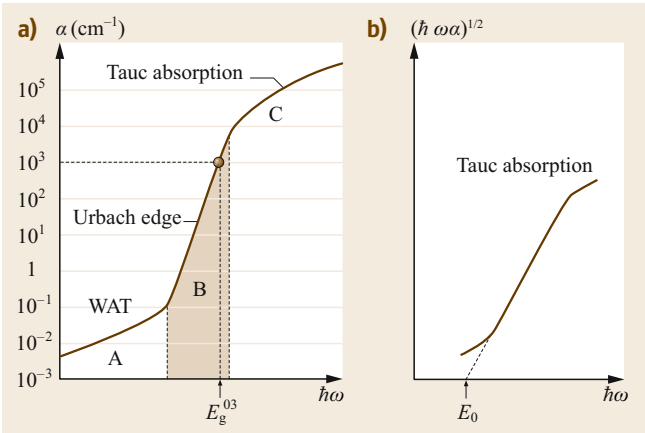


Fig. 3.11a,b Typical spectral dependence of the optical absorption coefficient in amorphous semiconductors. **(a)** In the A and B regions, the absorption is controlled by optical transitions between tail and tail states and tail and extended states respectively, and in the C region it is dominated by transitions from extended to extended states. In domain B, the absorption coefficient follows the Urbach rule (3.39). In region C, the absorption coefficient follows Tauc's relation (3.40) in a-Si:H, as shown in **(b)**

localized tail states in amorphous semiconductors are contributed to by defects. To some extent, the absolute value of absorption in region A may be used to estimate the density of defects in the material. In region B, where typically $10^{-1} < \alpha < 10^{4}\,\mathrm{cm}^{-1}$, the absorption is related to transitions from the localized tail states above the valence band edge to extended states in the conduction band and/or from extended states in the valence band to localized tail states below the conduction band. The spectral dependence of α usually follows the so-called Urbach rule, given in (3.39). For many amorphous semiconductors, ΔE has been related to the width of the valence (or conduction) band tail states, and may be used to compare the *widths* of such localized tail states in different materials; ΔE is typically 0.05–0.1 eV. In region C, the absorption is controlled by transitions from extended to extended states. For many amorphous semiconductors, the α versus $\hbar\omega$ behavior follows the Tauc relation given in (3.40). The optical bandgap, E_0, determined for a given material from the α versus $\hbar\omega$ relations obtained in (3.40), (3.42) and (3.43), can vary as shown in Table 3.7 for a-Si:H alloys.

3.3.4 Exciton Absorption

Excitons in Crystalline Semiconductors

Optical absorption in crystalline semiconductors and insulators can create an exciton, which is an electron and hole pair excited by a photon and bound together through their attractive Coulomb interaction. This means that the absorbed optical energy remains held within the solid for the lifetime of the exciton. There are two types of excitons that can be formed in nonmetallic solids: Wannier or Wannier–Mott excitons and Frenkel excitons. The concept of Wannier–Mott excitons is valid for inorganic semiconductors such as Si, Ge and GaAs, because in these materials the large overlap of interatomic electronic wavefunctions enables the electrons and holes to be far apart but bound in an excitonic state. For this reason, these excitons are also called large-radii orbital excitons. Excitons formed in organic crystals are called Frenkel excitons. In organic semiconductors/insulators or molecular crystals, the intermolecular separation is large and hence the overlap of intermolecular electronic wavefunctions is very small and electrons remain tightly bound to individual molecules. Therefore, the electronic energy bands are very narrow and closely related to individual molecular electronic energy levels. In these solids, the absorption of photons occurs close to the individual molecular electronic states and excitons are also formed within the molecular energy levels [3.59]. Such excitons are therefore also called molecular excitons. For a more detailed look at the theory of Wannier and Frenkel excitons, readers may like to refer to *Singh* [3.59].

In Wannier–Mott excitons, the Coulomb interaction between the hole and electron can be viewed as an effective hydrogen atom with, for example, the hole establishing the coordinate reference frame about which the reduced mass electron moves. If the effective masses of the isolated electron and hole are m_e^* and m_h^* respectively, their reduced mass μ_x is given by

$$\mu_x^{-1} = \left(m_e^*\right)^{-1} + \left(m_h^*\right)^{-1} . \tag{3.51}$$

Note that in the case of so-called hydrogenic impurities in semiconductors (both shallow donor and acceptor impurities), the reduced mass of the nucleus takes the place of one of the terms in (3.51) and hence the re-

Table 3.7 Optical bandgaps of a-$Si_{1-x}C_x$:H films obtained from Tauc's (3.40), Sokolov et al.'s (3.42) and Cody's (3.43) relations. (After [3.58])

	E_g at $\alpha = 10^3\,\mathrm{cm}^{-1}$	E_g at $\alpha = 10^4\,\mathrm{cm}^{-1}$	$E_g = E_0$ (Tauc)	$E_g = E_0$ (Cody)	$E_g = E_0$ (Sokolov)	ΔE (meV)
a-Si:H	1.76	1.96	1.73	1.68	1.60	46
a-$Si_{0.88}C_{0.18}$:H	2.02	2.27	2.07	2.03	1.86	89

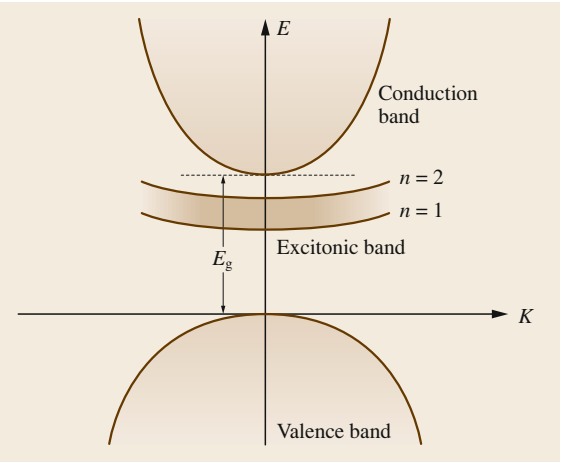


Fig. 3.12 Schematic illustration of excitonic bands for $n = 1$ and 2 in semiconductors. E_g represents the energy gap

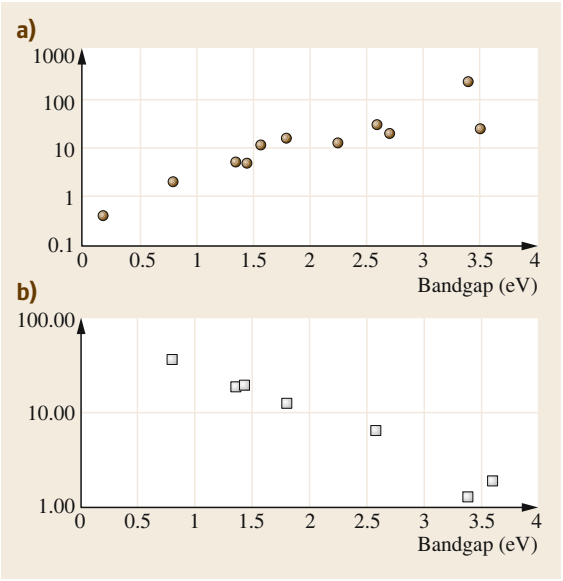


Fig. 3.13a,b Dependence of the **(a)** exciton binding energy (R_y^x) (3.29) and **(b)** size (in terms of the ratio of the excitonic Bohr radius to lattice constant $[a_B^*/a_0]$) as a function of the semiconductor bandgap. Exciton binding energy increases along with a marked drop in exciton spreading as bandgap increases. The Wannier-based description is not appropriate above a bandgap of about 2 eV

duced mass is given to good approximation by the effective mass of the appropriate carrier. When the exciton is the carrier, the effective masses are comparable and hence the reduced mass is markedly lower; accordingly, the exciton binding energy is markedly lower than that for hydrogenic impurities. The energy of a Wannier–Mott exciton is given by (e.g., [3.59])

$$E_x = E_g + \frac{\hbar^2 K^2}{2M} - E_n \,, \tag{3.52}$$

where E_g is the bandgap energy, $\hbar K$ is the linear momentum, M $(= m_e^* + m_h^*)$ is the effective mass associated with the center of mass of an exciton, and E_n is the exciton binding energy given by

$$E_n = \frac{\mu_x e^4 \kappa^2}{2\hbar^2 \varepsilon^2} \frac{1}{n^2} = \frac{\mathrm{Ry}^x}{n^2} \,, \tag{3.53}$$

where e is the electronic charge, $\kappa = 1/(4\pi\varepsilon_0)$, ε is the static dielectric constant of the solid, and n is the principal quantum number associated with the internal excitonic states $n = 1(s), 2(p), \ldots$ (Fig. 3.12). According to (3.53), the excitonic states are formed within the bandgap below the conduction band edge. However, as the exciton binding energy is very small (a few meV in bulk Si and Ge crystals), exciton absorption peaks can only be observed at very low temperatures. Ry^x is the so-called effective Rydberg of the exciton given by $\mathrm{Ry}(\mu_x/m_e)/\varepsilon^2$, where $\mathrm{Ry} = 13.6\,\mathrm{eV}$. For bulk GaAs, the exciton binding energy ($n = 1$) corresponds to about 5 meV. Extrapolating from the hydrogen atom model, the extension of the excitonic wavefunction can be found from an effective Bohr radius a_B^* given in terms of the Bohr radius as $a_B = h^2/(\pi e^2)$; that is, $a_B^* = a_B(\varepsilon/\mu_x)$. For GaAs this corresponds to about 12 nm or about 21 lattice constants – in other words the spherical volume of the exciton radius contains $\approx (a_B^*/a_0)^3$ or ≈ 9000 unit cells, where a_0 is the lattice constant of GaAs. As $\mathrm{Ry}^* \ll E_g$ and $a_B^* \gg a_0$, it is clear from this example that the excitons in GaAs are large-radii orbital excitons, as stated above. It should be noted that the binding energies of excitons in semiconductors tend to be a strong function of the bandgap. The dependences of Ry^x (exciton binding energy) and a_B^*/a_0 on the semiconductor bandgap are shown in Fig. 3.13a and b respectively. For excitons with large binding energies and correspondingly small radii (approaching the size of a lattice parameter), the excitons become localized on a lattice site, as observed in most organic semiconductors. As stated above, such excitons are commonly referred to as Frenkel excitons or molecular excitons. Unlike Wannier-based excitons, which are typically dissociated at room temperature, these excitations are stable at room temperature. For the binding energies of Frenkel excitons, one can refer to *Singh* for example [3.59].

Excitons can recombine radiatively, emitting a series of hydrogen-like spectral lines, as described by (3.52). In bulk (3-D) semiconductors such as GaAs, ex-

citon lines can only be observed at low temperatures: they are easily dissociated by thermal fluctuations. On the other hand, in quantum wells and other structures of reduced dimensionality, the spatial confinement of both the electron and hole wavefunctions in the same layer ensures strong excitonic transitions of a few meV below the bandgap, even at room temperature. Excitonic absorption is well located spectrally and very sensitive to optical saturation. For this reason, it plays an important role in nonlinear semiconductor devices (nonlinear Fabry–Perot resonators, nonlinear mirrors, saturable absorbers, and so on). If the valence hole is a heavy hole, the exciton is called a heavy exciton; conversely, if the valence hole is light, the exciton is a light exciton. For practical purposes, the excitonic contribution to the overall susceptibility around the resonance frequency ν_{ex} can be written as

$$\chi_{exc} = -A_0 \frac{(\nu - \nu_{ex}) + i\Gamma_{ex}}{(\nu - \nu_{ex})^2 + \Gamma_{ex}^2(1+S)} , \tag{3.54}$$

where Γ_{ex} is the linewidth and $S = I/I_S$ the saturation parameter of the transition. For instance, in GaAs MQW, the saturation intensity I_S is as low as $1\,kW/cm^2$, and Γ_{ex} ($\approx 3.55\,meV$ at room temperature) varies with the temperature according to

$$\Gamma_{ex} = \Gamma_0 + \frac{\Gamma_1}{\exp(\hbar\omega_{LO}/kT) - 1} , \tag{3.55}$$

where $\hbar\Gamma_0$ is the inhomogeneous broadening ($\approx 2\,meV$), $\hbar\Gamma_1$ is the homogeneous broadening ($\approx 5\,meV$), and $\hbar\omega_{LO}$ is the longitudinal optical phonon energy ($\approx 36\,meV$). At high carrier concentrations (provided either by electrical pumping or by optical injection), the screening of the Coulomb attractive potential by free electrons and holes provides an efficient mechanism for saturating the excitonic line.

The above discussion refers to so-called free excitons formed between conduction-band electrons and valence-band holes. According to (3.52), such an excitation is able to move throughout a material with a given center-of-mass kinetic energy (second term on the right-hand side). It should be noted, however, that free electrons and holes move with a velocity $\hbar(dE/dk)$ where the derivative is taken for the appropriate band edge. To move through a crystal, both the electron and the hole must have identical translational velocities, restricting the regions in k-space where these excitations can occur to those with $(dE/dk)_{electron} = (dE/dk)_{hole}$, commonly referred to as critical points.

A number of more complex pairings of carriers can also occur, which may also include fixed charges or ions. For example, for the case of three charged entities with one being an ionized donor impurity (D+), the following possibilities can occur: (D+)(+)(−), (D+)(−)(−) and (+)(+)(−) as excitonic ions, and (+)(+)(−)(−) and (D+)(+)(−)(−) as biexcitons or even bigger excitonic molecules [3.60]. Complexity abounds in these systems, as each electronic level possesses a fine structure corresponding to allowed rotational and vibrational levels. Moreover, the effective mass is often anisotropic. Note that when the exciton or exciton complex is bound to a fixed charge, such as an ionized donor or acceptor center in the material, the exciton or exciton complex is referred to as a bound exciton. Indeed, bound excitons may also involve neutral fixed impurities. It is usual to relate the exciton in these cases to the center binding them; thus, if an exciton is bound to a donor impurity, it is usually termed a donor-bound exciton.

Excitons in Amorphous Semiconductors

The concept of excitons is traditionally valid only for crystalline solids. However, several observations in the photoluminescence spectra of amorphous semiconductors have revealed the occurrence of photoluminescence associated with singlet and triplet excitons [3.37]. Applying the effective mass approach, a theory for the Wannier–Mott excitons in amorphous semiconductors has recently been developed in real coordinate space [3.37, 46, 55, 61]. The energy of an exciton thus derived is obtained as

$$W_x = E_0 + \frac{P^2}{2M} - E_n(S) , \tag{3.56}$$

where P is the linear momentum associated with the exciton's center of motion and $E_n(S)$ is the binding energy of the exciton, given by

$$E_n(S) = \frac{\mu_x e^4 \kappa^2}{2\hbar^2 \varepsilon'(S)^2 n^2} , \tag{3.57}$$

where

$$\varepsilon'(S) = \varepsilon \left[1 - \frac{(1-S)}{A}\right]^{-1} , \tag{3.58}$$

where S is the spin (S being $= 0$ for singlet and $= 1$ for triplet) of an exciton and A is a material-dependent constant representing the ratio of the magnitude of the Coulomb and exchange interactions between the electron and the hole of an exciton. Equation (3.57) is analogous to (3.53) obtained for excitons in crystalline solids for $S = 1$. This is because (3.53) is derived within the large-radii orbital approximation, which neglects the exchange interaction and hence is valid only for triplet excitons [3.59, 62]. As amorphous solids lack

long-range order, the exciton binding energy is found to be larger in amorphous solids than in their crystalline counterparts; for example, the binding energy is higher in hydrogenated amorphous silicon (a-Si:H) than in crystalline silicon (c-Si). This is the reason that it is possible to observe both singlet and triplet excitons in a-Si:H [3.63] but not in c-Si.

Excitonic Absorption

Since exciton states lie below the edge of the conduction band in a crystalline solid, absorption to excitonic states is observed below this edge. According to (3.53), the difference in energy in the bandgap and the excitonic absorption gives the binding energy. As the exciton-photon interaction operator and excited electron and hole pair and photon interaction operator depend only on their relative motion, then these interactions take the same form for band-to-band and excitonic absorption. Therefore, to calculate the excitonic absorption coefficient, one can use the same form of interaction as that used for band-to-band absorption, but one must use the joint density of states. Using the joint density of states, the absorption coefficient associated with the excitonic states in crystalline semiconductors is obtained as [3.37]

$$\alpha\hbar\omega = A_x(\hbar\omega - E_x)^{1/2} \tag{3.59}$$

with the constant

$$A_x = \frac{4\sqrt{2}\,e^2|p_{xv}|^2}{nc\sqrt{\mu_x}\,\hbar^2}\ ,$$

where p_{xv} is the transition matrix element between the valence and excitonic bands. Equation (3.59) is similar to that seen for direct band-to-band transitions, discussed above (3.60), and is only valid for the photon energies $\hbar\omega \geq E_x$. There is no absorption below the excitonic ground state in pure crystalline solids. Absorption of photons to excitonic energy levels is possible through either the excitation of electrons to higher energy levels in the conduction band and then nonradiative relaxation to the excitonic energy level, or through the excitation of an electron directly to the exciton energy level. Excitonic absorption occurs in both direct and indirect semiconductors.

In amorphous semiconductors, the excitonic absorption and photoluminescence can be quite complicated. According to (3.56), the excitonic energy level is below the optical bandgap by an energy equal to the binding energy given in (3.57). However, there are four transition possibilities:

1. Extended valence to extended conduction states
2. Valence to extended conduction states
3. Valence extended to conduction tail states and
4. Valence tail to conduction tail states.

These possibilities will have different optical gap energies, E_0, and different binding energies. Transition 1 will give rise to absorption in the free exciton states, transitions 2 and 3 will give absorption in the bound exciton states, because one of the charge carriers is localized in the tail states, and absorption through transition 4 will create localized excitons, which are also called geminate pairs. This can be visualized as follows: if an electron-hole pair is excited by a high-energy photon through transition 1 and forms an exciton, initially its excitonic energy level and the corresponding Bohr radius will have a reduced mass corresponding to both charge carriers being in extended states. As such an exciton relaxes downward nonradiatively, its binding energy and excitonic Bohr radius will change because its effective mass changes in the tail states. When both charge carriers reach the tail states (transition 4), their excitonic Bohr radius will be maintained although they are localized.

The excitonic absorption coefficient in amorphous semiconductors can be calculated using the same approach as presented in Sect. 3.3.3, and similar expressions to (3.40) and (3.43) are obtained. This is because the concept of the joint density of states is not applicable in amorphous solids. Therefore, by replacing the effective masses of the charge carriers by the excitonic reduced mass and the distance between the excited electron and hole by the excitonic Bohr radius, one can use (3.40) and (3.43) to calculate the excitonic absorption coefficients for the four possible transitions above in amorphous semiconductors. However, such a detailed calculation of the excitonic transitions in amorphous semiconductors is yet to be performed.

3.3.5 Impurity Absorption

Impurity absorption can be observed as the absorption coefficient peaks lying below fundamental (band-to-band) and excitonic absorption. It is usually related to the presence of ionized impurities or, simply, ions. The peaks occur due to electronic transitions between ionic electronic states and the conduction/valence band or due to intra-ionic transitions (within d or f shells, between s and d shells, and so on). The first case leads to intense and broad lines, while the characteristics of the features arising from the latter case depend on whether or not these transitions are allowed by parity selection rules. For allowed transitions, the absorption peaks are quite intense and broad, while forbidden transitions produce weak and narrow peaks. General reviews of this topic may be found in *Blasse* and *Grabmaier* [3.64], *Henderson* and *Imbusch* [3.65] and *DiBartolo* [3.66].

In the following section, we concentrate primarily on the properties of rare earth ions, which are of great importance in modern optoelectronics.

Optical Absorption of Trivalent Rare Earth Ions: Judd–Ofelt Analysis

Rare earths (REs) is the common name used for the elements from Lanthanum (La) to Lutetium (Lu). They have atomic numbers of 57 to 71 and form a separate group in the periodic table. The most notable feature of these elements is an incompletely filled 4f shell. The electronic configurations of REs are listed in Table 3.8. The RE may be embedded in different host materials in the form of divalent or trivalent ions. As divalent ions, REs exhibit broad absorption-emission lines related to allowed 4f→5d transitions. In trivalent form, REs lose two 6 s electrons and one 4f or 5d electron. The Coulomb interaction of a 4f electron with a positively charged core means that the 4f level gets split into a complicated set of manifolds with energies, to a first approximation, that are virtually independent of the host matrix because the 4f level is well screened from external influences by the 5s and 5p shells [3.67]. Figure 3.14 shows an energy level diagram for the low-lying $4f^N$ states of the trivalent ions embedded in $LaCl_3$. To a second approximation, the exact construction and precise energies of the manifolds depend on the host material, via crystal field and via covalent interactions with the ligands surrounding the RE ion. A ligand is an atom (or molecule or radical or ion) with one or more unshared pairs of electrons that can attach to a central metallic ion (or atom) to form a coordination complex. Examples of ligands include ions (F^-, Cl^-, Br^-, I^-, S^{2-}, CN^-, NCS^-, OH^-, NH_2^-) and molecules (NH_3, H_2O, NO, CO) that donate a pair of electrons to a metal atom or ion. Some ligands that share electrons with metals form very stable complexes.

Optical transitions between 4f manifold levels are forbidden by a parity selection rule that states that the wavefunctions of the initial and final states of an atomic (ionic) transition must have different parities for them to be permitted. Parity is a property of any function (or quantum mechanical state) that describes the function after mirror reflection. Even functions (states) are symmetric (identical after reflection, for example a cosine function), while odd functions (states) are antisymmetric (for example a sine function). The parity selection rule may be partially removed for an ion (or atom) embedded in host material due to the action of the crystal field, which gives rise to *forbidden lines*. The crystal field is the electric field created by a host material at the position of the ion.

The parity selection rule is weakened by the admixture of 5d states with 4f states and by the disturbed RE ion symmetry due to the influence of the host, which increases with the covalency. Higher covalency implies stronger sharing of electrons between the RE ion and the ligands. This effect is known as the nephelauxetic effect. The resulting absorption-emission lines are characteristic of individual RE ions and quite narrow because they are related to forbidden inner shell 4f transitions.

Judd–Ofelt (JO) analysis allows the oscillator strength of an electric dipole (ED) transition between two states of a trivalent rare earth (RE) ion embedded in a particular host lattice to be calculated. The possible states of an RE ion are often referred to as $^{2S+1}L_J$, where $L = 0, 1, 2, 3, 4, 5, 6 \ldots$ determines the electron's total angular momentum, and is conventionally represented by the letters S, P, D, F, G, I. The term $(2S + 1)$ is called the spin multiplicity and represents the number of spin configurations, while J is the total angular momentum, which is the vector sum of the overall (total) angular momentum and the overall spin ($J = L + S$). The value $(2J + 1)$ is called the multiplicity and corresponds to the number of possible combinations of overall angular momentum and overall spin that yield the same J. Thus, the notation $^4I_{15/2}$ for the ground state of Er^{3+} corresponds to the term $(J, L, S) = (15/2, 6, 3/2)$, which has a multiplicity of $2J + 1 = 16$ and a spin multiplicity of $2S + 1 = 4$. If the wavefunctions $|\psi_i\rangle$ and $|\psi_f\rangle$ correspond to the initial ($^{2S+1}L_J$) and final ($^{2S'+1}L_{J'}$) states of an electric dipole transition of an RE ion, the line strength of this transition, according to JO theory, can be calculated

Table 3.8 Occupation of outer electronic shells for rare earth elements

57	**La**	$4s^2$	$4p^2$	$4d^{10}$	–	$5s^2$	$5p^6$	**$5d^1$**	$6s^2$
58	**Ce**	$4s^2$	$4p^2$	$4d^{10}$	**$4f^1$**	$5s^2$	$5p^6$	**$5d^1$**	$6s^2$
59	**Pr**	$4s^2$	$4p^2$	$4d^{10}$	**$4f^3$**	$5s^2$	$5p^6$	–	$6s^2$
60	**Nd**	$4s^2$	$4p^2$	$4d^{10}$	**$4f^4$**	$5s^2$	$5p^6$	–	$6s^2$
…									
68	**Er**	$4s^2$	$4p^2$	$4d^{10}$	**$4f^{12}$**	$5s^2$	$5p^6$	–	$6s^2$
…									
70	**Yb**	$4s^2$	$4p^2$	$4d^{10}$	**$4f^{14}$**	$5s^2$	$5p^6$	–	$6s^2$
71	**Lu**	$4s^2$	$4p^2$	$4d^{10}$	**$4f^{14}$**	$5s^2$	$5p^6$	**$5d^1$**	$6s^2$

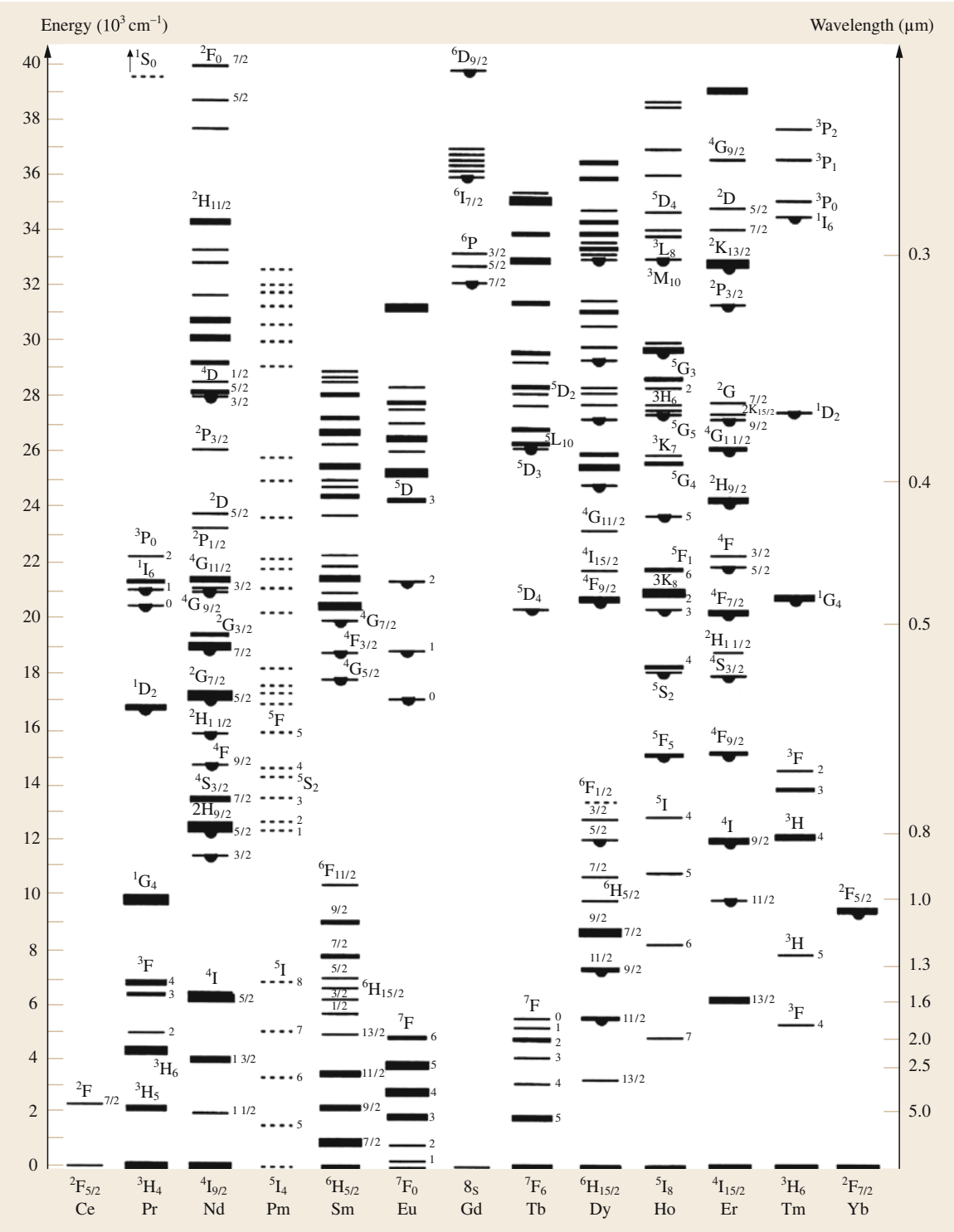


Fig. 3.14 Energy level diagram of the low-lying $4f^N$ states of trivalent ions doped in $LaCl_3$. The pendant semicircles indicate fluorescent levels. (After [3.65, 68, 69])

using

$$S_{\mathrm{ed}} = \left|\langle\psi_f\left|H_{\mathrm{ed}}\right|\psi_i\rangle\right|^2 = \sum_{k=2,4,6}\Omega_k\left|\left\langle f_\nu^N S'L'J'\left|U^{(k)}\right|f_\nu^N SLJ\right\rangle\right|^2 , \tag{3.60}$$

where H_{ed} is the ED interaction Hamiltonian, Ω_k are coefficients reflecting the influence of the host material, and $U^{(k)}$ are reduced tensor operator components, which are virtually independent of the host material, and their values are calculated using the so-called intermediate coupling approximation [3.70]. The theoretical values of S_{ed} calculated from this are compared with the values derived from experimental data using

$$S_{\mathrm{exp}} = \frac{3hcn}{8\pi^3 e^2\langle\lambda\rangle}\frac{2J+1}{\chi_{\mathrm{ed}}}\int_{\mathrm{band}}\frac{\alpha(\lambda)}{\rho}\mathrm{d}\lambda , \tag{3.61}$$

where $\langle\lambda\rangle$ is the mean wavelength of the transition, h is the Plank constant, c is the speed of light, e is the elementary electronic charge, $\alpha(\lambda)$ is the absorption coefficient, ρ is the RE ion concentration, n is the refraction index and the factor $\chi_{\mathrm{ed}} = (n^2+2)^2/9$ is the so-called local field correction. The key idea of JO analysis is to minimize the discrepancy between experimental and calculated values of line strength, S_{ed} and S_{exp}, by choosing the coefficients Ω_k, which are used to characterize and compare materials, appropriately. The complete analysis should also include the magnetic dipole transitions [3.71]. The value of Ω_2 is of prime importance because it is the most sensitive to the local structure and material composition and is correlated with the degree of covalence. The values of Ω_k are used to calculate radiative transition probabilities and appropriate radiative lifetimes of excited states, which are very useful for numerous optical applications. More detailed analysis may be found in, for example, [3.71]. Ω_k values for different ions and host materials can be found in *Gschneidner Jr.* and *Eyring* [3.72].

3.3.6 Effects of External Fields

Electroabsorption and the Franz–Keldysh Effect

Electroabsorption is the absorption of light in a device where the absorption is induced by an applied (or changing) electric field within the device. Such a device is an electroabsorption modulator. There are three fundamental types of electroabsorption processes. In the Franz–Keldysh process, a strong applied field modifies the photon-assisted probability of an electron tunneling from the valence band to the conduction band, and thus it corresponds to an effective reduction in the *bandgap energy*, inducing the absorption of light with photon energies of slightly less than the bandgap. It was first observed for CdS, in which the absorption edge was observed to shift to lower energies with the applied field; that is, photon absorption shifts to longer wavelengths with the applied field. The effect is normally quite small but is nonetheless observable. In this type of electroabsorption modulation, the wavelength is typically chosen to be slightly smaller than the bandgap wavelength so that absorption is negligible. When a field is applied, the absorption is enhanced by the Franz–Keldysh effect. In free-carrier absorption, the concentration of free carriers N in a given band is changed (modulated), for example, by an applied voltage, changing the extent of photon absorption. The absorption coefficient is proportional to N and to the wavelength λ of the light raised to some power, typically 2–3. In the confined Stark effect, the applied electric field modifies the energy levels in a quantum well. The energy levels are reduced by the field by an amount proportional to the square of the applied field. A multiple quantum well (MQW) pin-type device has MQWs in its intrinsic layer. Without any applied bias, light with photon energy just less than the quantum well (QW) exciton excitation energy will not be significantly absorbed. When a field is applied, the energy levels are lowered and the incident photon energy is then sufficient to excite an electron and hole pair in the QWs. The relative transmission decreases with the reverse bias V_{r} applied to the pin device. Such MQW (pin) devices are usually not very useful in the transmission mode because the substrate material often absorbs the light (for example a GaAs/AlGaAs MQW pin would be grown on a GaAs substrate, which would absorb the radiation that excites the QWs). Thus, a reflector would be needed to reflect the light back before it reaches the substrate; such devices have indeed been fabricated.

The Faraday Effect

The Faraday effect, originally observed by Michael Faraday in 1845, is the rotation of the plane of polarization of a light wave as it propagates through a medium subjected to a magnetic field parallel to the direction of propagation of the light. When an optically inactive material such as glass is placed in a strong magnetic field and plane-polarized light is sent along the direction of the magnetic field, the emerging light's plane of polarization is rotated. The magnetic field can be applied, for example, by inserting the material into the core of a magnetic coil – a solenoid. The specific rotatory power induced, given by θ/L, has been found to be proportional to the magnitude of the applied magnetic

Table 3.9 Verdet constants for some materials

Material	Quartz	Flint glass	Tb-Ga garnet	Tb-Ga garnet	ZnSe	Crown glass	NaCl
λ (nm)	589.3	632	632	1064	633	589.3	589.3
ϑ (rad m^{-1} T^{-1})	4.0	4.0	−134	−40	118	6.4	10

field $\boldsymbol{B}$, which gives the amount of rotation as

$$\theta = \vartheta \boldsymbol{B} L \,, \tag{3.62}$$

where L is the length of the medium, and ϑ is the so-called Verdet constant, which depends on the material and the wavelength of the light. The Faraday effect is typically small. For example, a magnetic field of $\approx$ 0.1 T causes a rotation of about 1° through a glass rod of length 20 mm. It appears that *optical activity* is induced by the application of a strong magnetic field to an otherwise optically inactive material. There is, however, an important distinction between natural optical activity and the Faraday effect. The sense of rotation θ observed in the Faraday effect for a given material (Verdet constant) depends only on the direction of the magnetic field $\boldsymbol{B}$. If ϑ is positive, for light propagating parallel to $\boldsymbol{B}$, the optical field $\boldsymbol{E}$ rotates in the same sense as an advancing right-handed screw pointing in the direction of $\boldsymbol{B}$. The direction of light propagation does not change the absolute sense of rotation of θ. If we reflect the wave to pass through the medium again, the rotation increases to 2θ. The Verdet constant depends not only on the wavelength λ but also on the charge-to-mass ratio of the electron and the refractive index $n(\lambda)$ of the medium through

$$\vartheta = -\frac{(e/m_\mathrm{e})}{2c} \lambda \frac{\mathrm{d}n}{\mathrm{d}\lambda} \,. \tag{3.63}$$

Verdet constants for some materials are listed in Table 3.9. Terbium gallium garnet is commercially used in optical isolators.

3.4 Optical Characterization

Various methods have been developed for the determination of optical constants of solids. They differ with respect to sample geometry (bulks, thin films, multilayers, powders etc.) and optical properties (absorbing and nonabsorbing). In this section we are mainly concerned with bulk samples and thin films with flat surfaces. From available experimental techniques we select those based on two fundamental properties of light (used as a probe) – its intensity and polarization. Generally, both the light intensity and polarization change upon light interaction with the sample and if measured in suitable configurations they can be used for the determination of the material refractive index and the extinction coefficient. Optical transmittance, reflectance and ellipsometry parameters are convenient measurable quantities.

Normal incidence optical transmittance is defined as a ratio of light intensity of transmitted and incident waves, $T = I^\mathrm{t}/I^\mathrm{i}$. Reflectance is usually measured for oblique incidence and therefore depends on light polarization and angle of incidence. It is convenient to define optical reflectance for two linear polarizations: p represents the electric intensity vector oscillations in parallel direction with respect to the plane of incidence and s represents the electric intensity vector oscillations in perpendicular direction with respect to the plane of incidence. Reflectance is then defined by the relation $R_\mathrm{p,s} = I^\mathrm{r}_\mathrm{p,s}/I^\mathrm{i}_\mathrm{p,s}$ as the ratio of reflected and incident beam intensities for specific polarizations. Experimentally, easily accessible transmittance and reflectance can be related to transmission and reflection coefficients of a sample

$$T = |t|^2 \,, \qquad t = \frac{\hat{E}^\mathrm{t}}{\hat{E}^\mathrm{i}} \tag{3.64}$$

$$R_\mathrm{p,s} = |r_\mathrm{p,s}|^2 \,, \qquad r_\mathrm{p,s} = \frac{\hat{E}^\mathrm{r}_\mathrm{p,s}}{\hat{E}^\mathrm{i}_\mathrm{p,s}} \,, \tag{3.65}$$

where amplitudes of electric fields of incident, reflected and transmitted waves are complex amplitudes $\hat{E} = E_0 \mathrm{e}^{\mathrm{i}\phi}$ of a plane electromagnetic wave

$$E = E_0 \mathrm{e}^{\mathrm{i}\phi} \mathrm{e}^{\mathrm{i}(\omega t - kr)} = \hat{E} \mathrm{e}^{\mathrm{i}(\omega t - kr)} \,. \tag{3.66}$$

All discussed cases of transmission and reflection are presented in Figs. 3.15–3.18 together with explicit expressions of incident, reflected and transmitted waves. It should be noted here that transmission and reflection coefficients are generally complex numbers providing information not only on the amplitude change but also on the phase change of the wave. On the other hand, transmittance and reflectance are real quantities that do not carry phase change information as they measure the ratio of light intensities

$$I = \frac{1}{2} c \varepsilon_0 E_0^2 \,. \tag{3.67}$$

An elegant way to get access to a phase change is by measuring the polarization state of the reflected or

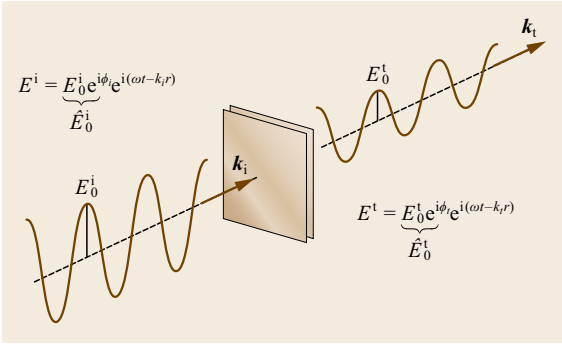


Fig. 3.15 Incident and transmitted waves in the case of normal incidence transmittance measurement

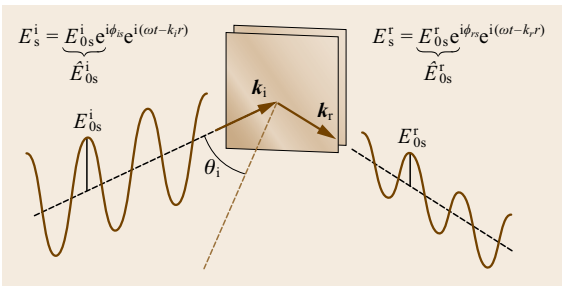


Fig. 3.16 Incident and reflected s-polarized waves in the case of oblique incidence reflectance measurement

transmitted wave with respect to the polarization state of the incident wave. The change of polarization is often represented by two angles Ψ, Δ that are, in the case of reflection, defined by the relation

$$\tan\Psi \mathrm{e}^{\mathrm{i}\Delta} = \frac{r_{\mathrm{p}}}{r_{\mathrm{s}}} \,. \tag{3.68}$$

The polarization state of a reflected wave – that is analyzed – generally takes an elliptical form, therefore this experimental method is called *ellipsometry*. The geometrical meaning of angles Ψ, Δ follows from their definition. Consider a linearly polarized incident wave with equal contributions (equal amplitudes) of p- and s-components that are in phase (Fig. 3.18). In the reflected wave, due to different reflection coefficients r_{p} and r_{s}, the p- and s- components are generally mutually shifted by an angle difference $\phi_{\mathrm{rp}} - \phi_{\mathrm{rs}}$ that equals the ellipsometric angle Δ. Furthermore, the amplitudes of p- and s- components in a reflected wave are no more equal and their ratio $E_{0\mathrm{p}}^{\mathrm{r}}/E_{0\mathrm{s}}^{\mathrm{r}}$ equals $\tan\Psi$. This interpretation is consistent with following mathematical treatment of relation (3.68)

$$\frac{r_{\mathrm{p}}}{r_{\mathrm{s}}} = \frac{E_{0\mathrm{p}}^{\mathrm{r}} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{rp}}} / E_{0\mathrm{p}}^{\mathrm{i}} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{ip}}}}{E_{0\mathrm{s}}^{\mathrm{r}} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{rs}}} / E_{0\mathrm{s}}^{\mathrm{i}} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{is}}}} = \frac{E_{0\mathrm{p}}^{\mathrm{r}}}{E_{0\mathrm{s}}^{\mathrm{r}}} \mathrm{e}^{\mathrm{i}(\phi_{\mathrm{rp}} - \phi_{\mathrm{rs}})} = \tan\Psi \mathrm{e}^{\mathrm{i}\Delta} \,. \tag{3.69}$$

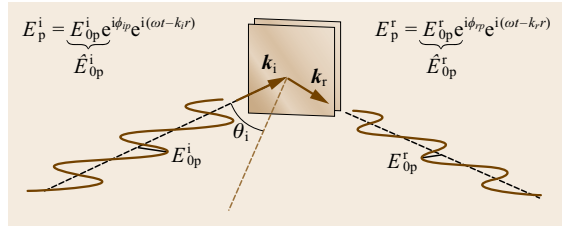


Fig. 3.17 Incident and reflected p-polarized waves in the case of oblique incidence reflectance measurement

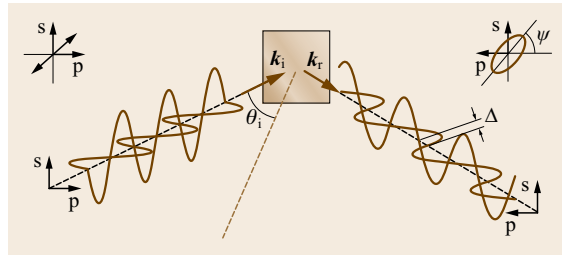


Fig. 3.18 Linearly polarized incident wave with equal contributions of s- and p-components reflected from the sample surface. The geometrical meaning of ellipsometric angles Ψ and Δ is indicated

For selected sample geometries, transmission and reflection coefficients can be expressed analytically in the frame of phenomenological electromagnetic theory [3.73]. Optical constants are explicitly covered in these expressions and therefore the measurable quantities of optical reflectance, transmittance and ellipsometry angles can be used for determination of the material refractive index and the extinction coefficient. This will be discussed in more detail for the case of single interface (bulk) and thin film materials in the following sections.

A large number of books and articles cover instrumental aspects of reflectance and transmittance measurements in different spectral ranges. For a broad overview of this field the reader is referred to [3.74] and references therein. Instrumentation of spectroscopic ellipsometers is discussed in detail by monographs devoted to spectroscopic ellipsometry [3.75–77]. Historically, the first ellipsometers that were developed were null ellipsometers. In this configuration, the orientation of the polarizer and compensator are adjusted such that the light reflected from the sample is linearly polarized. The analyzer is then rotated to the position that the light intensity on the detector is extinguished or *nulled*. From the positions of the polarizer, compensator and analyzer, the ellipsometric angles Ψ and Δ can be calculated. This configuration, if performed in four equivalent zones, can be very accurate and with low systematic errors. However, even when automated, this

approach is relatively slow and measurements are time consuming. In order to speed up measurements, rotating analyzer/polarizer ellipsometers were developed. In these systems, either the analyzer or polarizer is continuously rotated at a constant angular velocity (typically about 10–100 Hz) about the optical axis. The operating characteristics of both of these configurations are similar. However, the rotating polarizer system requires the light source to be totally unpolarized. Any residual polarization in the source results in a source of measurement error unless corrected. Similarly, a rotating analyzer system is susceptible to the polarization sensitivity of the detector. However, solid-state semiconductor photodetectors have extremely high polarization sensitivities. Thus, commercial systems tend to use rotating analyzer systems where residual polarization in the source is not an issue. Fourier analysis of the variation of the amplitude of the detector signal provides values for Ψ and Δ. Such systems can provide high-speed and accurate measurements. Spectroscopic ellipsometers extend the concepts developed for measurements at a single wavelength to measurements at multiple wavelengths. Being able to measure the dispersion in optical constants with wavelength adds another dimension to the analysis, permitting unambiguous determination of material and structure parameters.

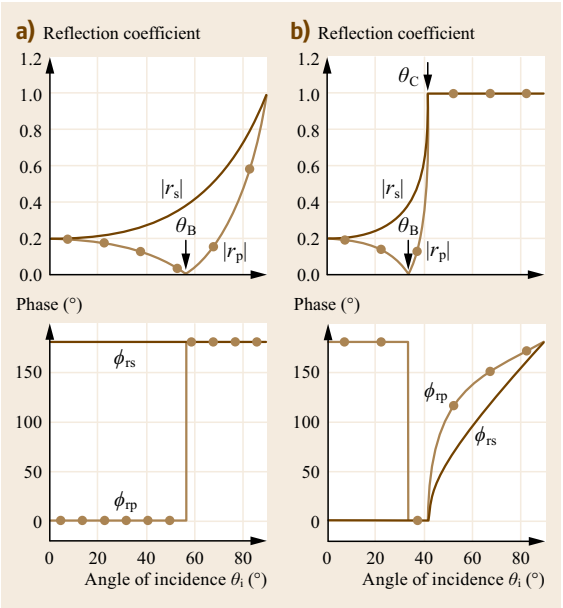


Fig. 3.19a,b Incident angle dependence of single interface reflection coefficient (its absolute value and phase) for a p- and s-polarized wave. Fresnel relations with $n_i = 1$; $n_t = 1.5$ **(a)** and $n_i = 1.5$; $n_t = 1$ **(b)** were used. Brewster θ_B and critical θ_C angles are also indicated

3.4.1 Bulk Samples

The term *bulk* is used here for samples with (at least locally) flat surfaces and for samples that are made from:

1. Absorbing materials thick enough with respect to penetration depth of light – used as a probe – that it senses only the sample surface (and not the backside of the sample or its deeper interfaces).
2. Nonabsorbing materials thick enough with respect to coherence length of the light – used as a probe – or shaped in such a geometry that light interference in the sample can be neglected.

For optical characterization of samples of type 1, reflection configuration is the most convenient, whereas for samples of type 2, refraction (possibly combined with reflection) is usually used.

First we will consider the determination of optical constants of absorbing bulk materials. In this case light senses only the sample surface, hence reflection coefficients for p- and s-polarized waves get the form of Fresnel equations derived for single interface

$$r_p = \frac{N_t \cos\theta_i - N_i \cos\theta_t}{N_t \cos\theta_i + N_i \cos\theta_t}\,,$$
$$r_s = \frac{N_i \cos\theta_i - N_t \cos\theta_t}{N_i \cos\theta_i + N_t \cos\theta_t}\,, \qquad (3.70)$$

where N_i and N_t are (complex) refractive indices of the ambient and sample respectively [3.73]. It is useful to shortly discuss the dependence of reflection coefficients on the incidence angle θ_i (angle of refraction is denoted by θ_t and can be calculated by Snell's law). Figure 3.19 represents incident angle dependence of absolute values of p- and s-reflection coefficients and their phases. Two cases are distinguished: (a) the ambient is optically rarer; and (b) the ambient is optically denser with respect to the sample. The angle of incidence where the r_p coefficient becomes zero is called the *Brewster angle* θ_B. However, for optically absorbing samples, the r_p coefficient does not reach a zero value. Instead, its angle dependence forms a nonzero minimum there and its position is called the *principal angle*. When the ambient is optically denser then the critical angle θ_c can be distinguished. For incidence angles greater than the critical, both reflection coefficients are equal in absolute value to unity. Total reflection is observed under these conditions. Calculations of reflectance and ellipsometry parameters for a specific angle of incidence is straightforward using Fresnel equations (3.70) and definition relations (3.65) and (3.69).

However, precaution must be used when approximating a surface of a bulk material with a single interface. A single interface is an ideal structure that presents an abrupt and sharp change of refractive index

across a planar interface between two semi-infinitive spaces – in our case, the ambient and sample. In practice it is rather difficult to prepare such surfaces. Mechanical polishing of a sample surface usually causes a so-called *Beilby overlayer* that accumulates various defects due to the mechanism of polishing [3.78]. This damaged layer can be reduced in thickness if a proper polishing process is selected or removed by electropolishing. The Beilby overlayer may present different optical properties than the bulk and, if not carefully considered, erroneous optical constants are determined. Alternatively, sample cutting or cleaving in a vacuum or the deposition of an optically thick film is possible. Ellipsometry as a phase sensitive method is extremely sensitive to the surface quality; and it is made use of for ultrathin film or overlayer characterization. The following sections cover selected approaches used to derive optical constants from reflectance and ellipsometry measurements carried out on sample surfaces that can be successfully approximated by the single interface model. For further use we provide here also the Fresnel formulae of transmission coefficients for a single interface

$$t_\mathrm{p} = \frac{2N_\mathrm{i}\cos\theta_\mathrm{i}}{N_\mathrm{t}\cos\theta_\mathrm{i} + N_\mathrm{i}\cos\theta_\mathrm{t}} ,$$
$$t_\mathrm{s} = \frac{2N_\mathrm{i}\cos\theta_\mathrm{i}}{N_\mathrm{i}\cos\theta_\mathrm{i} + N_\mathrm{t}\cos\theta_\mathrm{t}} . \qquad (3.71)$$

Normal Incidence Reflectivity and Kramers–Kronig Analyses

For normal incidence, p- and s-polarization cannot be distinguished. The reflection coefficient is then polarization-independent and if air is considered as an ambient we get

$$r = \frac{1-n+\mathrm{i}K}{1+n-\mathrm{i}K} . \qquad (3.72)$$

Hence, determination of the refractive index n and extinction coefficient K of the studied material requires knowledge of both the absolute value and phase of the normal incidence reflection coefficient

$$r = |r|\,\mathrm{e}^{\mathrm{i}\phi} = \sqrt{R}\mathrm{e}^{\mathrm{i}\phi} . \qquad (3.73)$$

Experimentally accessible normal incidence reflectivity R provides only the absolute value of the reflection coefficient. However when reflectivity is known in the whole spectral range then the phase of the reflection coefficient can be calculated with the help of the Kramers–Kronig relation [3.79]

$$\phi(\omega) = -\frac{2\omega}{\pi} P \int_0^\infty \frac{\ln\sqrt{R(\omega')}}{\omega'^2-\omega^2}\mathrm{d}\omega' . \qquad (3.74)$$

The final relations for the refractive index and extinction coefficient of the probed material are obtained from

$$n = \frac{1-R}{1+R+2\sqrt{R}\cos\phi} ,$$
$$K = \frac{2\sqrt{R}\sin\phi}{1+R+2\sqrt{R}\cos\phi} . \qquad (3.75)$$

However, in reality, reflectivity cannot be measured in the *whole* spectral range covering all frequencies. Nevertheless, the Kramers–Kronig relation (3.74) can still be used when the extrapolation of the experimentally available spectrum of reflectivity is carefully done in both spectral limits. Often this method combines visible (VIS), ultraviolet (UV) and vacuum ultraviolet (VUV) spectrophotometry or synchrotron facility to cover as wide a spectral range as possible. For the high energy of photons, the surface quality (mainly surface roughness) becomes extremely important; for nonnegligible roughness, a correction on light scattering lost in the reflected beam must be considered [3.80]. As an example of an application of the Kramers–Kronig approach we provide a normal incidence reflectivity spectrum recorded on PZT and the corresponding calculated spectra of the PZT refractive index and extinction coefficient in Fig. 3.20.

Oblique Incidence Reflectivity R_p and R_s

Even if a wide spectral range is not experimentally available for a normal incidence reflectivity measurement, we can still make use of reflectance measurements – carried out for oblique incidence – and determine optical constants of material from these spectra. Fresnel relations (3.70) inserted into definitions (3.65) make a link between reflectance for p- and s-waves and optical constants of the material

$$R_\mathrm{s} = \frac{(Y-u)^2+v^2}{(Y+u)^2+v^2} ,$$
$$R_\mathrm{p} = \frac{(\varepsilon_1-Zu)^2+(\varepsilon_2-Zv)^2}{(\varepsilon_1+Zu)^2+(\varepsilon_2+Zv)^2} , \qquad (3.76)$$

where

$$Y = n_\mathrm{i}\cos\theta_\mathrm{i} , \quad Z = \frac{n_\mathrm{i}}{\cos\theta_\mathrm{i}} ,$$
$$\varepsilon_1 = n_\mathrm{t}^2 - K_\mathrm{t}^2 , \quad \varepsilon_2 = 2n_\mathrm{t}K_\mathrm{t} , \quad \varepsilon_3 = (n_\mathrm{i}\cos\theta_\mathrm{i})^2 ,$$
$$2u^2 = \varepsilon_1 - \varepsilon_3 + \sqrt{(\varepsilon_1-\varepsilon_3)^2+\varepsilon_2^2} ,$$
$$2v^2 = \varepsilon_3 - \varepsilon_1 + \sqrt{(\varepsilon_1-\varepsilon_3)^2+\varepsilon_2^2} .$$

These equations are highly nonlinear and moreover transcendental, therefore only approximate inverse solutions exist [3.81]. Presently, numerical methods and

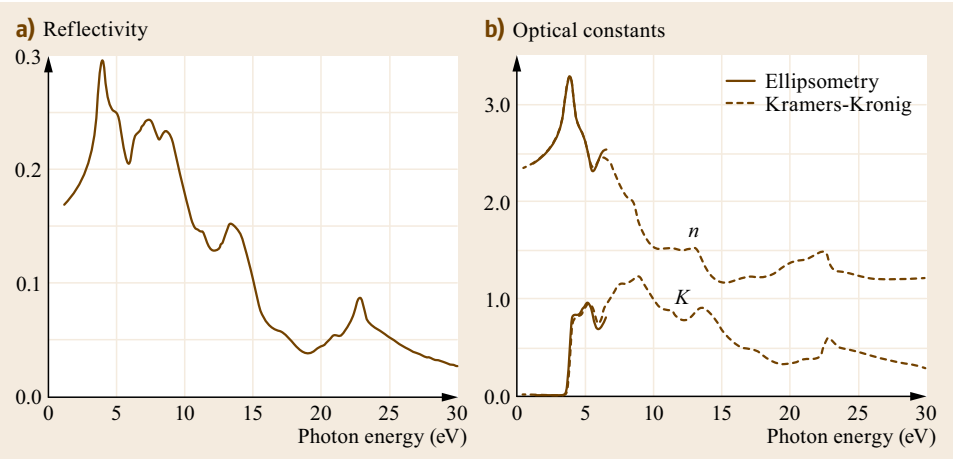


Fig. 3.20a,b Normal incidence reflectivity of PZT **(a)** and its complex refractive index **(b)** determined by Kramers–Kronig relations. Experimental data recorded on a synchrotron facility UVSOR, Japan. Comparison with ellipsometry is also presented

nonlinear fitting of experimental spectra of R_s and R_p can be routinely performed to adjust searched optical constants.

Ellipsometry

The determination of optical constants of a material from ellipsometry parameters Ψ and Δ recorded on a bulk sample is straightforward. It can be shown [3.82] that the complex refractive index of a material can be expressed as

$$n - \mathrm{i}K = \sqrt{\sin^2\theta_\mathrm{i}\left(1+\tan^2\theta_\mathrm{i}\left(\frac{1-\rho}{1+\rho}\right)^2\right)}\,, \quad (3.77)$$

where ρ represents the ratio of Fresnel reflection coefficients for p- and s-polarized waves; $\rho = r_p/r_s$ and the ambient is assumed to be air. Since the angle of incidence θ_i is set in the experiment, the two parameters measured from the experiment (Ψ and Δ) for a given wavelength can be used to deduce the complex parameter ρ and further, two remaining unknown variables in the equation above – namely, n and K. Alternatively, we can reformulate relation (3.77) and express the real and imaginary part of the material dielectric permittivity as an explicit function of ellipsometry parameters

$$\varepsilon_1 = n^2 - K^2 = \sin^2\theta_\mathrm{i}\left[1+\left(\tan^2\theta_\mathrm{i} \times \frac{\cos^2(2\Psi)-\sin^2(2\Psi)\sin^2\Delta}{[1+\sin(2\Psi)\cos\Delta]^2}\right)\right],$$

$$\varepsilon_2 = 2nK = \frac{\sin^2\theta_\mathrm{i}\tan^2\theta_\mathrm{i}\sin(4\Psi)\sin\Delta}{[1+\sin(2\Psi)\cos\Delta]^2}\,. \quad (3.78)$$

The single interface is the only case where the inverse solution is analytically available. For other structures such as thin films, multilayers, and so on, we have to use numerical methods to adjust searched optical and geometrical parameters of a sample. Even if the incidence angle θ_i appears explicitly in relation (3.77), the value of the complex refractive index does not depend on its selection. However, the most precise results are obtained when θ_i is set close to the principal value (or Brewster angle) where $r_p \approx 0$ and r_s and r_p coefficients have significantly different values (Fig. 3.19). Figure 3.21 represents an example of Ψ and Δ measured for incidence angles 55° and 65° on the surface of a pellet pressed from blue organic pigment and the corresponding optical constants determined by relation (3.77). For both incidence angles, the obtained optical constants are nearly identical. On the other hand, if determined n and K are incidence angle-dependent then the approximation of a single interface is no longer valid. In this case, the sample surface may have a more complex structure (for example the presence of an overlayer).

Alternatively, the searched spectral dependence of a complex refractive index may be parameterized by an appropriate function (model dielectric function). Its parameters may be then determined by nonlinear fitting of experimental data Ψ and Δ. For this purpose, ellipsometric spectra for several incidence angles are usually recorded and fitted simultaneously to increase the number of experimental data and hence increase the reliability of the results. Several examples of model dielectric function are provided above in Sect. 3.2. Other examples for specific materials and spectral ranges are available for example in monographs focused on spectroscopic ellipsometry [3.75, 76]. Figure 3.22 presents a model dielectric function of polycrystalline ZnSe defined as a sum of five oscillators whose parameters were fitted against experimental ellipsometric spectra recorded for several incidence angles.

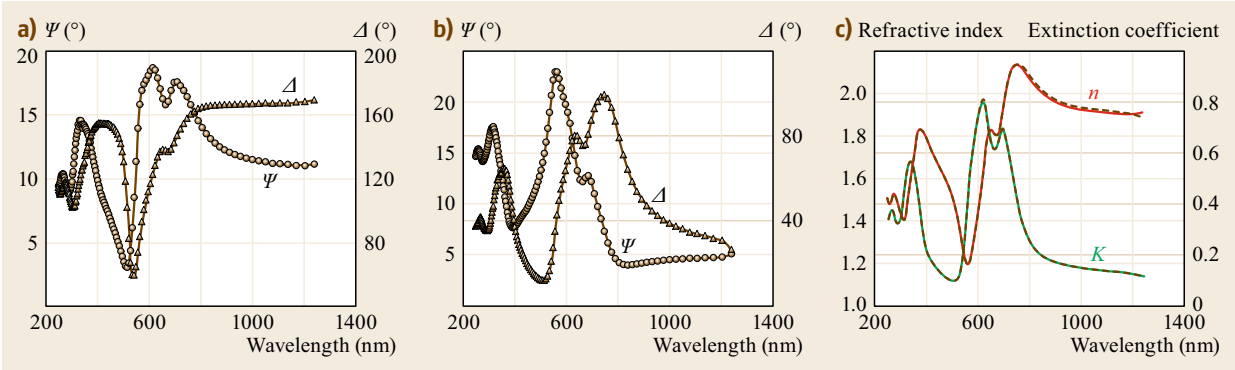


Fig. 3.21a–c Ellipsometric parameters recorded on pressed pellets of blue organic pigment for incidence angle of 55° **(a)** and 65° **(b)** together with refractive index **(c)** determined by analytical formula (3.70)

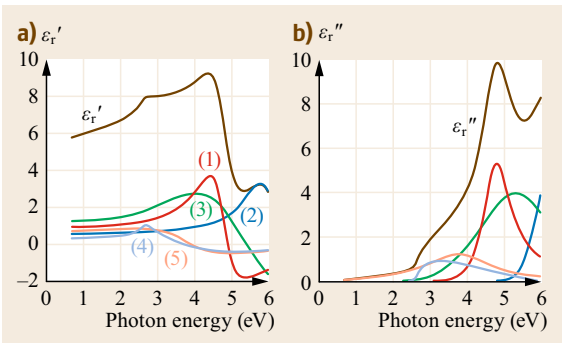


Fig. 3.22a,b Real **(a)** and imaginary **(b)** part of dielectric permittivity of ZnSe determined by spectroscopic ellipsometry. Its deconvolution to five oscillators is also indicated

Ellipsometry is a very precise and sensitive spectroscopic tool but when operated solely in reflection configuration then its sensitivity to low values of material absorption is limited. This is due to the fact that reflected light senses only the sample surface and its close vicinity. It is argued that ellipsometry can provide reliable results for absorption coefficient $\alpha > 10^5\,\mathrm{cm}^{-1}$ [3.83]. If α is in the range 10^3–$10^5\,\mathrm{cm}^{-1}$ usually transmittance spectrum is added and treated simultaneously with ellipsometry data to ensure accurate determination of the absorption coefficient. In the region of low absorption $\alpha < 10^3\,\mathrm{cm}^{-1}$ other techniques such as photocurrent measurement or photothermal deflection spectroscopy are more appropriate [3.84].

Combined Measurement of *T* and *R*

The combined measurement of reflectance and transmittance also enables material absorption determination. This is often done for the study of absorption edge onset of amorphous semiconductors (Fig. 3.11) and glasses where the absorption coefficient is relatively weak. The determination of the absorption coefficient in a spectral range where the penetration depth decreases as the absorption increases is usually realized by measurements of a series of samples with various thicknesses. Samples are usually in the form of plane-parallel plates (both sides polished to defined thickness), hand blown glass bubbles or films deposited on substrates. The absorption coefficient is then expressed by following relation

$$\alpha = \frac{1}{d}\ln\left(\frac{(1-R)^2 + \sqrt{(1-R)^4 + 4R^2T^2}}{2T}\right), \quad (3.79)$$

where d stands for sample thickness.

There is a wide class of materials that in some limited spectral range are nonabsorbing and their optical transparency is extensively used in various applications (optical elements, integrated optics, etc.). In this case, the refractive index has a real value, because its imaginary part (extinction) can be neglected. Of course, the refractive index of these nonabsorbing materials can be determined by most methods mentioned above that were based on light reflection but much more precise are methods based on light refraction, i. e., refractometry.

Minimal Deviation

The refractive indices of dielectrics and semiconductors in a spectral range of their transparency can be determined from a prism of the material of an apex angle, A, by measuring the angle of minimal deviation, D, of radiation passing through the prism as in Fig. 3.23. The refractive index, n, is then given by the relation

$$n = \frac{\sin\frac{1}{2}(A+D)}{\sin\frac{1}{2}A}\,. \quad (3.80)$$

It should be noted here that high quality samples (prisms) free of any volume imperfections and defects

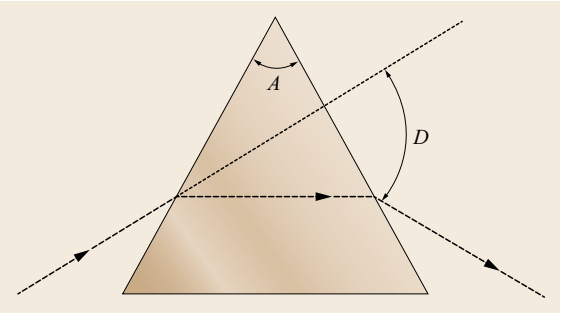


Fig. 3.23 Angle of deviation technique for measuring the refactive index

are required, otherwise scattered radiation in the transmitted beam does not enable a precise determination of the angle of minimal deviation. This technique can give a refraction index in absolute precision of about 0.001.

To complete this section we present one more convenient method for determination of refractive index of nonabsorbing materials based on the critical angle measurement.

Critical Angle

Consider a bulk material with an unknown refractive index, optically coupled to the prism with a known index of refraction (Fig. 3.24). If the angle of incidence is scanned, then an abrupt change of intensity is detected when the angle of incidence reaches the critical angle (onset of total reflection on prism/sample interface) (Fig. 3.19). From the condition of critical angle,

$$n = n_{\mathrm{p}} \sin\theta_{\mathrm{c}} \tag{3.81}$$

we can determine the refractive index of the bulk sample with a high precision of about 0.0005. The condition of total reflection requires that the refractive index of the sample is higher than that of the prism. This technique can also be used for refractive index and thickness determination of transparent films deposited on a substrate. In this case, for some discrete values of incidence angle, the light is coupled as a guided wave to the film. From these angle values, film parameters of interest can be determined and the method itself is referred to as a prism coupler method.

3.4.2 Thin Film Optics

Thin film optics involves multiple reflections of light entering a thin film dielectric (typically on a substrate) so that the reflection and transmission coefficients are determined by multiple wave interference phenomena, as shown in Fig. 3.25.

Consider a thin film coated on a substrate. Assuming that the incident wave has an amplitude of E_0, then

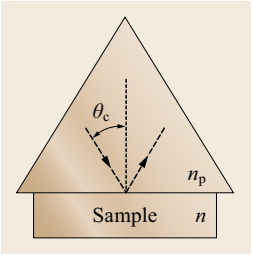


Fig. 3.24 Experimental setup for measuring the refractive index by the critical angle technique

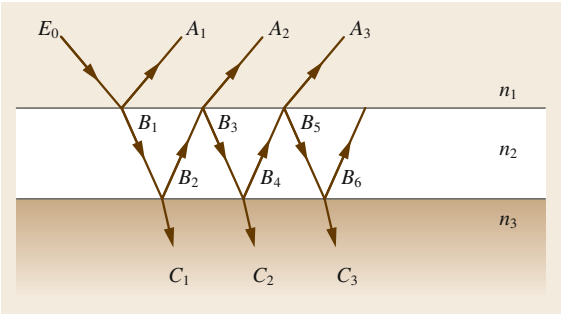


Fig. 3.25 Thin film coated on a substrate and multiple reflections of incident light, where n_1, n_2 and n_3 are the refractive indices of the medium above the thin film, the thin film, and the substrate respectively

there are various transmitted and reflected waves, as shown in Fig. 3.25. We then have the following amplitudes based on the definitions of the reflection and transmission coefficients

$$\begin{aligned}
A_1 &= E_0 r_{12} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{A1}}} \\
A_2 &= E_0 t_{12} r_{23} t_{21} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{A2}}} \\
A_3 &= E_0 t_{12} r_{23} r_{21} r_{23} t_{21} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{A3}}} \\
B_1 &= E_0 t_{12} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{B1}}} \\
B_2 &= E_0 t_{12} r_{23} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{B2}}} \\
B_3 &= E_0 t_{12} r_{23} r_{21} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{B3}}} \\
B_4 &= E_0 t_{12} r_{23} r_{21} r_{23} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{B4}}} \\
B_5 &= E_0 t_{12} r_{23} r_{21} r_{23} r_{21} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{B5}}} \\
B_6 &= E_0 t_{12} r_{23} r_{21} r_{23} r_{21} r_{23} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{B6}}} \\
C_1 &= E_0 t_{12} t_{23} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{C1}}} \\
C_2 &= E_0 t_{12} r_{23} r_{21} t_{23} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{C2}}} \\
C_3 &= E_0 t_{12} r_{23} r_{21} r_{23} r_{21} t_{23} \mathrm{e}^{\mathrm{i}\phi_{\mathrm{C3}}}
\end{aligned} \tag{3.82}$$

and so on, where r_{12} is the reflection coefficient of a wave in medium 1 incident on medium 2, and t_{12} is the transmission coefficient from medium 1 into 2. For simplicity, we will assume normal incidence. The phase change upon traversing the thin film thickness d is $\phi_{\mathrm{C1}} = -(2\pi/\lambda) n_2 d$, where λ is the free space wave-

length. In the following relations the phase ϕ is defined as $\phi = -\phi_{C1}$.

The reflection and transmission coefficients under normal incidence are given by

$$r_1 = r_{12} = \frac{n_1 - n_2}{n_1 + n_2} = -r_{21} \,,$$
$$r_2 = r_{23} = \frac{n_2 - n_3}{n_2 + n_3} \,, \tag{3.83a}$$

and

$$t_1 = t_{12} = \frac{2n_1}{n_1 + n_2} \,, \quad t_2 = t_{21} = \frac{2n_2}{n_1 + n_2} \,,$$
$$t_3 = t_{23} = \frac{2n_3}{n_2 + n_3} \,, \tag{3.83b}$$

where

$$1 - t_1 t_2 = r_1^2 \,. \tag{3.84}$$

The total reflection coefficient is then obtained as

$$r = \frac{A_{\text{reflected}}}{E_0} = r_1 - \frac{t_1 t_2}{r_1} \sum_{k=1}^{\infty} \left(-r_1 r_2 \mathrm{e}^{-\mathrm{i}2\phi}\right)^k \,, \tag{3.85}$$

which can be summed to be

$$r = \frac{r_1 + r_2 \mathrm{e}^{-\mathrm{i}2\phi}}{1 + r_1 r_2 \mathrm{e}^{-\mathrm{i}2\phi}} \,. \tag{3.86}$$

The total transmission coefficient is obtained as

$$t = \frac{C_{\text{transmitted}}}{E_0} = -\frac{t_1 t_3 \mathrm{e}^{\mathrm{i}\phi}}{r_1 r_2} \sum_{k=1}^{\infty} \left(-r_1 r_2 \mathrm{e}^{-\mathrm{i}2\phi}\right)^k$$
$$= \left(\frac{t_1 t_3 \mathrm{e}^{\mathrm{i}\phi}}{r_1 r_2}\right) \frac{r_1 r_2 \mathrm{e}^{-\mathrm{i}2\phi}}{1 + r_1 r_2 \mathrm{e}^{-\mathrm{i}2\phi}} \,, \tag{3.87}$$

which can be summed to be

$$t = \frac{t_1 t_3 \mathrm{e}^{-\mathrm{i}\phi}}{1 + r_1 r_2 \mathrm{e}^{-\mathrm{i}2\phi}} \,. \tag{3.88}$$

Equations (3.86) and (3.88) are very useful when studying the optical properties of thin films coated on a substrate as discussed by *Kasap*, *Ray* and *Gould* in Chap. 28 on Thin Films in this Handbook. Reflection and transmission coefficients of thin film systems for oblique incidence are also expressed by (3.86) and (3.88) respectively. However, single interface reflection and transmission coefficients (r_1, r_2, t_1, t_3) have to be replaced by Fresnel relations for s- and p-polarized waves (Eqs. (3.70) and (3.71)). Moreover, phase change has to be modified to $\phi = (2\pi/\lambda) d (n_2^2 - n_1^2 \sin^2\theta)^{1/2}$, where θ is the incidence angle, to account for oblique propagation of light in the layer. Reflectance of the thin film system is equal to the square of the absolute value of the reflection coefficient of the thin film (3.65). On the other hand, when considering transmittance of the film deposited on a substrate, it must be remembered that expression (3.88) relates the amplitudes of the waves in the media bounding the film (ambient and substrate), but transmitted wave intensity is not measured in the substrate but in free space with a refractive index n_1. Therefore, normal incidence transmittance is related to the transmission coefficient by relation

$$T = \frac{n_3}{n_1} |t|^2 \,. \tag{3.89}$$

In practice, the two most popular approaches for thin film optical characterization are either to analyze optical transmission spectra, which may be observed using a standard spectrophotometer, or ellipsometric investigations in reflection. Both approaches are briefly explained below.

Swanepoel's Analysis of Optical Transmission Spectra

One of the simplest and most practically realizable approaches to the problem was developed by Swanepoel [3.85]. He showed that for a uniform film, with a thickness d, a refractive index n and an absorption coefficient α, deposited on the substrate with a refractive index s, the transmittance can be expressed as

$$T(\lambda) = \frac{Ax}{B - Cx\cos\varphi + Dx^2} \,, \tag{3.90a}$$

where

$$A = 16n^2 s \,, \tag{3.90b}$$
$$B = (n+1)^3 \left(n + s^2\right) \,, \tag{3.90c}$$
$$C = 2\left(n^2 - 1\right)\left(n^2 - s^2\right) \,, \tag{3.90d}$$
$$D = (n-1)^3 \left(n - s^2\right) \,, \tag{3.90e}$$
$$\varphi = 4\pi n d/\lambda \,, \tag{3.90f}$$
$$x = \exp(-\alpha d) \,, \tag{3.90g}$$

and n is a function of λ.

For a nonuniform film with a wedge-like cross-section, (3.90a) must be integrated over the thickness

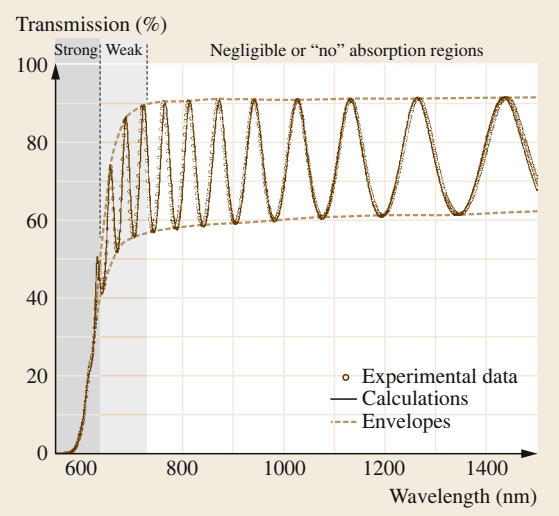


Fig. 3.26 Optical transmission of an a-Se thin film. Calculations are done using (3.90) with the $n(\lambda)$ and $\alpha(\lambda)$ relations shown in Fig. 3.27. The film was prepared by the thermal evaporation of photoreceptor-grade selenium pellets. Film thickness was 2 μm. Tentative regions of strong, weak and negligible absorption are also shown

of the film, giving

$$T_{\Delta d}(\lambda) = \frac{1}{\varphi_2 - \varphi_1} \int_{\varphi_1}^{\varphi_2} \frac{Ax}{B - Cx\cos\varphi + Dx^2} \mathrm{d}x \,, \quad (3.91)$$

where

$$\varphi_1 = 4\pi n\,(d - \Delta d)\,/\lambda$$

and

$$\varphi_2 = 4\pi n\,(d + \Delta d)\,/\lambda \,, \quad (3.92)$$

where d is the average thickness of the film and Δd is the variation of the thickness throughout the illumination (testing) area. The concept behind Swanepoel's method is to construct two envelopes $T_{\mathrm{M}}(\lambda)$ and $T_{\mathrm{m}}(\lambda)$ that pass through the maxima and minima of $T(\lambda)$ and to split the entire spectral range into three regions: negligible absorption, weak absorption and strong absorption regions, as shown in Fig. 3.26.

In the region of negligible absorption, (3.91) yields

$$T_{\mathrm{M/m}} = \frac{\lambda}{2\pi n \Delta d} \frac{a}{\sqrt{1-b^2}} \tan^{-1}\left[\frac{1 \pm b}{\sqrt{1-b^2}} \tan\left(\frac{2\pi n \Delta d}{\lambda}\right)\right] , \quad (3.93)$$

where $a = A/(B+D)$ and $b = C/(B+D)$. The *plus* sign in (3.93) corresponds to transmission maxima and

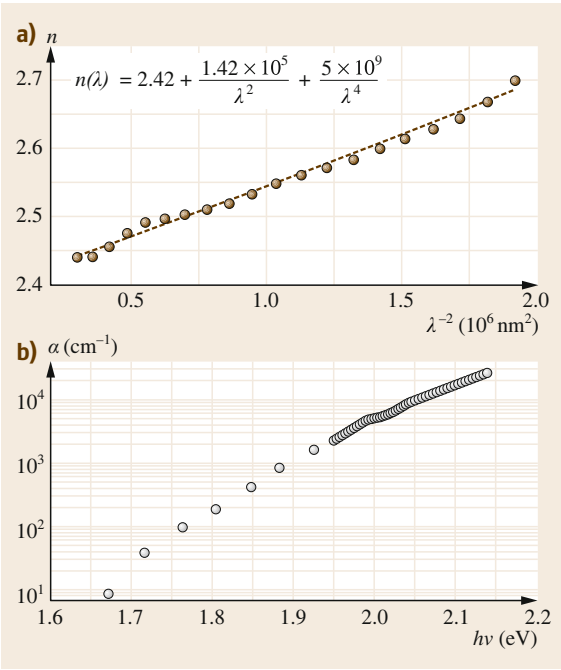


Fig. 3.27 **(a)** The spectral dependence of the refractive index of the a-Se thin film from Fig. 3.26. The *line* corresponds to the Cauchy approximation with the parameters shown on the figure. **(b)** The spectral dependence of the absorption coefficient of the same a-Se thin film

minus to minima. Equation (3.74) is used to find Δd and n.

In the region of weak absorption,

$$T_{\mathrm{M/m}} = \frac{\lambda}{2\pi n \Delta d} \frac{a_x}{\sqrt{1-b_x^2}} \tan^{-1}\left[\frac{1 \pm b_x}{\sqrt{1-b_x^2}} \tan\left(\frac{2\pi n \Delta d}{\lambda}\right)\right] , \quad (3.94)$$

where $a_x = Ax/(B+Dx^2)$ and $b_x = Cx/(B+Dx^2)$. Equation (3.94) allows us to find x and n using the previously found Δd.

The previous results are used to construct $n(\lambda)$ and to create approximations using the Cauchy, Sellmeier or Wemple–DiDominico dispersion equations. An example of a Cauchy approximation is shown in Fig. 3.27.

The positions of the extrema are given by the equation

$$2nd = m\lambda \,, \quad (3.95)$$

where $m = 1, 2, 3, \ldots$ for maxima and $m = 1/2, 3/2, \ldots$ for minima. Therefore, d can be found from the slope of (n/λ) versus m.

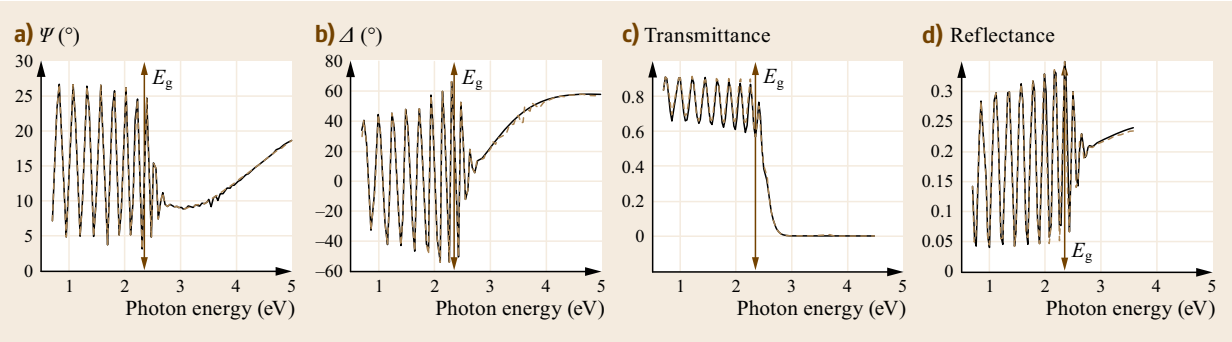


Fig. 3.28a–d Ellipsometric (a,b) transmittance (c) and reflectance (d) spectra recorded on $As_{50}Se_{50}$ amorphous thin film deposited on float glass substrate (experimental data – *black solid lines*) compared with best-fit theoretical spectra (*broken lines*)

In the region of strong absorption, the absorption coefficient is calculated as

$$x = \frac{A - \sqrt{A^2 - 4T_i^2 BD}}{2T_i D}\,, \tag{3.96}$$

where

$$T_i = \frac{2T_M T_m}{T_M + T_m} \tag{3.97}$$

and A, B and D are calculated using the above-mentioned $n(\lambda)$ approximation.

It is worth noting that the division into negligible absorption, weak absorption and strong absorption regions is quite arbitrary and should be checked using trial-and-error methods.

Ellipsometry

An example of ellipsometric spectra recorded in reflection configuration on a thin film is presented in Fig. 3.28. For photon energies lower than bandgap E_g of thin film material, we can distinguish interference phenomena (interference fringes) due to multiple reflections in the film. When photon energy in the spectrum increases and has a higher value than E_g, the onset of light absorption reduces the amplitude of fringes; and for even higher absorption values (higher photon energies) interference fringes completely disappear. In this spectral range, the incident light probes only the film surface, hence the approach derived for a single interface in the previous section can be used. This qualitative interpretation is valid also for the transmittance spectra treated previously by the Swanepoel method.

It should be noted here that in optics term *thin film* is usually used for films with thicknesses significantly smaller than coherence length of interacting (probe) light. In this case, the formalism of totally coherent light can be applied to the study of interference phenomena in the film. Quantitative analyses of ellipsometric spectra can provide information on the geometrical (film thickness) and optical (film complex refractive index) properties of a sample. However, desired optical and geometrical parameters cannot be expressed analytically as explicit functions of experimental Ψ and Δ values as was done previously in the case of a single interface (bulk samples). The solution of this inverse problem requires construction of a sample model. Selected parameters are then adjusted by a numerical procedure where differences between theoretically calculated and experimental spectra are iteratively minimized (nonlinear fitting). Theoretical spectra of Ψ and Δ are easily obtained when r_p and r_s reflection coefficients – determined for the thin film (see (3.86) and related remarks) – are inserted into the definition relation of ellipsometric angles (3.68). Often, it is helpful to add transmittance and reflectance experimental spectra and treat them simultaneously with ellipsometry spectra (recorded for several incidence angles). In this way, we can reduce possible correlations between searched parameters.

As an example of ellipsometry data treatment we present in Fig. 3.28 experimental ellipsometric, transmittance and reflectance spectra compared with best-fit theoretical ones. Spectra were recorded on a chalcogenide $As_{50}Se_{50}$ amorphous film deposited onto float glass. The back side of the substrate was grounded before ellipsometry and reflectance measurements to suppress spurious reflections from this interface. Sample model structure was then designed as a homogenous film sandwiched between two semi-infinitive media: ambient and substrate. In the case of transmittance, the effect of finite substrate was accounted for by equation (3.89). The refractive index of float glass was determined formerly by spectroscopic ellipsometry carried out on a naked substrate considering it as a bulk material (see previous section). Optical constants of the chalcogenide film were parameterized by

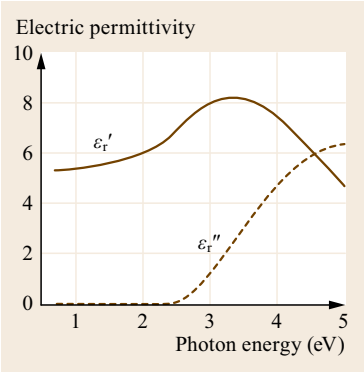


Fig. 3.29 Determined electrical permittivity spectra of $As_{50}Se_{50}$ film with thickness of 1085.5 ± 1.1 nm. Adjusted parameters of Tauc–Lorentz parameterization are $A = 155.16 \pm 1.39$ eV; $E_0 = 4.55 \pm 0.03$ eV; $C = 6.09 \pm 0.09$ eV; $E_g = 2.351 \pm 0.001$ eV; $\varepsilon_\infty = 1.05 \pm 0.01$

a Tauc–Lorentz formula (that is, a combination of the Lorentz oscillator and Tauc absorption edge models) that is the appropriate parameterization for amorphous semiconductors. The imaginary part of electric permittivity ε_r'' at a photon energy E is then expressed as

$$\varepsilon_r'' = \frac{AE_0C\left(E-E_g\right)^2}{\left(E^2-E_0^2\right)^2+CE^2}\frac{1}{E} \qquad E > E_g$$
$$\varepsilon_r'' = 0 \qquad E < E_g$$

and the real part ε_r' can be calculated with the help of Kramers–Kronig relations ((3.9a) and (3.9b)) or by an analytic solution [3.86, 87]. Parameters of the Tauc–Lorentz formula together with the film thickness were free parameters adjusted by the fitting procedure. Figure 3.29 represents the determined spectral dependence of the relative permittivity of $As_{50}Se_{50}$ with adjusted values of searched parameters.

Of course, thin films usually have various defects and nonidealities, for example nonuniform thickness, surface roughness, internal nonuniformities, refractive index gradient, and so on. Ellipsometry is a sensitive method capable of detecting most of the mentioned defects and, moreover, theoretical approaches have been developed to account for them and analyze them. Monographs focused on spectroscopic ellipsometry [3.75–77] and related articles treat these topics.

3.5 Optical Materials

3.5.1 Abbe Number or Constringence

In an optical medium, the Abbe number is defined as the inverse of its dispersive power; that is, it represents the relative importance of refraction and dispersion. There are two common definitions based on the use of different standard wavelengths. The Abbe number ν_d is defined by

$$\nu_d = (n_d - 1)/(n_F - n_C)\,, \tag{3.98}$$

where n_F, n_d and n_C are the refractive indices of the medium at the Fraunhofer standard wavelengths corresponding to the helium d-line ($\lambda_d = 587.6$ nm, yellow), the hydrogen F-line ($\lambda_F = 486.1$ nm, blue) and the hydrogen C-line ($\lambda_C = 656.3$ nm, red) respectively. The Abbe number of a few glasses are listed in Table 3.10. The Abbe number ν_e, on the other hand, is defined by

$$\nu_e = (n_e - 1)/(n_{F'} - n_{C'})\,, \tag{3.99}$$

where n_e, $n_{F'}$ and $n_{C'}$ are the refractive indices at wavelengths of the e-line (546.07 nm), the F′-line (479.99 nm) and the C′-line (643.85 nm) respectively.

The relationship between the refractive index n_d and the Abbe number ν_d is given by the Abbe diagram, where n_d is plotted on the y-axis against its corresponding ν_d value for different glasses, usually with the Abbe numbers decreasing on the x-axis, as shown in Fig. 3.30.

3.5.2 Optical Materials

Optical materials are those crystalline solids and glasses that are commonly used in the construction of optical components, such as lenses, prisms, windows, mirrors and polarizers. In addition to having the required optical properties, such as a well-defined refractive index n with a known dependence on wavelength $n = n(\lambda)$ and temperature, optical materials should also have various other desirable material properties, such as good ho-

Table 3.10 Abbe numbers for a few glasses. PC denotes polycarbonate, PMMA is polymethylmethacrylate, and PS represents polystyrene

Optical glass →	SF11	F2	BaK1	Crown glass	Fused silica	PC	PMMA	PS
ν_d →	25.76	36.37	57.55	58.55	67.80	34	57	31

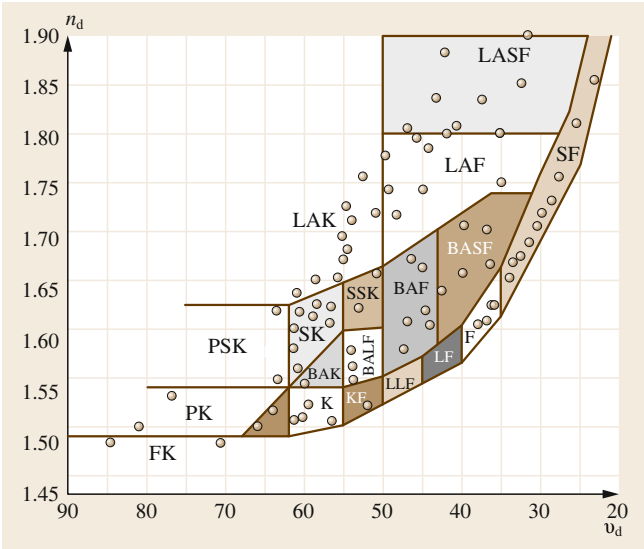


Fig. 3.30 The Abbe diagram is a diagram in which the refractive indices n_d of glasses are plotted against their Abbe numbers in a linear n_d versus ν_d plot and, usually, with the Abbe number decreasing along the x-axis, rather than increasing. A last letter of F or K represents flint or crown glass. Other symbols are as follows: S, dense; L, light; LL, extra light; B, borosilicate; P, phosphate; LA or La, lanthanum; BA or Ba, barium. Examples: BK, dense flint; LF, light flint; LLF, extra light flint; SSK, extra dense crown; PK, phosphate crown; BAK, barium crown; LAF, lanthanum flint

mogeneity (including negligible macroscopic variations in the refractive index), negligible thermal expansion (small α_L, thermal expansion coefficient), resistance to mechanical damage (resistance to scratching), and resistance to chemical degradation (chemical corrosion, staining). Also, in the case of optical glasses, it is desirable to have negligible air (or gas) bubbles incorporated into the glass structure during fabrication (during shaping for example), negligible stress-induced birefringence, negligible nonlinear properties (unless used specifically due to its nonlinearity), and negligible fluorescence. In all of these cases, negligible implies less than a tolerable quantity within the context of the particular optical application. For example, at a given wavelength and for many glass materials used in optics, n should not vary by more than 10^{-5}, while 10^{-6} is required for certain optics applications in astronomy. In addition, they should have good or reliable manufacturability at an affordable cost. There are various useful optical materials that encompass not only single crystals (such as CaF_2, MgF_2, quartz, sapphire) but also a vast range of glasses (which are supercooled liquids with high viscosity, such as flint and crown glasses as well as fused silica). Higher refractive index materials have more refractive power and allow lens designs that need less curvature to focus light, and hence tend to give fewer aberrations. Flint glasses have a larger refractive index than crown glasses. On the other hand, crown glasses are chemically more stable, and can be produced more to specification. While most optical materials are used for their optical properties (such as in optical transmission), certain *optical* materials (auxiliary materials) are used in optical applications such as mirror substrates and optical spacers for their nonoptical properties, such as their negligible thermal expansion coefficients. Some optical properties of selected optical materials and their applications are listed in Table 3.11.

3.5.3 Optical Glasses

Optical glasses are a range of noncrystalline transparent solids used to fabricate various optical components, such as lenses, prisms, light pipes and windows. Most (but not all) optical glasses are either crown (K) types or flint (F) types. K-glasses are usually soda-lime-silica glasses, whereas flint glasses contain substantial lead oxide; hence F-glasses are denser and have higher refractive powers and dispersions. Barium glasses contain barium oxide instead of lead oxide and, like lead glasses, have high refractive indices, but lower dispersions. There are other high refractive index glasses, such as lanthanum- and rare earth-containing glasses. Optical glasses can also be made from various other glass formers, such as boron oxide, phosphorus oxide and germanium oxide. The Schott glass code or number is a special number designation (511 604.253 for Schott glass K7) in which the first three numbers (511) represent the three decimal places in the refractive index ($n_d = 1.511$), the next three numbers (604) represent the Abbe number ($\nu_d = 60.4$), and the three numbers after the decimal (253) represent the density ($\rho = 2.53\,g/cm^3$). A different numbering system is also used, where a colon is used to separate n_d and ν_d; for example, 517:645 for a particular borosilicate crown means $n_d = 1.517$, $\nu_d = 64.5$ (Sect. 3.5.1).

In the Schott glass coding system, optical glasses are represented by letters in which a last letter of K refers to crown, and F to flint. The first letters usually represent the most important component in the glass, such as P in the case of phosphate. The letters Kz (*Kurz*), L (*leicht*) and S (*schwer*) before K or F represent short, light and dense (heavy) respectively (from German). S after K or F means *special*. Examples include: BK, borosilicate crown; FK, fluor crown; PK,

Table 3.11 The refractive indices, n_d, and Abbe numbers, v_d, (3.98) of selected optical materials (compiled from the websites of Oriel, Newport and Melles-Griot); n_d at $\lambda_d = 587.6\,\text{nm}$, α_L is the linear thermal expansion coefficient

Glasses	Transmission (typical, nm)	n_d	v_d	Applications	Comment
Fused silica	175–2000	1.45846	67.8	Lenses, windows, prisms, interferometric FT-IR components. UV lithography	Synthetic. Has UV properties; transmittance and excellent thermal low α_L. Resistant to scratching
SF 11, flint	380–2350	1.78472	25.76	Lenses, prisms	Flint glasses have $v_d < 50$
LaSFN9, flint	420–2300	1.85025	32.17	Lenses, prisms	High refractive index. More lens power for less curvature
BK7, borosilicate crown	380–2100	1.51680	64.17	Visible and near-IR optics. Lenses, windows, prisms, interferometric components	All around excellent optical lens material. Not recommended for temperature-sensitive applications
BaK1, barium crown	380–2100	1.57250	57.55	Visible and near-IR optics. Lenses, windows, prisms, interferometric components	All around excellent optical lens material. Not recommended for temperature-sensitive applications
Optical crown	380–2100	1.52288	58.5	Lenses, windows, prisms, interferometric components	Lower quality than BK7
Pyrex, borosilicate glass		1.43385	66	Mirrors	Low thermal expansion
Crystals					
CaF_2 crystal	170–7000	1.43385	94.96	Lenses, windows for UV optics, especially for excimer laser optics	Sensitive to thermal shock
MgF_2 crystal	150–7000	$n_0 = 1.37774$ $n_e = 1.38956$		Lenses, windows, polarizers, UV transmittance	Positive birefringent crystal. Resistant to thermal and mechanical shock
Quartz, SiO_2 crystal	150–2500	$n_0 = 1.54431$ $n_e = 1.55343$		UV optics. Wave plates. Polarizers	Positive uniaxial birefringent crystal
Sapphire, Al_2O_3 crystal	150–6000	1.7708 (546.1 nm)		UV–far-IR windows, high power laser optics	High surface hardness, scratch resistant. Chemically inert
Auxiliary optical materials					
ULE SiO_2-TiO_2 glass				Optical spacers	Very small thermal expansion
Zerodur, glass ceramic composite		1.5424	56–66	Mirror substrates. Not suitable for transmission optics due to internal scattering	Ultra-low α_L. Fine mixture of glass and ceramic crystals (very small size)

phosphate crown; PSK, dense phosphate crown; BaLK, light barium crown; BaK, barium crown; BaSK, dense barium crown; SSK, extra dense barium crown; ZnK, zinc crown; LaK, lanthanum crown, LaSK, dense lanthanum crown; KF, crown flint; SF, dense flint; SFS, special dense flint; BaF, barium flint; BaLF, barium light flint; BaSF, dense barium flint; LLF, extra light flint; LaF, lanthanum flint.

Acknowledgments. Authors are grateful to Stephen Karrer O'Leary (University of British Columbia, Okanagan Campus) for many insightful discussions on the subject and his invaluable help in the first edition chapter. This work was supported by the grants LM2015082 and CZ.1.05/4.1.00/11.0251 from the Ministry of Education, Youth and Sports of the Czech Republic.

References

3.1 S. Adachi: *Properties of Group IV, III–V and II–VI Semiconductors* (Wiley, Chichester 2005)
3.2 M. Fox: *Optical Properties of Solids*, 2nd edn. (Oxford Univ. Press, Oxford 2010)
3.3 J.H. Simmons, K.S. Potter: *Optical Materials* (Academic, San Diego 2000)
3.4 D.E. Aspnes, A.A. Studna: Phys. Rev. **B27**, 985 (1983)
3.5 D.Y. Smith, M. Inokuti, W. Karstens: J. Phys. Condens. Matter **13**, 3883 (2001)
3.6 A.L. Cauchy: Bull. Sci. Math. **14**, 6 (1830)
3.7 A.L. Cauchy: *M'emoire sur la Dispersion de la Lumiere* (Calve, Prague 1836)
3.8 J.W. Fleming: Appl. Opt. **23**, 4486 (1984)
3.9 K.L. Wolf, K.F. Herzfeld, H. Geiger, K. Scheel (Eds.): *Handbuch der Physik*, Vol. 20 (Springer, Berlin, Heidelberg 1928)
3.10 M. Herzberger: Opt. Acta **6**, 197 (1959)
3.11 P.J. Herve, L.K.J. Vandamme: J. Appl. Phys. **77**, 5476 (1995)
3.12 D.F. Edwards, E. Ochoa: Appl. Opt. **19**, 4130 (1980)
3.13 W.L. Wolfe: In: *The Handbook of Optics*, ed. by W.G. Driscoll, W. Vaughan (McGraw-Hill, New York 1978)
3.14 D. Dale, F. Gladstone: Philos. Trans. **153**, 317 (1863)
3.15 S.H. Wemple, M. DiDominico Jr.: Phys. Rev. **3**, 1338 (1971)
3.16 W.J. Turner, W.E. Reese: Phys. Rev. **127**, 126 (1962)
3.17 W.G. Spitzer, H.Y. Fan: Phys. Rev. **106**, 882 (1957)
3.18 E. Hagen, H. Rubens: Ann. Phys. **14**, 986 (1904)
3.19 R.J. Elliott, A.F. Gibson: *An Introduction to Solid State Physics and Its Applications* (Macmillan, London 1974)
3.20 J.D. Wiley, M. DiDomenico: Phys. Rev. **B1**, 1655 (1970)
3.21 H.R. Riedl: Phys. Rev. **127**, 162 (1962)
3.22 R.L. Weihler: Phys. Rev. **152**, 735 (1966)
3.23 H.B. Briggs, R.C. Fletcher: Phys. Rev. **91**, 1342 (1953)
3.24 C.R. Pidgeon: Optical properties of solids. In: *Handbook on Semiconductors*, Vol. 2, ed. by M. Balkanski (North Holland, Amsterdam 1980) p. 223
3.25 H.E. Ruda: J. Appl. Phys. **72**, 1648 (1992)
3.26 H.E. Ruda: J. Appl. Phys. **61**, 3035 (1987)
3.27 W. Kaiser, R.J. Collins, H.Y. Fan: Phys. Rev. **91**, 1380 (1953)
3.28 I. Kudman, T. Seidel: J. Appl. Phys. **33**, 771 (1962)
3.29 A.E. Rakhshani: J. Appl. Phys. **81**, 7988 (1997)
3.30 R.H. Bube: *Electronic Properties of Crystalline Solids* (Academic, San Diego 1974)
3.31 F. Urbach: Phys. Rev. **92**, 1324 (1953)
3.32 J. Pankove: Phys. Rev. **140**, 2059 (1965)
3.33 R.A. Street: *Hydrogenated Amorphous Silicon* (Cambridge Univ. Press, Cambridge 1991)
3.34 D.A. Papaconstantopoulos, E.N. Economou: Phys. Rev. **24**, 7233 (1981)
3.35 H. Cohen, H. Fritzsche, S.R. Ovshinsky: Phys. Rev. Lett. **22**, 1065 (1969)
3.36 N.F. Mott, E.A. Davis: *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford 1979)
3.37 J. Singh, K. Shimakawa: *Advances in Amorphous Semiconductors* (Taylor Francis, London 2003)
3.38 S. Sherman, S. Wagner, R.A. Gottscho: Appl. Phys. Lett. **69**, 3242 (1996)
3.39 T. Tiedje, J.M. Cebulla, D.L. Morel, B. Abeles: Phys. Rev. Lett. **46**, 1425 (1981)
3.40 K. Winer, L. Ley: Phys. Rev. **36**, 6072 (1987)
3.41 D.P. Webb, X.C. Zou, Y.C. Chan, Y.W. Lam, S.H. Lin, X.Y. Lin, K.X. Lin, S.K. O'Leary, P.K. Lim: Solid State Commun. **105**, 239 (1998)
3.42 W.B. Jackson, S.M. Kelso, C.C. Tsai, J.W. Allen, S.-H. Oh: Phys. Rev. **31**, 5187 (1985)
3.43 S.K. O'Leary, S.R. Johnson, P.K. Lim: J. Appl. Phys. **82**, 3334 (1997)
3.44 S.M. Malik, S.K. O'Leary: J. Non-Cryst. Solids **336**, 64 (2004)
3.45 S.R. Elliott: *The Physics and Chemistry of Solids* (Wiley, Sussex 1998)
3.46 J. Singh: Nonlinear Opt. **29**, 119 (2002)
3.47 J. Tauc: Phys. Status Solidi **15**, 627 (1966)
3.48 F. Orapunt, S.K. O'Leary: Appl. Phys. Lett. **84**, 523 (2004)
3.49 A.P. Sokolov, A.P. Shebanin, O.A. Golikova, M.M. Mezdrogina: J. Phys. Condens. Matter **3**, 9887 (1991)
3.50 G.D. Cody: Semicond. Semimet. **21**, 11 (1984)
3.51 K. Shimakawa, Y. Ikeda, S. Kugler: Optoelectronic materials and devices. In: *Non-Crystalline Materials for Optoelectronics*, ed. by G. Lucovsky, M.A. Popescu (INOE Publ., Bucharest 2004) p. 103
3.52 K. Morigaki: *Physics of Amorphous Semiconductors* (World Scientific, London 1999)
3.53 L. Ley: In: *The Physics of Hydrogenated Amorphous Silicon II*, ed. by J.D. Joannopoulos, G. Lukovsky (Springer, Berlin, Heidelberg 1984) p. 61
3.54 T. Aoki, H. Shimada, N. Hirao, N. Yoshida, K. Shimakawa, S.R. Elliott: Phys. Rev. **59**, 1579 (1999)
3.55 J. Singh: J. Mater. Sci. Mater. Electron. **14**, 171 (2003)
3.56 S.M. Malik, S.K. O'Leary: J. Mater. Sci. Mater. Electron. **16**, 177 (2005)
3.57 S. Abe, Y. Toyozawa: J. Phys. Soc. Jpn. **50**, 2185 (1981)
3.58 A.O. Kodolbas: Mater. Sci. Eng. **98**, 161 (2003)
3.59 J. Singh: *Excitation Energy Transfer Processes in Condensed Matter* (Plenum, New York 1994)
3.60 J. Singh: Nonlinear Opt. **18**, 171 (1997)
3.61 J. Singh, T. Aoki, K. Shimakawa: Philos. Mag. **82**, 855 (2002)
3.62 R.J. Elliott: In: *Polarons and Excitons*, ed. by K.G. Kuper, G.D. Whitfield (Oliver Boyd, London 1962) p. 269
3.63 T. Aoki, S. Komedoori, S. Kobayashi, T. Shimizu, A. Ganjoo, K. Shimakawa: Nonlinear Opt. **29**, 273 (2002)
3.64 G. Blasse, B.C. Grabmaier: *Luminescent Materials* (Springer, Berlin, Heidelberg 1994)
3.65 B. Henderson, G.F. Imbusch: *Optical Spectroscopy of Inorganic Solids* (Clarendon, Oxford 1989)
3.66 B. DiBartolo: *Optical Interactions in Solids* (Wiley, New York 1968)
3.67 W.T. Carnall, G.L. Goodman, K. Rajnak, R.S. Rana: J. Chem. Phys. **90**, 3443 (1989)
3.68 P.C. Becker, N.A. Olsson, J.R. Simpson: *Erbium-Doped Fiber Amplifiers. Fundamentals and Technol-*

ogy (Academic, New York 1999)
3.69 S. Hüfner: *Optical Spectra of Rare Earth Compounds* (Academic, New York 1978)
3.70 M.J. Weber: Phys. Rev. **157**, 262 (1967)
3.71 E. Desurvire: *Erbium-Doped Fibre Amplifiers* (Wiley, New York 1994)
3.72 K.A. Gschneidner Jr., L. Eyring (Eds.): *Handbook on the Physics and Chemistry of Rare Earths*, Vol. 25 (Elsevier, Amsterdam 1998)
3.73 M. Born, E. Wolf: *Principles of Optics* (Cambridge Univ. Press, Cambridge 1999)
3.74 M. Dressler, G. Gruner: *Electrodynamics of Solids, Optical Propetries of Electrons in Matter* (Cambridge Univ. Press, New York 2003)
3.75 H.G. Tompkins, E.A. Irene (Eds.): *Handbook of Ellipsometry* (Springer, Heidelberg 2005)
3.76 H. Fujiwara: *Spectroscopic Ellipsometry, Principles and Applications* (Wiley, Chichester 2007)
3.77 R.M.A. Azzam, N.M. Bashara: *Ellipsometry and Polarized Light* (Elsevier, Amsterdam 2003)
3.78 G. Beibly: *Aggregation and Flow of Solids* (Macmilian, London 1921)
3.79 F. Wooten: *Optical Properties of Solids* (Academic Press, London 1972)
3.80 I. Ohlidal, D. Franta: Prog. Opt. **41**, 181 (2000)
3.81 L. Ward: *The Optical Constants of Bulk Materials and Films* (Adam Hilger, Bristol, Philadelphia 1988)
3.82 D.E. Aspnes: Spectroscopic ellipsometry of solids. In: *Optical Properties of Solids: New Developments*, ed. by B.O. Seraphin (North-Holland, Amsterdam 1976) p. 801
3.83 R.W. Collins, K. Vedam: Optical properties of solids. In: *Encyclopedia of Applied Physics*, Vol. 12, ed. by G.L. Trigg (Wiley-VCH, New York 1995) p. 285
3.84 W.B. Jackson, N.M. Amer, A.C. Boccara, D. Fournier: Appl. Opt. **20**, 1333 (1981)
3.85 R. Swanepoel: J. Phys. E **17**, 896 (1984)
3.86 G.E. Jellison Jr., F.A. Modine: Appl. Phys. Lett. **69**, 371 (1996)
3.87 G.E. Jellison Jr., F.A. Modine: Appl. Phys. Lett. **69**, 2137 (1996)

4. Magnetic Properties: From Traditional to Spintronic

Charbel Tannous, Jacek Gieraltowski

This chapter reviews basic concepts used in the traditional macroscopic magnetism in order to understand current and future developments of submicronic spin-based electronics where interplay of electronic and magnetic properties is crucial. Traditional magnetism is based on macroscopic observation and physical quantities are deduced from classical electromagnetism. Physical interpretations are usually made with reference to atomic magnetism where localized magnetic moments and atomic physics prevail, despite the fact that standard ferromagnetic materials such as Fe, Co, and Ni are not localized-type magnets (they have extended s and localized d electronic states). While this picture might be enough to understand some aspects of traditional storage and electromechanics, it is no longer sufficient for the description of condensed matter systems with smaller length scales progressing toward the nanometer range. The precise nature of magnetism (localized, free, or itinerant like Fe, Co, and Ni transition metals) with simultaneous presence of charge and spin of carriers should be considered. In addition, when we deal with thin films or multilayers as in conventional electronics or with reduced dimensionality objects such as wires, pillars, dots, or grains, magnetic properties are expected to be different from three-dimensional conventional bulk systems.

4.1 **Traditional Magnetism** 87
4.1.1 Fundamental Magnetic Quantities 88
4.1.2 The Hysteresis Loop 89
4.1.3 Intrinsic Magnetic Properties of a Material 93
4.1.4 Traditional Types of Magnetism and Classes of Magnetic Materials 96
4.2 **Nonconventional Magnetism and Progress Toward Spintronics and Quantum Devices** 99
4.2.1 Types of Exchange and Coupling in Magnetic Materials 99
4.2.2 Thin Magnetic Films 100
4.2.3 Electronic Properties: Localized, Free, Itinerant Magnetism and Spin-Polarized Band Structure 101
4.3 **Spintronics and Quantum Information Devices** 104
4.3.1 Magnetic Signal Processing Devices 106
4.3.2 Magnetic Quantum Dot Arrays 107
4.3.3 Magnetic Vortex Properties and Applications 107
References 109

Digital information technology is based on three major constituents:

- Processing of information (transistors, logic gates, CPU, RAM, digital signal processor (DSP), etc.)
- Communication of information (networks, switches, cables, fibers, antennas, etc.)
- Storage of information (hard disks, CD, DVD, Blu-ray, Flash, etc.).

Traditionally, magnetism has been confined to safe storage at the time of the development of bubble and ferrite core technologies when RAM memory belonged to the realm of magnetism. RAM memory is a special type of information storage called primary (nonpermanent) storage to distinguish it from secondary (permanent) or mass storage such as hard disks, floppies, CDs, DVDs, Blu-rays, Flash, etc.

Presently, magnetism is undergoing a very deep transformation because of several recent developments, among which is the progress toward the nanometer scale in integrated circuits. On that length scale, quantum effects become extremely important, and the spin of carriers becomes of interest since it might be conserved over that length scale and, consequently, can be used to carry and manipulate information.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_4

Moreover, magnetism is endowed with at least two particularly interesting features:

1. The absence of dipolar length scale stemming from the absence of monopoles in contrast with the electrical dipole, which leads to easy adaptation to device scaling.
2. The atomic nature of dynamic response, which is equivalent to being universal and of potentially high speed.

Those properties pave the way toward the fabrication of new integrable and fast devices based on charge and spin (spintronic devices) instead of charge only, like in traditional microelectronics. New types of junctions, transistors, logic gates, sensors, and devices can be built in a way such that magnetism will be able to tackle processing of information along with its storage.

Quantum effects have already delivered many products since they are the basis of the GMR effect (Table 4.1) that underlies the recent surge of extremely higher densities in hard disks; nevertheless they are slated for other deeper developments geared toward the building of the basic parts of the *quantum computer* and *quantum communication* devices. The *quantum computer* is based on the *qubit* or the quantum bit, which is the basic unit of information replacing the classical bit. If we consider a sphere, the classical bit can be viewed as an object with two possible states at the north and the south pole of the sphere, whereas a qubit is an object that can sit anywhere on the surface of the sphere (called the *Bloch sphere*, see *Nielsen* and *Chuang* [4.28]). A quantum computer can do massive parallel computations since quantum mechanics is linear. This means one is allowed at any time to access any superposition of states, in contrast to a classical Von-Neumann-type computer where one has access to only a single state at any time (Table 4.1).

Quantum communications are extremely secure since a caller may build a coherent state with the callee and any intrusion can perturb the coherence, providing a very efficient detection.

From the fundamental point of view, one may start with a single magnetic moment representing a magnetic material and study its behavior before indulging into many interacting moments that are the building blocks of magnetic materials. From an applied point of view, the orientation of the moment (left–right or up–down) defines the value of the bit. Once the orientation of the moment has been linked to a bit value, it becomes

Table 4.1 Magnetic topics and corresponding selected references. *Kittel* [4.27] and *Hurd* [4.26] are recommended as general introductory references

Topic	Applications/Comments	Reference
Amorphous magnets	Shielding, sensing, transformers, transducers, etc.	*Boll* and *Warlimont* [4.1]
Coherent rotation model	Hysteresis loop	*Stoner* and *Wohlfarth* [4.2]
Coupling and exchange in multilayers	Biquadratic exchange, exchange bias, etc.	*Platt* et al. [4.3], *Slonczewski* [4.4], *Koon* [4.5]
Giant magnetoimpedance	Sensing, detection, etc.	*Tannous* and *Gieraltowski* [4.6]
Giant magnetoresistance (GMR)	Sensors, spin filters, magnetic recording heads, etc.	*White* [4.7]
Giant magnetostriction	Smart wings, actuators, transducers, resonators, etc.	*Schatz* et al. [4.8], *Dapino* et al. [4.9]
Itinerant magnetism	Magnetism in transition metals	*Himpsel* et al. [4.10]
Localized magnetism	Atoms/molecules/ions/insulators	*Jansen* [4.11]
Losses in magnetic materials	Eddy currents, hysteresis loss, etc.	*Goodenough* [4.12]
Magnetic recording	Information storage technology	Chap. 49 in this Handbook, *Richter* [4.13]
Magnetic thin films	Growth and characterization	*Himpsel* et al. [4.10]
Magnetoelastic effects	Cantilevers, microelectromechanical systems, etc.	*Farber* et al. [4.14], *Dapino* et al. [4.9]
Microwave devices	Communications, bubble memory, etc.	*Coeure* [4.15], *Wigen* [4.16]
Permanent magnets	Relays, motors, transformers, etc.	*Gutfleisch* [4.17]
Quantum computing/ communications	Quantum devices, magnetic RAM, etc.	*Burkard* and *Loss* [4.18]
Sensors	Field detector, probes, etc.	*Hauser* et al. [4.19]
Soft magnets	Shielding, sensing, transformers, transducers, etc.	*Boll* and *Warlimont* [4.1]
Spintronics	Spin diode, spin LED, spin transistor, magnetic RAM, etc.	*Prinz* [4.20], *Zutic* et al. [4.21], *Wu* et al. [4.22] and *Dietl* et al. [4.23]
Technology overview	Permanent and soft magnets	*Kronmueller* [4.24], *Simonds* [4.25]
Types of magnetic order	Ferromagnetism, antiferromagnetism, diamagnetism, paramagnetism, etc.	*Hurd* [4.26]

important to understand the physical processes, the energetics, and the dynamics of the moment orientation change (moment reversal from left to right or moment flip from up to down) in order to be able to control, alter, and predict the bit value.

This work is organized as follows. We present the traditional elements of magnetism such as the hysteresis loop, conventional types of magnetism and magnetic materials, proceeding toward a brief description of non conventional magnetism useful for understanding new high-tech materials that will make the backbone of future spintronic and quantum information-based devices. In Sect. 4.3, we describe spintronics in more detail, drawing from analogies with standard electronics and describing magnetic quantum dot arrays with a single application based on magnetic vortex devices in the cylindrical dot case. Table 4.1 contains topical works for readers who are interested in additional information that is briefly mentioned or not covered in this work.

4.1 Traditional Magnetism

Classical magnetism relates to magnetic moments and their behavior under external field or mechanical stress or some other perturbing effect. Basically, one writes an energy for the magnetic moment and tries to understand its behavior with time. In order to describe the different magnetic energy terms that control the behavior of a moment, we can start from a single isolated moment in vacuum, at zero temperature $T = 0\,\mathrm{K}$, zero applied magnetic field ($\boldsymbol{H} = 0$), and zero frequency ($f = 0$). It is obvious that the energy is zero from the classical point of view (quantum mechanically, one might say that even at $T = 0\,\mathrm{K}$, there exist quantum fluctuations that might disorient, flip, or reverse the moment).

The next step is to apply a magnetic field ($T = 0\,\mathrm{K}$, $f = 0$); then we obtain the Zeeman energy $E_Z = -\boldsymbol{M} \cdot \boldsymbol{H}$, with $\boldsymbol{M}$ representing the moment. If we place the moment in an unbounded (infinite sized) crystal, there appears a magnetocrystalline anisotropic energy [4.27] characterized (in the uniaxial case such as cobalt) by a constant K (4.5) since the moment will point in a direction (called an easy axis direction) imposed by the crystal's internal symmetry, in contrast to vacuum where $\boldsymbol{M}$ is free to point in any direction. When the crystal has a set of easy axes (or easy planes), the moment will point in one of several directions (or any direction in one or several planes). This occurs in ferromagnetic materials such as iron or nickel possessing cubic anisotropy.

If the body containing the moment has a finite size, a new energy will appear: the demagnetization energy. Magnetic surface charges (poles) induced on the surface bounding the body create a demagnetizing field inside it (outside, it is called a stray field). The demagnetization energy is also called shape anisotropy energy or magnetostatic energy because it is expressed like anisotropy energy as $2\pi N_{\alpha\beta}\boldsymbol{M}_\alpha\boldsymbol{M}_\beta$, where $N_{\alpha\beta}$ is a set of factors (demagnetization coefficients) that depend on body shape (Einstein summation convention is used for repeated indices). In the ellipsoidal shaped body case, the latter expression is exact.

Now suppose we include several local moments in a material. If sites i and j carrying moments $\boldsymbol{M}_i$ and $\boldsymbol{M}_j$, respectively, are close enough (say less than 1 nm), we have an interaction energy called exchange energy (of electrostatic origin) $-A_{ij}\boldsymbol{M}_i.\boldsymbol{M}_j$ that will align moments $\boldsymbol{M}_i$ and $\boldsymbol{M}_j$ if A_{ij} is positive (ferromagnetic interaction). If A_{ij} is negative, the moments will align antiparallel to each other (antiferromagnetic interaction). In a ferromagnet, we have a nonzero net moment, whereas in an antiferromagnet the net moment is zero.

The above description considers localized magnetism in distinct atoms (such in a gas), ions, molecules, or in special materials (like insulators or rare-earth solid state compounds possessing external shell f electrons with a highly atomic-like character). If we have a conducting material with free electrons interacting with localized atoms/ions/molecules, a different type of magnetism, called itinerant magnetism, appears (e.g., transition metals with s and d electrons or magnetic semiconductors, etc.). Nevertheless, it is possible to extend the notion of magnetic moment to that case accounting for the combined effects of free and localized charges modeled as an effective number of Bohr magnetons (Sect. 4.1.1 and Table 4.1).

The different physical mechanisms and types of magnetism briefly described above operate at different length scales. In order to put things into perspective and ponder about what lies ahead in terms of possible developments and hurdles, the summary in Fig. 4.1 displays different mechanisms, characterization methods, and manufacturing processes along with the corresponding length scale. Note that, on the nanometer scale, device size becomes comparable to most ranges of interactions encountered in magnetic materials and this will trigger development of novel effects and fabrication of new devices (Sect. 4.3).

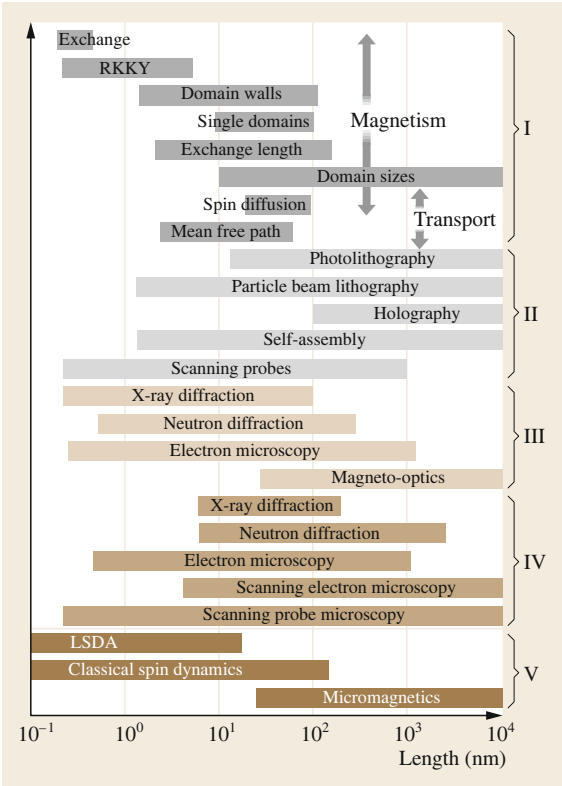


Fig. 4.1 Typical lengths (in nm) encountered in magnetic materials and spintronics with corresponding basic magnetic phenomena. Structures, interactions, and transport are gathered under group I. Group II is for growth and fabrication methods, whereas diagnostic techniques appear under groups III and IV. Group III is for depth-profiling techniques whereas group IV is for surface-probing techniques. Finally, group V is for theoretical analysis and characterization techniques. LSDA (local spin density approximation) is the spin counterpart of the density function theory (DFT) used for traditional band structure calculation. Since most interaction lengths are in the 1−100 nm range, the nanoelectronics device size is comparable to the interaction range. (After [4.29])

4.1.1 Fundamental Magnetic Quantities

Magnetization is the fundamental property exhibited by a magnetic substance originating from its electrons, similar to the electric dipole moment. It can be intrinsic (existing without any external field) or induced by an external magnetic field. In atoms/molecules/ions and rare-earth solids, magnetization is made from individual atomic-like localized magnetic moments. In the solid state, the moment is generally defined from the band structure, and an effective moment can be defined in terms of a Bohr magneton $\mu_B = e\hbar/2m$, the magnetic moment carried by a single electron (e and m are electron absolute charge and mass). Magnetization is a thermodynamic quantity that changes with temperature, mechanical stress, and chemical processes such as alloying [4.30].

Table 4.2 Selected ferromagnetic solids with their saturation magnetization M_s, effective Bohr magneton number n_B, and Curie temperature T_c. (After [4.27])

Substance	M_s (emu/cm^3) at 300 K	M_s (emu/cm^3) at 0 K	n_B at 0 K	Curie T_c (K)
Fe	1707	1740	2.22	1043
Co	1400	1446	1.72	1388
Ni	485	510	0.62	627
Gd	–	2060	7.63	292
Dy	–	2920	10.2	88
MnAs	670	870	3.4	318
MnBi	620	680	3.52	630
CrO_2	515	–	2.03	386
EuO	–	1920	6.8	69

When a field is applied to a magnetic substance, the largest acquired magnetization measured along the direction of the applied field is the saturation magnetization M_s, implying that all moments are aligned parallel to the field. Since temperature T misaligns moments, one defines a $T = 0$ K saturation moment $Nn_B\mu_B$, with N the number of ions atoms or molecules (for a mole $N = N_A$ Avogadro number) and n_B the effective number of Bohr magnetons μ_B. The saturation magnetization M_s value is $N_A n_B \mu_B$ per unit volume (Table 4.2).

n_B is different for atoms/ions/molecules and solids and determined by the electronic state. For example, it is 5 for Fe^{3+} ions (according to Hund rules, see *Jansen* [4.11]) since we have a $3d^5$ atomic configuration (without orbital contribution), whereas for Fe atom it is 2 ($3d^8$ configuration). Trivalent rare earth ions with highly localized 4f orbitals possess magnetic moments determined by $g_J\sqrt{J(J+1)}$, with g_J being the total Landé factor and J the total angular momentum. Transition metal ions possess magnetic moments determined by $2\sqrt{S(S+1)}$ with S being the total spin of the ion (no orbital contribution because of orbital moment quenching, see *Kittel* [4.27]). In the solid state, such as ferromagnetic metals, n_B is determined from the band structure (Table 4.1). In lower dimensions (clusters, dots, thin films, etc.), n_B tends to increase because of lower symmetry and coordination, since high symmetry and coordination tend to decrease it. Among the elements, rare earth ions (Dy^{3+}, Ho^{3+}, and Er^{3+}) possess large values of n_B on the order of 10, whereas in solids, Gd-based garnets [4.16] have n_B on the order of 15 (at $T = 0$ K).

The moment due to an angular momentum is given in the localized case (atoms/ions/molecules) by $-g_L\mu_B L$ (for an orbital angular momentum $\boldsymbol{L}$) with $g_L = 1$ (the orbital Landé factor), or by $-g_S\mu_B S$ (for a spin angular momentum $\boldsymbol{S}$) with $g_S = 2$ (the spin Landé factor). In the case of an atom/ion with a total angular momentum $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$, the moment is given by $-g_J\mu_B J$ with

$$g_J = \frac{3}{2} + \frac{[S(S+1) - L(L+1)]}{[2J(J+1)]} \tag{4.1}$$

the total Landé factor. In the solid state, the Landé factor is determined from the band structure (spin splitting of electron or hole bands in the presence of a magnetic field). It may become negative, anisotropic (tensor), and depends on wavevector $\boldsymbol{k}$. In semiconductors such as InSb: $g = -44$, GaAs: $g = 0.32$, InAs: $g = -12$ at wavevector $\boldsymbol{k} = 0$ [4.11, 21].

4.1.2 The Hysteresis Loop

When a magnetic field is applied to a ferromagnetic material [4.27], a magnetization change takes place. The best way to understand the underlying phenomena is to plot the value of the magnetization $\boldsymbol{M}$ projected along the direction of the applied field $\boldsymbol{H}$. The locus of the magnetization $\boldsymbol{M}$ measured along the direction of the applied magnetic field $\boldsymbol{M}(\boldsymbol{H})$ versus the field and depicted in the $\boldsymbol{M}$–$\boldsymbol{H}$ plane is the hysteresis loop. The term hysteresis means that when the material is field cycled (i. e., the field $\boldsymbol{H}$ is increased then decreased), two different nonoverlapping curves $\boldsymbol{M}(\boldsymbol{H})$ are obtained. Another point of view involves rather the induction $\boldsymbol{B}$ versus $\boldsymbol{H}$ since it is $\boldsymbol{B}$ that is, in fact, measured. Several points in the $\boldsymbol{M}$–$\boldsymbol{H}$ plane (Fig. 4.2) are worth mentioning. Point d is where the magnetization reaches its largest value M_s (saturation magnetization). Points i and f denote the magnetic fields $\pm H_c$ for which the magnetization is zero. Magnetic field H_c is called the coercive field. It is large for hard materials and small for soft magnets. Points e and h denote remanent magnetization M_r, that is, the leftover magnetization after the applied field is switched off, thus it is the hallmark of permanent magnets.

The ratio M_r/M_s, called squareness, is close to 1 when the applied magnetic field is close to the easy axis and therefore the hysteresis loop is squarer. Once the easy axis (or set of easy axes) is determined, the angle between the easy axis and the magnetic field (say ϕ) is varied and the hysteresis loop is graphed for different angles. When the angle ϕ is increased, the opening of the hysteresis loop is reduced; it is largest when the magnetic field is most parallel to the easy axis and smallest when the magnetic field is most parallel to the so-called hard axis. As shown in Fig. 4.3, characteristics of the hysteresis loop are depicted and for a given temperature and frequency of the applied field H, quantities

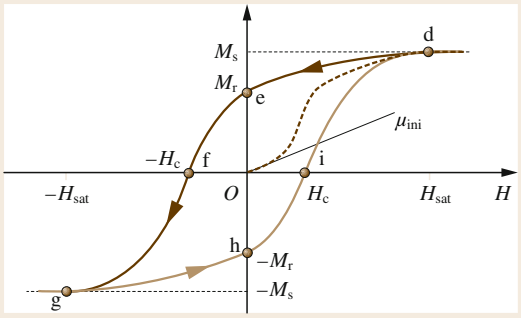


Fig. 4.2 Hysteresis loop with the fundamental points in the $\boldsymbol{M}$–$\boldsymbol{H}$ plane such as saturation magnetization M_s, remanent magnetization M_r, coercive field H_c (at which $M = 0$), and the saturation field $\boldsymbol{H}_{sat}$ (for which $M = M_s$). For a virgin (unmagnetized) material, the initial magnetization curve (*dashed line*) with permeability μ_{ini} are displayed. (After [4.31])

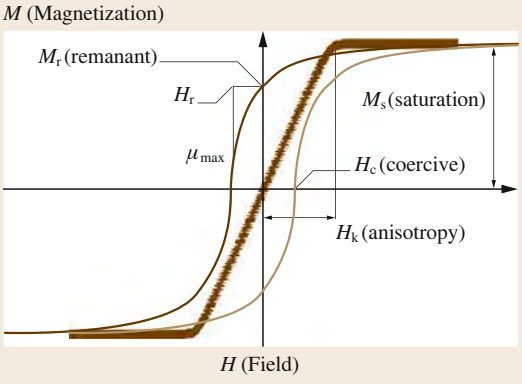


Fig. 4.3 Hysteresis loop obtained for an arbitrary angle ϕ, between the magnetic field and the anisotropy axis (easy axis). Associated quantities such as coercive field H_c, remanent magnetization M_r, and field H_r (given by the intersection of the tangent to the loop at $-H_c$ and the M_r horizontal line) are shown. The tangent to the loop at $-H_c$ is called maximum differential permeability μ_{max}. The *thick line* is the hysteresis loop when the field is along the hard axis and (the anisotropy field H_k) is the field value at the slope break where the magnetization reaches its saturation value M_s. Quantities such as H_c, H_r, M_r, and μ_{max} depend on the angle ϕ. When the field is along the easy axis, the coercive field reaches its largest value (the hysteresis loop is broadest) at which the magnetization jumps (at $\pm H_c$)

such as the remanent magnetization M_r, the remanent field H_r, the coercive field H_c, and the maximum differential permeability μ_{max} all vary with angle ϕ.

In addition, when the temperature varies or the frequency of the field is varied, the hysteresis loop changes and might even disappear altogether above a given temperature (Curie temperature) or be seriously altered by the frequency of the field (see below).

Magnetization Process Point of View

When a magnetic field is applied to a material, a magnetization process is initiated. That means a magnetization change can occur taking any of the following routes. At low applied magnetic fields (and low frequencies), domain boundaries bulge (this is a reversible regime implying that if we decrease the field again, the magnetization will go back to the initial state following the same path). This is called domain nucleation (high-slope path) or pinning process (small slope path). At higher fields, the walls are depinned and free to move. This is an irreversible regime (in contrast to the low field case) meaning that if we decrease the field, the magnetization versus field variation will not follow the same path. This regime corresponds to free domain wall movement. At higher fields (or frequencies), magnetization changes through moment rotation and thus we have two possibilities: if the material is homogeneous, all happens as if we had a single moment in the material. This is the coherent rotation regime (well described by the Stoner–Wohlfarth model, see Table 4.1). On the other hand, if magnetization changes in a nonuniform way (different points in the sample undergo different magnetization changes), we have a curling process. Finally, if the magnetic field is applied along the anisotropy axis (also called the easy axis) and suddenly reversed, the magnetization change is referred to as a switching process. This means the moment jumps from one value to its negative without any rotation process since magnetization is already along the easy axis (Fig. 4.4).

Energetic Point of View

Hysteretic behavior can be viewed as stemming from the motion of the energy minimum versus field. In Fig. 4.5, the field is cycled in a clockwise fashion. The energy minimum is displayed for several values of the field. Hysteresis is shown to arise from the asymmetric behavior of the energy versus magnetization as the field is cycled. Hysteresis loss is given by the area circumscribed by the hysteresis loop. Since soft materials have thinner hysteresis loops (equivalent to smaller coercivity) with respect to hard materials, their losses are smaller.

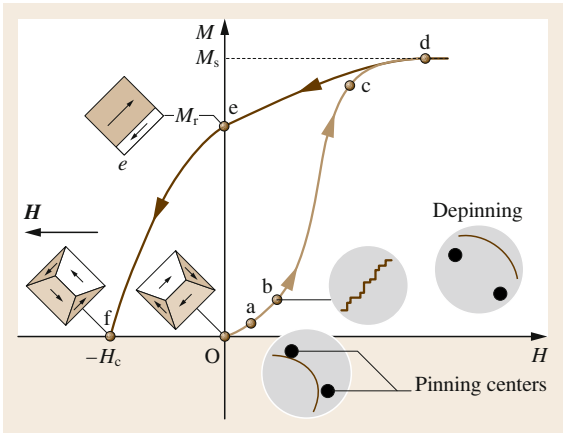


Fig. 4.4 Reversible and irreversible motion of the magnetization M with the field H. For small fields, $H < H_c$ and initial magnetization, Rayleigh's law is valid along a–b but for large H, saturation in M is induced by irreversible domain boundary motion along b–c, followed by rotation of magnetization along c–d. When the field is reversed along d–e–f toward $-H_c$, magnetization motions are not exactly those we had previously, and this is why a different $M(H)$ curve is obtained. Reversible movement at a low field means that domain walls are pinned by impurities and simply bulge under the action of H (0–a), whereas jerky irreversible movement is associated with depinning of the domain walls (a–b), creating Barkhausen noise. Free irreversible movement beyond pinning centers is depicted (*right*). (After [4.31])

Signal Processing Point of View

Consider a time-dependent applied field $\boldsymbol{H}(t)$ as an input signal to the magnetic material and the magnetization $\boldsymbol{M}(t)$ as the output. Then relation $\boldsymbol{M}(t) = \boldsymbol{F}(\boldsymbol{H}(t))$ pertains to a nonlinear filter. A magnetic material acts as a nonlinear filter since $\boldsymbol{M}$ is not proportional to $\boldsymbol{H}$ except at very low fields. A simple illustration of nonlinearity is to observe $\boldsymbol{M}(t)$ as a square signal with a sinusoidal input excitation $\boldsymbol{H}(t)$ (Fig. 4.6). In addition, the material imposes a propagation delay to the signal proportional to the width of the hysteresis loop (twice the coercive field). Hysteretic behavior means that different values of the output are obtained when input excitation is varied (increasing or decreasing). This is generally exploited in control and power systems, for instance.

Information Storage Point of View

Since hysteresis means that different output behavior is observed as the input is increased or decreased, this implies that one obtains a nonzero response ($\boldsymbol{M} \neq 0$) at zero excitation ($\boldsymbol{H} = 0$) called remanence (with value M_r when $\boldsymbol{H} = 0$). The major advantage of remanence

(in addition to its usefulness in permanent magnets) is that information (value of M_r) is stored without any excitation source ($\boldsymbol{H} = 0$). Getting M_r when the signal is decreased (say M_{r1}), we ought to have a different M_r when the signal is increased (say M_{r2}) because of hysteresis. Thus we have a bit value (0 for M_{r1} and 1 for M_{r2}). This contrasts sharply with (non-Flash) electronic storage where a voltage has to be maintained in order to keep the charge (representing information) in place (Chap. 49).

Electromagnetic Compatibility and Frequency Synthesis Point of View

Hysteretic behavior can be viewed as a flux ϕ (corresponding to magnetization $\boldsymbol{M}$) induced in a circuit with an exciting current $i(t)$ at time t (corresponding to field H). The resulting relationship $\phi[i(t)]$ is a nonlinear characteristic that can be expanded as

$$\phi[i(t)] = a_0 + a_1 i(t+\tau) + a_2[i(t+\tau)]^2 + a_3[i(t+\tau)]^3 + a_4[i(t+\tau)]^4 + \cdots , \tag{4.2}$$

where τ accounts for the response delay of the flux (the larger the delay, the broader the hysteresis curve). Ordinary linear inductance of the circuit corresponds to first derivative $[\mathrm{d}\phi/\mathrm{d}i]$ that equals a_1 for a short delay τ. Higher order terms define a nonlinear inductor behavior such that (for a short delay)

$$L(i) = a_1 + 2a_2 i + 3a_3 i^2 + 4a_4 i^3 + \cdots . \tag{4.3}$$

A nonlinear inductor response is interesting in itself from the frequency synthesis point of view since it can generate harmonics and subharmonics off the exciting current $i(t)$ (see, for instance, *Chua* [4.32]).

Time-Dependent Point of View

In spite of different points of view describing hysteresis, some experimentally observed phenomena cannot be explained with the above, such as a particularly important phenomenon known as the magnetic after-effect. Experimentally, one observes a magnetization change with time $\boldsymbol{M}(t)$, despite the fact that the applied field H is kept constant in time, implying a time-dependent hysteresis loop stemming solely from $\boldsymbol{M}(t)$ (Fig 4.7). Therefore, the material must contain units (single domain grains) that respond differently in time, as if each unit had its intrinsic delay; thus magnetization might be written in the form $M(t) = M_0 S(t) \ln(t)$, where $S(t)$ is called the magnetic viscosity. If the response time of each grain is considered of the thermal activation

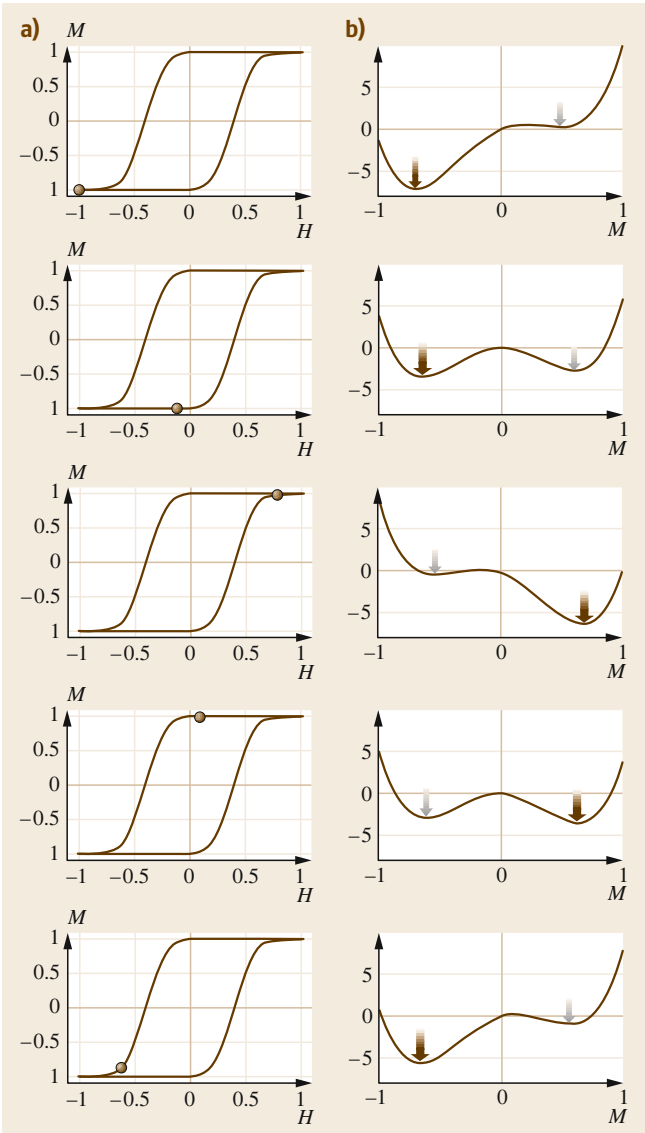


Fig. 4.5 Hysteresis loop (a) and energy as a function of magnetization (b). When field H is varied in the hysteresis loop (indicated by the *round spot*) in a clockwise fashion, the energy changes its form and its absolute minimum indicated by the *thick arrow* moves accordingly. Hysteresis is due to asymmetric energy behavior as the field is cycled

type, that is, following the Néel–Arrhenius formula (Chap. 49) $\tau = \tau_0 \exp(\Delta E/k_B T)$, where τ_0 is an attempt time to cross some energy barrier ΔE (Fig. 4.5) at a temperature T, then after-effect implies that we have a distribution of τ_0 and ΔE values. Calling the corresponding probability density function $g(\tau)$ and M_0 some scale initial magnetization, $\boldsymbol{M}(t)$ is consequently

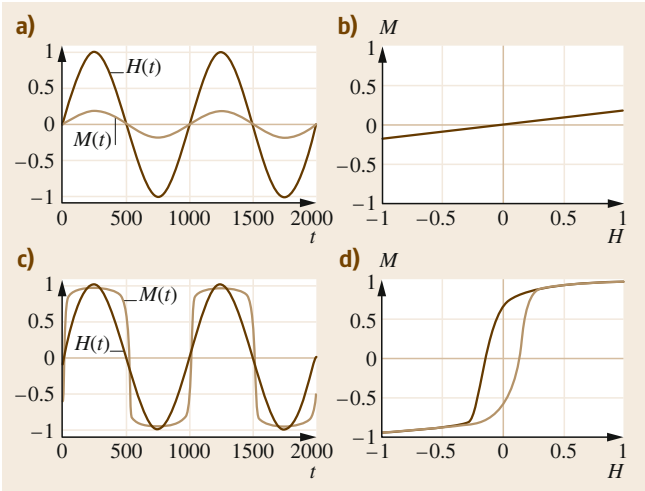


Fig. 4.6a–d Hysteresis loop as an input–output relationship. **(a,c)** The magnetic field $H(t)$ and magnetization $M(t)$ are displayed explicitly as functions of time t. **(b,d)** Resulting input–output relationship (***M*–*H*** characteristic) parametric curve $M(H)$. In **(a,b)** we have an output ($M(t)$), undelayed and proportional to the input $H(t)$ resulting in a paramagnetic (no hysteresis) ***M*–*H*** characteristic. **(c,d)** is an output $M(t)$, delayed and nonlinearly distorted version of the input $H(t)$ resulting in a hysteretic characteristic ***M*–*H***. (After [4.33])

defined as

$$M(t) = M_0 \int_0^\infty g(t) \exp\left(-\frac{t}{\tau}\right) \mathrm{d}\tau \,. \qquad (4.4)$$

In the simple case where we have a single response time τ^*, that is, $g(\tau) = 2\delta(\tau - \tau^*)$, $\boldsymbol{M}(t)$ behaves as $\boldsymbol{M}(t) = M_0 \exp(-t/\tau^*)$.

After analyzing different aspects of the hysteresis loop, we move on to tackle physical processes affecting magnetization behavior. Let us start with small magnetic fields $H < H_c$ with respect to the coercive field. The initial magnetization curve follows Rayleigh's law: $M(H) = \chi_0 H + \alpha_R H^2$, where χ_0 is the low-field susceptibility and α_R the Rayleigh coefficient. Since this law is valid for small fields, it describes the reversible motion of the magnetization.

We move next to description of a domain, which is some physical region inside a material where the magnetization is oriented along some direction.

Typically, when the extent of the material is smaller than the exchange length, one expects a single-domain structure (that contains about 10^{12}–10^{18} atoms, see also Fig. 4.1).

In magnetic materials described by a uniaxial anisotropy K, the expression of the exchange length is $\ell_{ex} = \sqrt{A/K}$ with A being the exchange stiffness constant (typically, $A \approx 10^{-6}$ erg/cm). In soft materials with anisotropy constant $K \approx 0$, the magnetostatic exchange length is used. It is defined as

$$\ell_{ex} = \sqrt{\frac{A}{2\pi M_s^2}}$$

given the magnetostatic energy density is $2\pi M_s^2$.

For instance, a recording medium is considered as being made of small grains made of single domains. As the recording density is increased, the grain size making the recording media decreases and if it is small enough, its magnetization becomes sensitive to thermal energy, that is, it can flip or reverse simply by the effect of temperature. This is called the superparamagnetic effect that traditionally limits longitudinal recording to 100 Gbit/in^2 and (current) perpendicular recording to 1 Tbit/in^2. It was partially solved recently through the introduction of antiferromagnetic coupling between the storage layer and a stabilization layer. Practically, a thin film of ruthenium (7–9 Å thick) is deposited between the storage layer and the stabilization layer (Chap. 49). If the extent of the magnetic material is large, one expects to have a multidomain structure separated by domain walls. A multidomain structure emerges in order to minimize the magnetostatic long-range dipolar interaction energy between the different moments existing in the different domains [4.34].

Since a domain is a piece of a material with the magnetization along a given direction, generally a magnetic material contains many regions where the magnetization is along some direction in order to minimize the magnetostatic energy. Regions with different magnetization orientations can be close to one another, albeit with a boundary called a domain wall (typically containing about 10^2–10^3 atoms, see Fig. 4.1). Saturation occurs when all these regions align along some direction imposed by the external applied field. The domain wall width (in CGS units) is equal to $\pi\sqrt{A/K}$, where A is the typical (nearest neighbor) exchange stiffness and K the typical uniaxial anisotropy (Sect. 4.1.3). It results from competing exchange and anisotropy energies. It is thinner for higher anisotropy or smaller exchange (in Fe it is about 30 nm, whereas in a hard material such as $Nd_2Fe_{14}B$, it is only 5 nm). Its energy is equal to $4\sqrt{AK}$. For bulk materials, it is of the Bloch type, whereas for thin films (i. e., with thickness on the order of the exchange length) it is of the Néel type (Fig. 4.8), again with a width proportional to exchange length $\sqrt{A/2\pi M_s^2}$. A single parameter $Q = 2K/M_s^2$ allows us to discriminate between simple ($Q < 1$) and complex wall profiles ($Q > 1$) [4.35].

During a magnetization process, the irreversible jerky movements of domain walls due to local instabil-

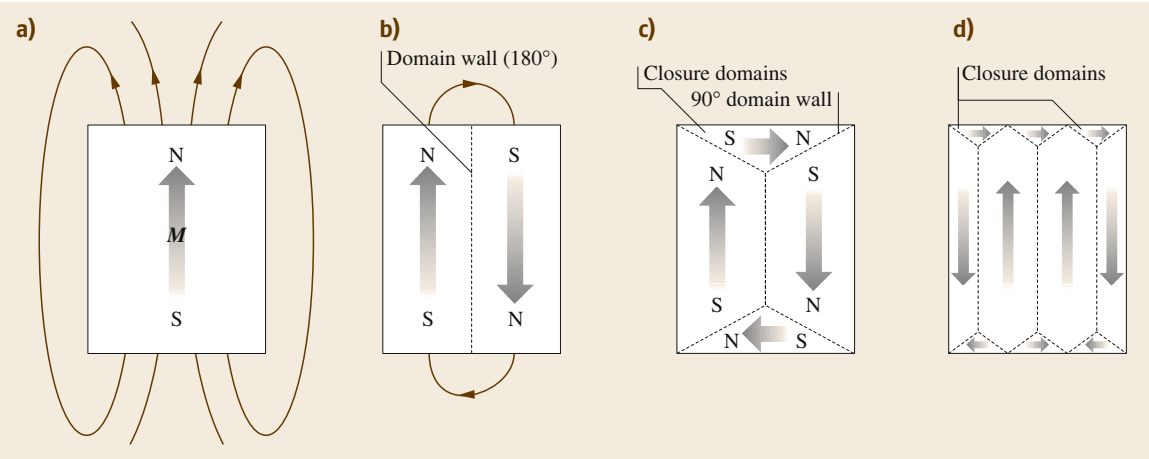


Fig. 4.7a–d Possible domain structures and wall structure in a magnetic material. **(a)** Initial magnetic configuration with a large stray field (outside the material). **(b)** In order to reduce magnetostatic energy (and stray field), domains form, as in a highly anisotropic material. If n domains are formed, the magnetostatic energy is reduced by n. (After [4.27]) **(c)** In a material with small anisotropy, 45° closure domains are formed to minimize the magnetostatic energy. **(d)** As more domains are formed, more walls are also formed until a compromise is reached since it is necessary to minimize the sum of the magnetostatic energy and wall energies. (After [4.31])

ities created by impurities, defects, inclusions, or interactions between domain walls lead to Barkhausen noise in a hysteresis loop cycle. In devices, it is preferred to induce magnetization change by rotation processes (implying that the external magnetic field must be applied along hard axis) that are less noisy.

4.1.3 Intrinsic Magnetic Properties of a Material

The induction $\boldsymbol{B}$ in a linear isotropic material is related to $\boldsymbol{H}$ through the relation $\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M})$, where μ_0 is the free space permeability and $\boldsymbol{M}$ is the magnetization of the material. Using the relation $\boldsymbol{M} = \chi \boldsymbol{H}$ relating the magnetization to the applied field, one gets $\boldsymbol{B} = \mu_0(1 + \chi)\mathbf{H}$. This leads to the definition of the total permeability $\mu = \mu_0(1 + \chi)$, which yields the constitutive relation $\boldsymbol{B}(\boldsymbol{H}) = \mu \boldsymbol{H}$. The quantity χ is called the susceptibility. In general, nonlinear material possesses nonlinear constitutive equations $\boldsymbol{M}(\boldsymbol{H})$ and $\boldsymbol{B}(\boldsymbol{H})$. A constitutive relation is like the $I(V)$ characteristic of an electronic device. For a linear material, the permeability μ is generally frequency dependent. If losses occur during propagation of an electromagnetic wave in a material resulting from the absorption of magnetic energy, it is possible to extend the definition of permeability to the complex plane (like in dielectrics). Losses are attributed to peaks in the imaginary part of the permeability at the absorption frequencies [4.12]. At low frequencies, losses are attributed to domains, whereas at high frequencies, losses are attributed to

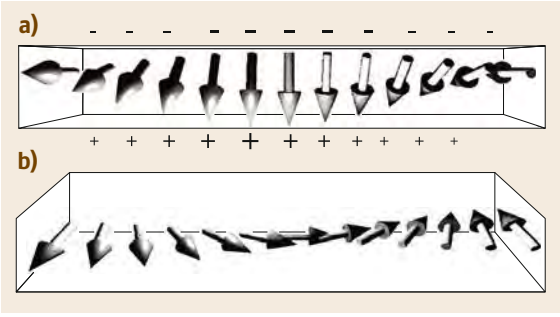


Fig. 4.8a,b Possible domain wall shapes in a magnetic material. **(a)** Bloch type for bulk materials with the magnetization rotating in a vertical plane with the associated poles. **(b)** Néel type for thin films. For these walls, the magnetization rotates in a horizontal plane when the width of the film is smaller than the exchange length. (After [4.36])

rotation processes. Since permeability relates the two vector quantities $\boldsymbol{B}$ and $\boldsymbol{H}$, in a linear anisotropic material (such as a linear crystal), the relation $\boldsymbol{B}(\boldsymbol{H}) = \mu \boldsymbol{H}$ becomes $\boldsymbol{B}_\alpha = \mu_{\alpha\beta}\boldsymbol{H}_\beta$ (Einstein summation convention) and $\mu_{\alpha\beta}$ is a rank-2 ($r = 2$) permeability tensor that has different components according to the relative directions of $\boldsymbol{B}$ and $\boldsymbol{H}$ (nine components in total or 3^r in three dimensions). For a nonlinear anisotropic material, we have $\mu_{\alpha\beta} = \partial B_\alpha / \partial H_\beta$.

When a material is run through a field cycle (hysteresis loop) for the first time, the magnetization $\boldsymbol{M}$ follows a path different from the above $\boldsymbol{M}(\boldsymbol{H})$ lines. That path represents the initial magnetization and the associ-

ated permeability is the initial permeability (Fig. 4.2). If a magnetic material is subjected to a static magnetic field $\boldsymbol{H}$ and a small time-dependent field $\boldsymbol{H}(t)$ (alternating at frequency f) parallel to $\boldsymbol{H}$, the magnetic response is called the longitudinal permeability (at frequency f). This kind of excitation is encountered in ferromagnetic resonance (FMR) problems [4.16].

Alternatively, one may describe magnetic behavior in terms of susceptibility. It relates the magnetization to the applied field in a linear material through $\boldsymbol{M} = \chi\boldsymbol{H}$. For a nonlinear isotropic material, we have $\chi = \partial M/\partial H$ and for a nonlinear anisotropic material, we have $\chi_{\alpha\beta} = \partial M_\alpha/\partial H_\beta$ with the $\boldsymbol{M}$ component along direction α and the $\boldsymbol{H}$ component along β. Any jump in χ signals an important magnetic change in the material. When the temperature is decreased, the Curie temperature signals the onset of ferromagnetic order from paramagnetic disorder, and the Néel temperature signals the onset of antiferromagnetic order from paramagnetic disorder. Magnetic susceptibility spans several orders of magnitude (from about 10^{-5} to $10^6\,\mathrm{cm}^3/\mathrm{mol}$).

The magnetization settles along a direction in order to minimize the sum of the magnetic anisotropy energy and the Zeeman energy provided by the applied field. Since the hysteresis loop changes its shape as we vary the applied field angle, it might reach its largest width for some angle. The field then indicates the easy axis direction, also called the anisotropy axis. The extreme value of the magnetic field is the coercive field $\pm H_\mathrm{c}$ at which the magnetization (switches) jumps to its saturation value $\pm M_\mathrm{s}$ (Fig. 4.9).

When the field is along the hard axis direction, it takes a longer time to align the magnetization slowly by rotation, in sharp contrast to fast switching that occurs when the field is along the easy axis direction. It is easily determined with the hysteresis loop when it is narrowest as the angle of the applied field is varied. At slope break we have the value of the anisotropy field $\pm H_\mathrm{K}$ (Fig. 4.10).

Crystals are generally anisotropic because their microscopic structure is not the same along different directions. Hence, different properties will show up along different orientations of the crystal. Magnetic properties are no special case, and there are special preferred directions along which the magnetization will point. This means there exists some energy contribution that will be minimized when the magnetization settles along these directions. This is the anisotropy energy. On top of being inherent to crystal structure, anisotropy can also be induced by an external field, a change of symmetry (e.g., surface with respect to bulk), or a mechanical deformation. Microscopically, anisotropy originates from spin–orbit coupling. Coercivity increases with anisotropy, that is, hardness increases with anisotropy (Fig. 4.11). Anisotropy energy can be evaluated from the hysteresis loop by determining the loop for various angles between the applied field and the easy axis. In general, it is an expansion of the form

$$\sum_{\alpha,\beta} K_{\alpha\beta} m_\alpha m_\beta + \sum_{\alpha,\beta,\gamma,\delta} K_{\alpha\beta\gamma\delta} m_\alpha m_\beta m_\gamma m_\delta \cdots , \tag{4.5}$$

where $K_{\alpha\beta}, K_{\alpha\beta\gamma\delta}$ are the anisotropy coefficients (second and fourth-order, respectively) and the coefficients m_α, m_β, $m_\gamma, \ldots$ are the components of $\boldsymbol{M}/M_\mathrm{s}$, the magnetization $\boldsymbol{M}$ normalized by the saturation magnetization M_s. α, β, γ are the indices along the crystal x, y, z axes. This means the norm of $\boldsymbol{m}$,

$$|\boldsymbol{m}| = \sqrt{[m_\mathrm{x}^2 + m_\mathrm{y}^2 + m_\mathrm{z}^2]} = 1 .$$

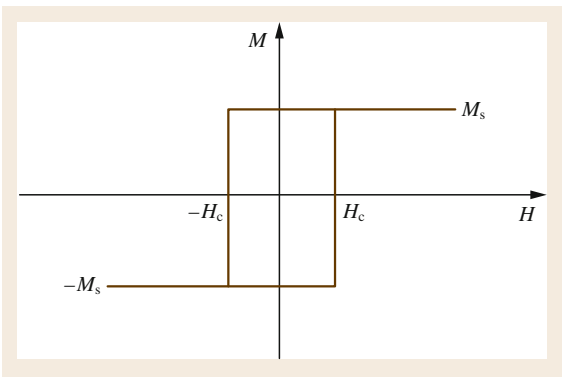


Fig. 4.9 Easy axis hysteresis loop. When the magnetic field is along the anisotropy axis (easy axis) of the material, the hysteresis loop is broadest, and the magnetization jumps (switches) to $\pm M_\mathrm{s}$ at the coercive field $\pm H_\mathrm{c}$

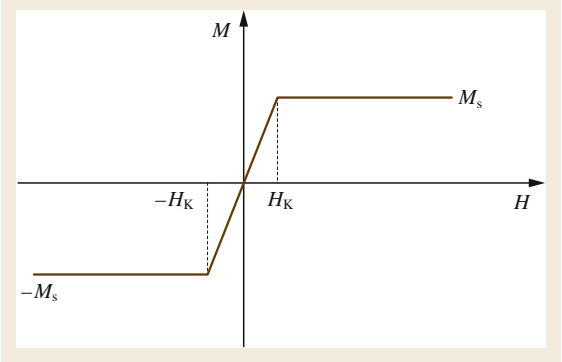


Fig. 4.10 Hard axis hysteresis loop. When the magnetic field is along the hard axis of the material, the hysteresis loop is the thinnest and the magnetization rotates smoothly from a value to another between $\pm H_\mathrm{K}$ the anisotropy field

Uniaxial and Biaxial Anisotropies

Anisotropy energy in the case of crystals with a single dominant axis such as hexagonal (like cobalt), tetragonal, and rhombohedral crystals is called uniaxial. To the lowest approximation, it is given by $K_1 \sin^2\theta$, where K_1 is the anisotropy strength, angle θ being the angle between the magnetization and the easy axis. The next higher approximation allows fourth-order terms to be added to the $K_1 \sin^2\theta$ term. In hexagonal crystals, the fourth-order term is of the form $K_2 \sin^4\theta$. Rhombohedral symmetry allows two fourth-order terms of the form $K_2 \sin^4\theta + K_4 \cos\theta \sin^3\theta \cos 3\phi$. In tetragonal crystals, the fourth-order terms are of the form $K_2 \sin^4\theta + K_4 \cos\theta \sin^4\theta \cos 4\phi$. The easy axis is taken along the z-axis, the angle ϕ is the magnetization polar angle in the x–y plane, and θ is the azimuthal angle.

Anisotropy energy in the case of crystals with two dominant axes such as monoclinic, triclinic, and orthorhombic crystals is called biaxial. It also means that two anisotropy directions exist and are competing. Some other cases are given below (see also Table 4.1 for other definitions).

Cubic Anisotropy. Anisotropy energy in the case of the most symmetric crystals, cubic crystals, is called as such because it relates to the cube's axes. The energy to fourth order is

$$K_1(m_1^2m_2^2 + m_1^2m_3^2 + m_2^2m_3^2) + K_2m_1^2m_2^2m_3^2 \,,$$

where the coefficients m_1, m_2, m_3 are the components of the normalized magnetization $\boldsymbol{M}/M_\mathrm{s}$ with respect to the crystal axes. That is $m_1 = \sin\theta\cos\phi$, $m_2 = \sin\theta\sin\phi$, $m_3 = \cos\theta$. The easy axis, as above, is taken along the z-axis and the angle ϕ is the polar angle in the x–y plane.

Shape Anisotropy. A finite-size magnetic body (ellipsoidal shape is assumed for simplicity) possessing a single uniform magnetization $\boldsymbol{M}$ (represented by its components $\boldsymbol{M}_\alpha$) contains a magnetic energy (also called magnetostatic energy) given by $2\pi N_{\alpha,\beta}\boldsymbol{M}_\alpha\boldsymbol{M}_\beta$ (Einstein summation). The $N_{\alpha\beta}$ coefficients are the demagnetization (akin to depolarization) coefficients of the body determined by its shape. The origin of the terminology is due to the resemblance to the familiar anisotropy energy of the form $K_{\alpha\beta}\boldsymbol{M}_\alpha\boldsymbol{M}_\beta$ (Einstein summation). If a magnetic material contains N moments (atoms, ions, or molecules each carrying a moment $\boldsymbol{\mu}_i$, the energy originating from dipolar coupling energy between the different moments is written as

$$W_\mathrm{dip} = \frac{1}{2}\sum_{i=1}^{\mathrm{N}}\sum_{j=1}^{\mathrm{N}} \frac{\boldsymbol{\mu}_i\cdot\boldsymbol{\mu}_j}{r_{ij}^3} - 3\frac{(\boldsymbol{\mu}_i\cdot\boldsymbol{r}_{ij})(\boldsymbol{\mu}_j\cdot\boldsymbol{r}_{ij})}{r_{ij}^5} \,. \quad (4.6)$$

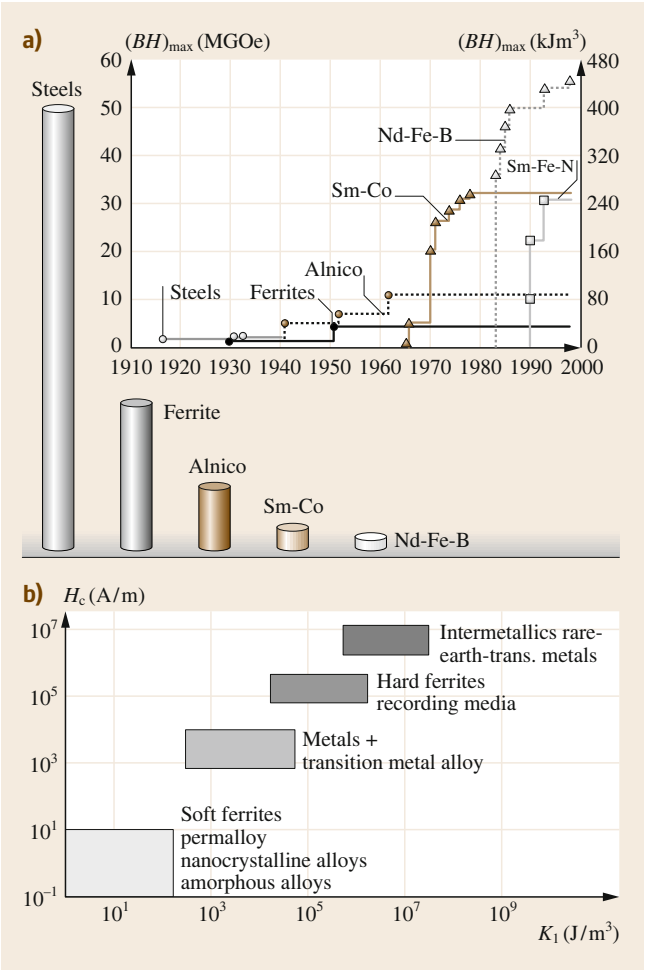


Fig. 4.11 **(a)** Development of hard materials for permanent magnets indicated by the value of the energy density $(BH)_\mathrm{max}$. Below progress is indicated by equivalent volume change with fixed energy density. **(b)** Correlation between coercivity H_c and anisotropy coefficient K_1 for various materials, their magnetic hardness, and corresponding applications. (After [4.17, 24])

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are two moments ($i \neq j$) in the material separated by a distance r_{ij}. This energy can be represented by constant coefficients ($N_{\alpha\beta}$) only in the case when the body has an ellipsoidal shape. Hence one can write

$$W_\mathrm{dip} \approx 2\pi N_{\alpha\beta}\boldsymbol{M}_\alpha\boldsymbol{M}_\beta \qquad \text{(Einstein convention)}$$

where the single magnetization $\boldsymbol{M}$ is the sum of all individual moments $\boldsymbol{\mu}_i$.

Surface Anisotropy. A finite-size magnetic body with a not too large bulk anisotropy will have a reorientation of the magnetization close to its surface in order to min-

imize the magnetostatic energy. This can be described by a surface anisotropy that is different from the bulk one. It originates from an abrupt change of symmetry at the interface between the bulk and free surface.

Helical Anisotropy. In a magnetic wire, the magnetization may prefer to lie radially in a plane perpendicular to the wire axis. This is the case of a radial easy axis (radial anisotropy). If it lies in the circumferential direction along the tangent to circles in the plane perpendicular to the wire axis, then we have circular anisotropy. However, if it lies in a plane tangent to the lateral surface of the wire (parallel to the wire axis) and making some angle different from 0° or 90° with any plane perpendicular to the wire axis, then we have helical anisotropy.

Anisotropy from Demagnetization

Demagnetization energy is expressed with coefficients that describe the demagnetization field $\boldsymbol{H}_\mathrm{d}$ inside a finite-size material and created by the bulk magnetization against an applied external magnetic field. The components of the demagnetization field (in the ellipsoidal case) are given by (with Einstein summation) $[\boldsymbol{H}_\mathrm{d}]_\alpha = -2\pi N_{\alpha\beta}\boldsymbol{M}_\beta$ (it is much like the depolarization field in the electrical case). The constant coefficients' ($N_{\alpha\beta}$) representation is valid only in the case the body has an ellipsoidal shape. The coefficients depend on the geometry of the material. Usually one has three positive coefficients along three orthogonal directions (N_{xx}, N_{yy}, N_{zz}) (assuming that the off-diagonal terms are all 0) for simple geometries such as wires, disks, thin films, and spheres. All three coefficients are positive and smaller than 1, and their sum is equal to 1. For a sphere, all three coefficients are equal to 1/3. For a thin film (or a disk), they are given by (0,0,1) if the z-axis is perpendicular to the film (or the disk) lying in the x–y plane. For an infinite length cylindrical wire with its axis lying along the z direction, the values are (1/2,1/2,0).

4.1.4 Traditional Types of Magnetism and Classes of Magnetic Materials

The main traditional types of magnetism are:

- Ferromagnetism
- Antiferromagnetism
- Ferrimagnetism
- Paramagnetism
- Diamagnetism.

However, a larger number of types are given in the review by *Hurd* [4.26] and with promising advances in materials research, we expect new emerging classes as described in Sect. 4.2 of this work dedicated to non conventional magnetic types (Table 4.1).

A ferromagnet is an assembly of magnetic moments interacting with a positive exchange integral that will minimize their energy by adopting a common parallel configuration with a net large value of the total magnetization. Such a definition is valid for localized magnetism and not for itinerant ferromagnets (such as Fe, Ni, and Co transition metals) since one does not have distinct localized moments to define an exchange integral. A ferromagnetic material (itinerant or localized) displays a characteristic hysteresis curve and remanence ($\boldsymbol{M} \neq 0$ for $\boldsymbol{H} = 0$) when one varies the applied magnetic field. When heated, the material loses, in general, this ordered alignment and becomes paramagnetic at the Curie temperature. Ferromagnets are usually metallic; however there are ferromagnetic insulators such as $CrBr_3$, EuO, EuS, and garnets [4.27, 37].

An antiferromagnet is made of an assembly of magnetic moments interacting with a negative exchange integral that will minimize their energy by adopting an antiparallel configuration. Again, such a definition is not valid for itinerant antiferromagnets (such as Cr and Mn transition metals) since one does not have distinct localized moments; thus an exchange integral cannot be defined. The net total magnetization is zero and, therefore, we do not have hysteresis. In the localized magnetism case, it is possible to consider the entire crystal as being made of two interpenetrating sublattices containing moments all parallel inside each sublattice but with the total magnetization of each sublattice opposite to one another. When heated, the material loses, in general, this alternately ordered alignment and becomes paramagnetic at the Néel temperature. Oxides are generally antiferromagnetic insulators except EuO [4.31, 37].

The intermediate order between ferro and antiferromagnets is a ferrimagnet used in microwave devices [4.16]. In the localized case, one considers the crystal as being made of two sublattices, as in the case of an antiferromagnet, with total magnetizations opposing one another. The magnitude of each magnetization is different, resulting in a net total magnetization (called uncompensated case) in contrast to an antiferromagnet.

Another intermediate case is a canted antiferromagnet, also called a weak ferromagnet. This is an assembly of magnetic moments alternating in direction and making a small inclination (canting). The small angle (Fig. 4.12d) between two neighboring moments $\boldsymbol{M}_i, \boldsymbol{M}_{i+1}$ results in a small ferromagnetic moment yielding weak ferromagnetism.

Canting originates from the Dzyaloshinski–Moriya exchange interaction, which is a vectorial exchange

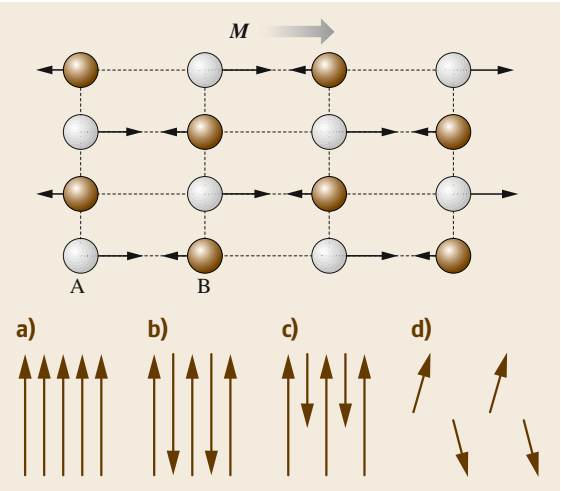


Fig. 4.12a–d Magnetic moment arrangement in an ferrimagnet. Total net magnetization is different from zero. **(a–d)** Comparison of the magnetic arrangements between a ferromagnet **(a)** an antiferromagnet **(b)**, a ferrimagnet **(c)**, and a weak ferromagnet **(d)**. (After [4.31])

interaction between two neighboring localized moments $\boldsymbol{M}_i$ and $\boldsymbol{M}_j$ of the form $D_{ij}\boldsymbol{M}_i \times \boldsymbol{M}_j$ that contrasts with the scalar ordinary exchange interaction of the form $A_{ij}\boldsymbol{M}_i \cdot \boldsymbol{M}_j$. It is responsible for canting (producing a small inclination) two neighboring antiferromagnetic moments that are usually antiparallel in an ordinary antiferromagnet. This is due to asymmetric spin–orbit effects [4.26].

An assembly of magnetic moments oriented at random with a zero total magnetic moment is called a paramagnet. A material of that sort does not display hysteresis or remanence when one varies the applied magnetic field. The magnetic field simply imposes some alignment to the randomly oriented moments. For a paramagnet, the susceptibility χ is positive, many orders of magnitude smaller than ferromagnets, and behaves as $1/(T - T_c)$, where T_c is the Curie temperature. Some materials never order magnetically, not even at low temperature (that implies $T_c = 0\,\mathrm{K}$) where order prevails. On the other hand, some materials are called superparamagnetic with no hysteresis and no remanence but with $T_c = \infty$. This does not mean that they are always ordered but their individual moment is so large (a classical moment corresponds to a quantum angular momentum $J = \infty$, resulting in $T_c = \infty$) that it takes an infinite temperature to destroy it completely [4.27, 37].

A material that combats the influence of a magnetic field by trying to maintain zero induction ($\boldsymbol{B} = 0$) in its midst is diamagnetic (a metal like copper is diamagnetic and superconductors are perfect diamagnets). Linguistically, the term diamagnetic is a misnomer since dia means across, implying the field gets across the material, whereas the opposite is true. When a diamagnetic substance such as a silicon crystal is placed in a magnetic field, the magnetization vector $\boldsymbol{M}$ in the material is in the opposite direction to the applied field. A negative susceptibility can be interpreted as the diamagnetic substance trying to expel the applied field from itself. A substance exhibits diamagnetism whenever the constituent atoms in the material have closed shells and subshells. This means that each constituent atom has no permanent magnetic moment in the absence of an applied field. Covalent crystals and many ionic crystals are typically diamagnetic materials because the constituent atoms have no unfilled subshells. Since the diamagnet tries to minimize the effect of $\boldsymbol{B}$, it expels field lines that can be exploited in magnetic levitation. A superconductor is a perfect diamagnet and a metal at high frequencies is partially diamagnetic since the applied field can penetrate it within the skin-depth layer only. For a diamagnet, the susceptibility χ is constant (no variation with temperature) and is slightly negative. Superconductors have $\chi = -1$ (since $\boldsymbol{B} = 0$ inside a superconductor below critical temperature due to the Meissner effect), whereas χ values of semiconductors are typically (Si: -0.26×10^{-6}, Ge: -0.58×10^{-6}, GaAs: -1.22×10^{-6} as taken from *Harrison* [4.38]) in $\mathrm{cm}^3/\mathrm{mol}$ at room temperature.

Materials with a relatively small coercive field (typically smaller than $1000\,\mathrm{A/m}$) employed in transformer cores and magnetic read heads are called soft magnetic materials. Typically, but not all, simple metals and transition metals and their compounds are soft. Permalloy, amorphous, and nanocrystalline alloys and some ferrites are soft. Amorphous materials are soft since their disordered structure does not favor any direction (no anisotropy energy) whereas nanocrystals possess anisotropy over a short length scale (albeit it can be enhanced with respect to the bulk counterpart). Softness is also measured by the maximum attainable permeability (Table 4.3, soft elements).

In the opposite case, a material with a relatively large coercive field (typically larger than $10\,000\,\mathrm{A/m}$) employed in permanent magnets, motors, and magnetic recording media (disks and tapes) is called a hard magnetic material. This means that stored data is not easily lost since it will require a large field to alter the magnetization value. Typically, but not all rare-earth metals, their compounds, and intermetallics are hard. There are also hard ferrites. Permanent magnets are used in power systems such as in relays and in motors and audio–video equipments such as headphones, videotapes, etc. (Table 4.1 and Fig. 4.11).

Table 4.3 Examples of soft magnetic materials and their hierarchy according to saturation magnetostriction coefficient λ_s. The main composition is successively based on Fe, NiFe, and finally Co. μ_{max} is the maximum differential permeability and H_c is the coercive field. $Fe_{80}\ B_{20}$ is also called Metglass 2605. $Fe_{40}\ Ni_{40}\ P_{14}\ B_6$ is also called Metglass 2826. The highest μ_{max} is attained by $(Fe_{0.8}\ Ni_{0.2})_{78}\ Si_8\ B_{14}$ reaching 2×10^6 after annealing. (After [4.37])

Alloy	H_c (mOe)	μ_{max} (50 Hz)	λ_s
$Fe_{80}\ B_{20}$	40	320 000	$\approx 30\times10^{-6}$
$Fe_{81}\ Si_{3.5}\ B_{13.5}\ C_2$	43.7	260 000	
$Fe_{40}\ Ni_{40}\ P_{14}\ B_6$	7.5	400 000	$\approx 10\times10^{-6}$
$Fe_{40}\ Ni_{38}\ Mo_4\ B_{18}$	12.5–50	200 000	
$Fe_{39}\ Ni_{39}\ Mo_4\ Si_6\ B_{12}$	12.5–50	200 000	
$Co_{58}\ Ni_{10}\ Fe_5\ (Si, B)_{27}$	10–12.5	200 000	$\approx 0.1\times10^{-6}$
$Co_{66}\ Fe_4\ (Mo, Si, B)_{30}$	2.5–5	300 000	

Ferrimagnetic ceramic-like alloys with composition $FeOFe_2O_3$ are called ferrites. The ferrite definition has been extended to $MOFe_2O_3$ with M being a divalent metal (note the trivalent state of Fe in Fe_2O_3). Ferrites are used in microwave engineering, recordings, etc., because of their very low eddy current loss [4.12] and they operate over a large frequency interval (kHz to GHz). Their resistivities may reach a ratio as high as 10^{14} with respect to typical metallic resistivities. They are made by sintering a mixture of metallic oxides $MOFe_2O_3$ with M = Mn, Mg, Fe, Zn, Ni, Cd, etc. Ferrite read heads are nonetheless limited to frequencies below 10 MHz as far as switching is concerned, and that is why several new types of read heads (thin films, anisotropic magneto-resistance (AMR), GMR, spin valves, magnetic tunnel junctions) have been developed in order to cope with faster switching (Chap. 49). Other conventional magnetic materials close to ferrites are discussed next.

Spinels

Alloys with a composition generalizing the ferrite one $(MO)_x(MO)_{1-x}\ Fe_2O_3$ with the structure of $MgAl_2O_4$ (origin of the word spinel) [4.16].

Garnets

Oxides with composition initially issued from the spinel family of the form $(3M_2O_3, 5Fe_2O_3)$ crystallize in the garnet cubic structure $(Ca_3Fe_2(SiO_4)_3)$. They are ferromagnetic insulators of the general formula $M_3Fe_5O_{12}$, where M is a metallic trivalent ion (e.g., $M = Fe^{3+}$). Garnets were been used in memory bubble technology, lasers, and microwave devices (because their ferromagnetic resonance linewidth versus field is small, being on the order of a fraction of an oersted while the resonance frequency is several tens of GHz, see *Wigen* [4.16]), on the basis of another generalization of the formula to $M_3Ga_5O_{12}$. For instance, $Gd_3Ga_5O_{12}$ called GGG (gadolinium gallium garnet) when produced as a thin film, a few microns thick, has perpendicular anisotropy with domains (bubbles) having their magnetization up or down (perpendicularly with respect to the thin film sample plane).

Thus a bit can be stored with a bubble and can be controlled with a small magnetic field. GGG is considered as one of the most perfect artificially made crystals since it can be produced with extremely few defects (less than 1 defect per cm^2). Another nomenclature called the [cad] notation is used with rare-earth iron garnets having the general formula $X_3Y_2Z_3O_{12}$. This notation means dodecahedral (c site is surrounded by 12 neighbors and represented by element X), octahedral (a site is surrounded by 8 neighbors and represented by element Y), and tetrahedral (d site is surrounded by four neighbors and represented by element Z). The most important characteristic of these garnets is the possibility to adjust their composition and, therefore, their magnetic properties according to selected substitutions on the c, a, or d sites. The element X is a rare earth, whereas Y and Z are Fe^{3+}. The magnetization is changed by placing nonmagnetic ions on the tetrahedral d site; increasing the amount of Ga^{3+}, Al^{3+}, Ge^{4+}, or Si^{4+} will decrease the magnetization. On the other hand, increasing the amount of Sc^{3+} or In^{3+} at the octahedral a site will increase the magnetization. Ion substitution can also be used to tailor other magnetic properties (anisotropy, coercivity, magnetostriction, etc.).

Garnets are typically grown with liquid-phase epitaxy with a growth speed that easily reaches 1 µm thickness in 1 min with a very high yield. They were not only in bubble materials but also in magneto-optical displays, printers, optical storage, microwave filters, and integrated optics components, etc. In spite of all these attractive properties, the ease of tunability, and the very high yield, problems soon arose with the access time limitation in bubble memories, since switching frequency was found to be bounded to less than about 10 MHz (see *Wigen* [4.16]; Chap. 49). More recently, garnets are being used in spin–charge interconversion as discussed in the last section.

4.2 Nonconventional Magnetism and Progress Toward Spintronics and Quantum Devices

4.2.1 Types of Exchange and Coupling in Magnetic Materials

In conventional magnetic materials, the strength of the magnetic interactions between two neighboring localized moments i and j (as in atoms/ions/molecules and rare-earths) is described by a direct exchange interaction. The latter is essentially a Coulomb (electrostatic) interaction between the electrons at the i and j sites. The word *exchange* is used since the corresponding overlap integral describing this interaction involves wavefunctions with permuted (exchanged) electron coordinates (in order to respect the Pauli exclusion principle). The exchange energy between sites i and j is given by $A_{ij}\boldsymbol{M}_i\boldsymbol{M}_j$, where $\boldsymbol{M}_i$ and $\boldsymbol{M}_j$ are, respectively, the magnetization at the i and j sites. When $A_{ij} > 0$ energy is minimized when the moments are parallel and when $A_{ij} < 0$ antiparallel configuration of the moments is favored. The energy arising from exchange over a distance r in the continuum limit is approximately AM^2/r^2.

Surrounding the localized moments i and j by an electron gas modifies the nature of their interaction into an RKKY (Ruderman–Kittel–Kasuya–Yoshida, see *Kittel* [4.27]) coupling. The oscillatory interaction occurring between the moments is mediated by the surrounding electron gas. It varies in three-dimensional (3-D) systems as $\cos(2k_F r)/r^3$, where r is the distance between the moments and k_F the Fermi wavevector of the electron gas. The finite value of k_F is responsible for the oscillatory behavior since it produces a cutoff in the Fourier transform [4.27]. It was discovered recently that a two-dimensional (2-D) counterpart of the RKKY interaction exists between two magnetic thin films separated by a metallic spacer [4.39]. The RKKY-like interaction between the two magnetic films across a metallic spacer is oscillatory with respect to the spacer thickness z. Thus it becomes possible to positively (ferro) or negatively couple magnetic layers by changing the thickness z of the sandwiched metallic spacer (Fig. 4.1 for typical lengths). This is extremely useful for thin-film devices (Chap. 49 and Table 4.1).

In planar magnetic technology, the interest is in magnetic thin films and their interactions. Information storage, sensing, spintronics, quantum computing, and other applications of magnetic thin-film devices are the main drive for understanding the nature and extent of magnetic exchange interactions and also coupling effects arising between magnetic 2-D layers in order to tailor appropriate devices. Novelty is expected since the device size is comparable to the interaction range (Fig. 4.1).

The coupling strength of the interaction between two neighboring magnetic films i and j can be modeled by a factor J_{ij}. It is similar to the exchange integral A_{ij} between two neighboring localized moments; however, it involves entire layers generally made from itinerant magnets and not single moments like with A_{ij}. The exchange interaction is of the form $J_{ij}\boldsymbol{M}_i.\boldsymbol{M}_j$, where $\boldsymbol{M}_i$ and $\boldsymbol{M}_j$ are the magnetization per unit surface of films i and j. The main interest of J_{ij} lies in the fact that its range is longer in reduced dimensions (one-dimensional (1-D) and 2-D) with respect to 3-D (for instance, an RKKY-like interaction between two magnetic films across a metallic spacer is oscillatory with a longer range than 3-D since it varies like $1/r^2$ instead of $1/r^3$) and its physics based on complex Fermi wavevectors [4.39] is entirely different from the standard RKKY interaction between localized moments [4.39]. The sign of the interaction depends on the thickness of the metallic spacer. Other types of exchange exist between films such as biquadratic or higher order ones with a generalized Heisenberg form $[J_{ij}\boldsymbol{M}_i \cdot \boldsymbol{M}_j]^n$ with $n \geq 2$ (see The entry related to exchange and coupling in multilayers in Table 4.1).

Interactions between two ferromagnetic (or other type of magnetic) films are not limited to metallic spacers; they may occur across an insulating or semiconducting spacer (Fig. 4.13). Magnetic pinholes generally

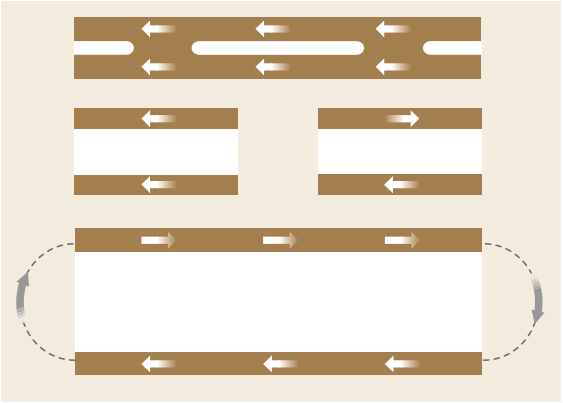


Fig. 4.13 *Top*: ferromagnetic coupling arising from pinholes existing across the spacer. *Middle*: exchange coupled layers that might be ferro (*left*) or antiferromagnetic (*right*) depending on the spacer thickness. *Bottom*: magnetostatic dipolar coupling between the top and bottom layers. (After [4.10])

Table 4.4 Surface energies γ in J/m^2 for magnetic and nonmagnetic materials, listed with respect to their atomic number for the low-energy cleavage surface. These are approximate values, which are difficult to measure in general and are dependent on surface orientation and reconstruction. (After [4.10])

Magnetic metal	Cr	Mn	Fe	Co	Ni	Gd				
γ (J/m^2)	2.1	1.4	2.9	2.7	2.5	0.9				
Transition metal	Ti	V	Nb	Mo	Ru	Rh	Pd	Ta	W	Pt
γ (J/m^2)	2.6	2.9	3.0	2.9	3.4	2.8	2.0	3.0	3.5	2.7
Simple or noble metal	Al		Cu		Ag		Au			
γ (J/m^2)	1.1		1.9		1.3		1.6			
Semiconductor	Diamond		Si		Ge		GaP		GaAs	
γ (J/m^2)	1.7		1.2		1.1		1.9		0.9	
Insulator	LiF		NaCl		CaF_2		MgO		Al_2O_3	
γ (J/m^2)	0.34		0.3		0.45		1.2		1.4	

mediate direct ferromagnetic coupling between neighboring multilayers. They are viewed as shorts across the insulating or semiconducting spacer. Néel proposed that conformal roughness (orange-peel model, see *Schulthess* and *Butler* [4.40]) at interfaces can result in ferromagnetic coupling for a moderate thickness of spacer material. Magnetostatic dipolar coupling occurs between the roughness features or between domain walls in the two magnetic layers. In this case, stray flux fields from walls in one film can influence the magnetization reversal process in the other. The Slonczewski loose-spin (see the exchange and coupling in multilayers in Table 4.1) model is based on angular momentum change as experienced by spin-polarized electrons tunneling across an insulating spacer, resulting in a magnetic exchange coupling. The magnitude and sign of the coupling oscillate, depending on the interfacial barrier properties (*Himpsel* et al. [4.10] and Table 4.1).

In the absence of a spacer, the exchange energy per unit surface of ferromagnetic and antiferromagnetic films gives birth to a new phenomenon called exchange bias (a shift of the hysteresis curve of the ferromagnetic film, see the entry in Table 4.1 related to exchange and coupling in multilayers). An important consequence of bias is pinning of the direction of the ferromagnetic moment used in low noise read heads because it hinders noisy (jerky) domain wall motion (Chap. 49). Exchange bias is also called anisotropic exchange and is still relatively poorly understood from the fundamental point of view despite the fact that it was discovered more than 50 years ago.

4.2.2 Thin Magnetic Films

A magnetic film is considered thin if its thickness is smaller than the (bulk defined) exchange length; in other words its thickness corresponds to the spatial extent of a single domain [4.10]. Magnetic thin films possess very attractive and distinct physical properties with respect to their bulk counterparts. For instance, reduced dimensionality, coordination and symmetry lead to magnetocrystalline anisotropy energies that are two or three orders of magnitude larger with respect to bulk (quenching of the orbital angular momentum as in transition metal ions is absent, see *Kittel* [4.27]). Saturation magnetization M_s is also enhanced with respect to bulk since the effective Bohr magneton number n_B per atom/molecule is larger in lower dimensions (for instance, in Nickel clusters n_B might reach 1.8, whereas in the bulk it is only 0.6). Couplings in thin films are also enhanced as in the case of the RKKY-like interaction across a metallic spacer. Coupling between two magnetic layers across a semiconducting spacer or an insulator leads to entirely new interactions, such as the Slonczewski loose-spin model or the Bruno quantum interference model. Other types of couplings or materials may be encountered besides exchange bias (between a ferro and an antiferromagnetic layer) such as spring magnets fabricated from alternating hard and soft thin films (Table 4.1).

Considering the growth of magnetic materials as thin films, we have to recall that these possess distinct physical properties with respect to the materials used in conventional microelectronics such as surface energies. A dominant role is played by surface free energies γ and the interface. They determine the growth modes in thermal equilibrium; the morphology of material B grown on material A depends on the balance between the free surface energies of substrate, overlayer, and interface [4.10]. Transition-metal-based magnetic materials exhibit a relatively high surface energy, owing to their partially filled *d* shell. Noble metal substrates have smaller surface energies and insulating substrates have even smaller ones (Table 4.4).

Additionally, when one performs epitaxial growth of materials, another concern is lattice matching for the different underlayers, as displayed in the tables below along with typical quantities of interest in representative magnetic materials (Table 4.5 and 4.6).

Table 4.5 Lattice-matched combinations of magnetic materials, substrates, and spacer layers. There are two main groups of lattice-matched systems with lattice constants close to 4.0 or 3.6 Å, respectively, after taking 45° rotations of the lattice or doubling the lattice constant. (After [4.10])

First group: magnetic metal $a\sqrt{2}$ (Å) [a (Å)]	Cr (bcc) 4.07 [2.88]	Fe (bcc) 4.05 [2.87]	Co (bcc) 3.99 [2.82]		
Simple or noble metal a (Å)	Al 4.05	Ag 4.09	Au 4.07		
Semiconductor $a/\sqrt{2}$ (Å) [a (Å)]	Ge 3.99 [5.65]	GaAs 4.00 [5.65]	ZnSe 4.01 [5.67]		
Insulator a (Å) [$a/\sqrt{2}$ (Å)]	LiF 4.02 [2.84]	NaCl 5.65 [3.99]	MgO 4.20 [2.97]		
Second group: Material a (Å)	Fe (fcc) 3.59	Co (fcc) 3.55	Ni (fcc) 3.52	Cu 3.61	Diamond 3.57

Table 4.6 Magnetic properties of Fe, Co, and Ni transition metals. Note that the domain wall (DW) thickness and energy are approximate. This is a complement to Table 4.2

Magnetic substance/property	Fe	Co	Ni
Exchange stiffness A (meV)	0.015	0.03	0.020
Anisotropy K_1 (J/m^3) at 300 K	4.8×10^4	45×10^4	-0.5×10^4
Lattice spacing (nm): a Lattice spacing (nm): c	0.29	0.25 0.41	0.35
DW thickness (nm) DW thickness (lattice parameters)	40 138	15 36	100 285
DW energy (J/m^2)	3×10^{-3}	8×10^{-3}	1×10^{-3}

4.2.3 Electronic Properties: Localized, Free, Itinerant Magnetism and Spin-Polarized Band Structure

Building a working device requires understanding not only magnetic properties but also electronic properties and their interplay. The new devices are expected to be made of a variety of magnetic (conventional and unconventional) materials as well as others already known in microelectronics. Insulating oxides (except EuO) and rare-earth compounds with well-localized external shell f electrons are the solid-state representative of atomic-like magnetism. Magnetism of atoms, ions, and molecules deals with well-defined, localized orbitals and individual moments arising from orbital, spin, or total angular momentum. When these moments get close to one another, as in the solid state, they interact Heisenberg wise: $A_{ij}\boldsymbol{M}_i.\boldsymbol{M}_j$ [4.31]. The latter is going to be altered by presence of the surrounding free electron gas. Therefore we have to understand magnetism in a free-electron gas, its counterpart arising from localized moments, and finally its nature when we have the hybrid case (itinerant magnetism) like in a transition metal (free s and localized d electrons). This problem being complicated due to its many-body nature, we rely upon a *one-electron approximation* and spin-polarized band structure [4.11].

In a free-electron gas, one assumes independent noninteracting electrons; therefore many-electron and nonlocal effects (arising from exchange) are not to be taken into account. Magnetism in this case is due to individual electron spin, and it is straightforward to establish so-called Pauli paramagnetism [4.27]. In addition, orbital effects give rise to Landau diamagnetism. The spin-polarized density of states is shifted by the field contribution $\pm\mu_B B$ depending on the spin orientation with respect to induction field $\boldsymbol{B}$. The magnetization in the free-electron Pauli case (simple metals) can be evaluated (at $T = 0$ K) from

$$M = \mu_B(N_\uparrow - N_\downarrow) = \frac{\mu_B}{2}\int_{-\mu_B B}^{E_F} \mathrm{d}EN(E+\mu_B B) - \frac{\mu_B}{2}\int_{\mu_B B}^{E_F} \mathrm{d}EN(E-\mu_B B) \tag{4.7}$$

that yields $M \approx \mu_B^2 N(E_F)B$, where μ_B is the Bohr magneton, $N(E)$ is the free electron, and $N_{\uparrow,\downarrow}(E) = N(E \pm \mu_B B)$ the spin-polarized density of states with respective occupations $N_{\uparrow,\downarrow}$. The Landau diamagnetism (orbital) contribution is a lot more complicated to evaluate, nevertheless it is exactly minus one-third the paramagnetic expression [4.27].

In transition metals (itinerant magnets), states derived from the 3d and 4s atomic levels are responsible for their physical properties. The 4s electrons are more spatially extended (of higher principal quantum number) and determine, for instance, compressibility, whereas the 3d states determine magnetic properties. The 3d electrons propagate throughout the material, hence the term *itinerant magnetism*. The spin-polarized band structure transition of metals is different from that of conventional electronic materials (see Figs. 4.14

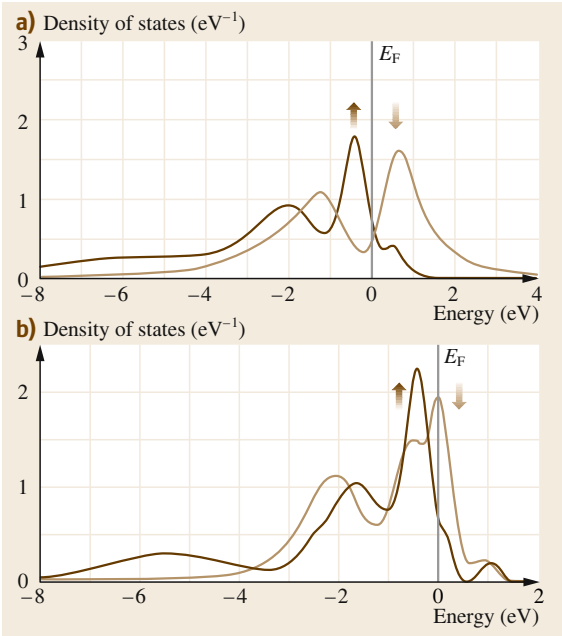


Fig. 4.14a,b Spin-polarized density of states for bcc Fe and fcc Ni. The bands for the spin up (*dark*) electrons are separated from the spin down (*light*) band by the exchange splitting energy (see text). The method used is superior to the LSDA method that yields a noisy density of states. The authors call it LDA + DMFT meaning local density approximation + dynamical mean field theory. (After [4.41])

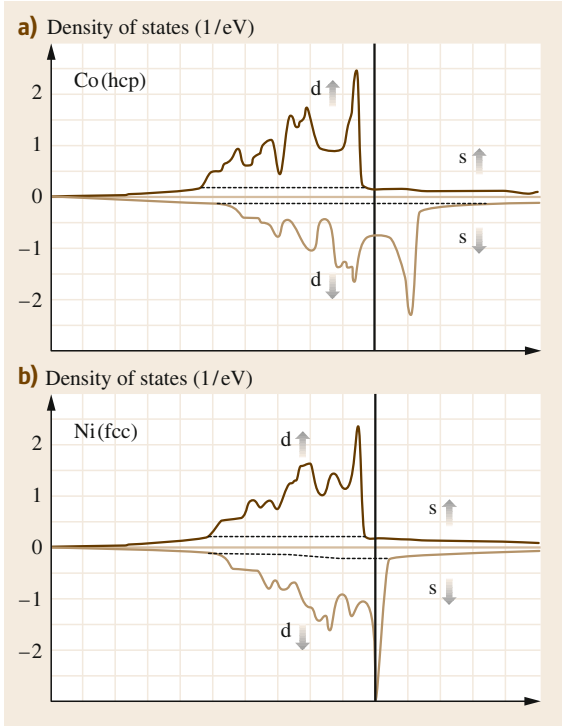


Fig. 4.15a,b Symmetric double-sided representation of the spin-polarized density of states for **(a)** Co and **(b)** Ni (obtained with the LMTO method, i. e., linearized muffin-tin orbitals) with the Fermi level indicated. (After [4.42])

and 4.15 for iron, nickel, and cobalt spin-polarized band structures).

Many metals have an odd number of electrons (in the atomic state), thus one would expect a magnetic material from unpaired spins; however, only five of them are actually magnetic (Cr, Mn, Fe, Ni, and Co are all transition metals with s–d hybridization). Note that copper is diamagnetic with $\chi = -0.77 \times 10^{-6}\,\mathrm{cm}^3/\mathrm{mol}$ at room temperature, Fe, Ni, and Co are ferromagnets, whereas Cr and Mn are antiferromagnets.

Magnetism and related properties in transition metal alloys are not straightforward to model owing to the electrons' itinerant nature. In solids, magnetism arises mainly from electrostatic (Coulomb) electron–electron interactions, namely the exchange interactions, and in magnetic insulators these can be described rather simply by associating electrons appropriately with particular atomic sites so that the Heisenberg exchange $A_{ij}\boldsymbol{M}_i \cdot \boldsymbol{M}_j$ can be used to describe the behavior of such systems. In metallic systems, it is not possible to distribute the itinerant electrons in this way, and such simple pairwise interactions between sites cannot be easily defined. Metallic magnetism is a complicated many-body effect and has attracted significant effort over a long period to understand and describe it [4.10]. A widespread approach is to map this problem onto one involving independent electrons moving in fields set up by all other electrons. It is this aspect which gives rise to the spin-polarized band structure that is often used to explain the properties of metallic magnets such as noninteger values in multiples of the Bohr magneton μ_B. In addition, the compositional structure plays a major role in itinerant magnetism, such as, for instance, in the Ni–Pt alloy. It is an antiferromagnet in the ordered case and ferromagnetic in the disordered case. Itinerant magnetism is comprised between two extreme limits: the localized (rare-earth and insulators) and the completely free electron case (more appropriate to alkalis) as described below. As far as transport is concerned, one may use the Stoner two-band model, which is a two-fluid model (like electrons and holes in semiconductors) with a population of electrons with spin-up and another of spin-down with the corresponding spin-polarized densities of states [4.10].

In ordinary crystalline nonmagnetic materials, the electronic band structure is built from the lattice (pe-

Table 4.7 Spin polarization in several ferromagnetic materials according to several authors. The discrepancies between the different results stem from the various approaches to estimate the density of states at the Fermi level and point to the difficulty of getting a unanimous figure. Note that for a half-metal like CrO_2 or NiMnSb, the polarization is 100%

	Meservey–Tedrow [4.43]	Moodera–Mathon [4.44]	Monsma–Parkin [4.45]
Fe	40%	44%	45%
Co	35%	45%	42%
Ni	23%	33%	31%
$Ni_{80}Fe_{20}$	32%	48%	45%
$Co_{50}Fe_{50}$	47%	51%	50%

Table 4.8 Magnetic splitting δk_{ex}, full width at half maximum $\delta k_\uparrow$ and $\delta k_\downarrow$ (± 0.01 Å^{-1}), and spin-dependent mean free paths $\lambda_\uparrow$ and $\lambda_\downarrow$ for NiFe and NiCr alloys. (After [4.46])

	δk_{ex} (Å^{-1})	$\delta k_\uparrow$ (Å^{-1})	$\delta k_\downarrow$ (Å^{-1})	$\lambda_\uparrow$ (Å)	$\lambda_\downarrow$ (Å)
Ni	0.14	0.046	0.046	> 22	> 22
$Ni_{0.9}Fe_{0.1}$	0.14	0.04	0.10	> 25	10
$Ni_{0.8}Fe_{0.2}$	0.14	0.03	0.22	> 33	5
$Ni_{0.93}Cr_{0.07}$	0.09	0.096	0.086	11	10
$Ni_{0.88}Cr_{0.12}$	≤ 0.05	0.12	0.11	8	9

riodic) potential seen by a single electron (having any state of spin) in the structure. Mathematically, it amounts to solving the Schrödinger equation corresponding to that potential and calculating its eigenfunctions (spin-independent wavefunctions). The eigenvalues versus wavevector are the ordinary bands. In a magnetic material, one has to include electronic exchange effects (originating from the Pauli exclusion principle applied to electrostatic interactions between pairs of electrons leading to nonlocal interactions) with explicit accounting for spin in the wavefunctions. Several methods exist such as the LSDA (local spin density approximation, see Fig. 4.1), which is a spin extension of the LDA (local density approximation) where one builds a local approximation to the exchange interaction from the local density of electrons [4.41]. The LSDA (Fig. 4.1) formalism provides a reliable description of magnetic properties of transition metal systems at low temperatures. It also provides a mechanism for generation of the noninteger effective number

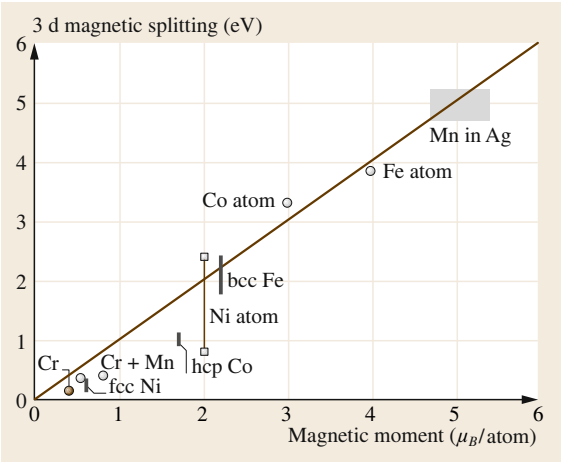


Fig. 4.16 Exchange splitting versus magnetic moment in the Bohr magneton μ_B per atom in 3d transition metals. The *diagonal line* is $1\text{eV}/\mu_B$ and the magnetic moment values (in Bohr magnetons) are given for the atom and the corresponding crystal. (After [4.10])

n_B of Bohr magnetons, together with a plausible account of the many-body nature of magnetic moment formation at $T = 0$ K. The obtained bands depend on spin as depicted in Figs. 4.14 and 4.15. From the spin-dependent band structure, one may obtain the spin-polarized density of states for each spin polarization (up $\uparrow$ or down $\downarrow$, also called majority and minority like in ordinary semiconductors). This nomenclature will eventually be confusing when one starts dealing with metals and semiconductors simultaneously. For the time being, this nomenclature is accepted (Table 4.7), as long as we are dealing solely with magnetic metals and one defines a new type of gap (originating from the exchange interaction) called exchange splitting (Fig. 4.16 and Table 4.8) or the (spin) gap between two spin-dependent bands (Fig. 4.14). This explains the existence of novel materials such as half-metals that ought to be contrasted with semimetals (graphite) where we have a negative electronic gap because of valence and conduction band overlap.

Half-metals (such as CrO_2 and NiMnSb) possess one spin-polarized band (up, for instance) full and the other (down) empty. These materials are very important for spintronics and specially for injecting spin-polarized carriers.

4.3 Spintronics and Quantum Information Devices

Presently, we are witnessing electronics extension toward the treatment of spin and charge instead of charge only, the realm of traditional electronics.

The reason spin becomes interesting and useful stems from the following. As device integration increases and feature length decreases toward the nanometer, the spin of individual carriers (electrons or holes) becomes a good quantum number. This means that the spin value is conserved over nanometric distances (spin diffusion length is typically 5–50 nm). This was not the case in microelectronics; consequently, spin can be used in the nanometer regime to carry useful information. Henceforth carriers transport energy, momentum, charge, and additionally spin. The quantum nature of spin leads to the natural consequence that quantum computers might be made successfully with spintronic components [4.28].

Spintronics, like microelectronics, is based on materials and their physical properties (mechanical, thermal, electrical, magnetic, etc.) as well as the know-how of growing/fabricating thin films and reduced dimensionality objects (quantum dots, quantum wires, quantum pillars, clusters, etc.). The different classical processing steps (epitaxial growth, deposition, oxidation, diffusion, doping, implantation, etching, passivation, thermal insulation, annealing, texturing, sputtering, patterning, etc.) that are well established in microelectronics have to be extended and adapted to tackle magnetism-related electronics. Magnetic interactions have an anisotropic vectorial character in contrast to scalar electrical interactions (based solely on charge) present in conventional microelectronic devices. Magnetism introduces the notions of anisotropy (magnitude, nature, and direction), coercivity, saturation magnetization, etc. that ought to be controlled during growth (magnetic field assisted growth or epitaxial growth has to be developed in order to favor magnetic anisotropy along one or several selected directions).

For instance, in some devices, there is a need to grow amorphous metallic magnetic layers (in order to have a very small anisotropy resulting in a magnetically soft layer), and these might be harder to grow than amorphous semiconductors or insulators (recall that a cooling speed of one million degrees per second is typically needed in order to produce an amorphous metal).

Overall, temperature is a very important parameter in magnetism since above the Curie (or Néel) temperature, ferromagnetism (antiferromagnetism) is lost.

Spin-polarized materials with novel properties emerge, such as half-metallic materials (CrO_2, NiMnSb, etc.), embodying carriers that are completely polarized spin wise (all up or all down).

Since we are dealing mostly with s–d hybridized transition metals, we expect at least three types of spin currents [4.22]: SPC (spin-polarized current) made of free (s-type) spin-polarized carriers, SWC (spin-wave current) carried by undulating localized spins (d-type) (with no dissipative Joule effect produced), and STT (spin-transfer torque) current stemming from the s–d

Table 4.9 Spintronic effects viewed as spin extension of ordinary unpolarized bulk and 2DEG electronics

Spintronic effect or device	Comments	Reference
Spin diode	Diode with spin polarized carriers	*Merchant* et al. [4.47]
Spin transistor	Bipolar magnetic junction transistor based on InMnAs (DMS) with magnetic field control of amplification	*Rangaraju* et al. [4.48]
Spin-Hall effect (SHE) and its inverse (ISHE)	Can be used to interconvert charge and spin currents	*Inoue* et al. [4.49]
Spin LED and lasers	LED and lasers with spin polarized carriers	*Holub* et al. [4.50]
Spin photovoltaic effect	High efficiency (50%) can be reached in DMS (AlP:Cr)	*Olsson* et al. [4.51]
Spin photogalvanic effect (SPGE)	Optical excitation of 2DEG system in quantum well structures	*Ganichev* et al. [4.52] and *Yang* et al. [4.53]
Spin Raman and nonlinear optical effects	Electrically induced Raman emission from 2DEG spin oscillator with nonlinear optical frequencies tuned by magnetic field gradient and electron density	*Nogaret* [4.54] (see also *Dietl* et al. [4.23])
Spin-Hall magnetoresistance (SMR)	Hall magnetoresistance with spin polarized carriers	*Sandweg* et al. [4.55]
Spin-Seebeck effect (SSE)	Thermal gradient effect with spin polarized carriers	*Bauer* et al. [4.56]
spin-Peltier effect (SPE)	Spin currents can be used for cooling or heating	*Flipse* et al. [4.57]
Spin-Gunn effect (SGE)	Instability at high electric fields against dynamic formation of spin-polarized current pulses	*Qi* et al. [4.58]
Spin breakdown	Spin counterpart of Avalanche/Zener breakdown	*Johnston–Halperin* et al. [4.59]

(double) exchange between free and localized carriers [4.60].

Other types of spin current can be generated by shining circularly polarized light [4.52, 53] over quantum well structures [4.52, 53] over quantum well structures exploiting a special effect originating from spin–orbit interaction in a 2DEG (two-dimensional electron gas) (Table 4.9) and called Rashba coupling [4.22, 23]. A spin current can also be produced with linearly polarized light [4.61] exciting antiferromagnetic materials such as NiO, Fe_2O_3 (hematite), or bismuth ferrite materials. The physical mechanism is based on a nonlinear optical effect called *shift current* (or photovoltaic drift current) produced by second-order interaction with the light electric field in noncentrosymmetric materials such as perovskites [4.61] ($BaTiO_3$, $LiNbO_3$, $KNbO_3$, etc.) of the form $J_i = \sigma_{ijk}E_jE_k$ (second-order extension of Ohm's law $J_i = \sigma_{ij}E_j$).

This is to be added to the wealth of magnetic interactions that can be based on para, dia, ferro, antiferro, or ferrimagnetic materials, which can be simultaneously electrically wise metallic (simple, transition, noble, or rare-earth), insulating, or semiconducting. As an example, multiferroics (see below) that are single-phase materials possessing interconnected electrical, magnetic, and elastic properties can be used as transducers in such multifunctional structures.

In analogy with ordinary electronics, spintronics is based on seven operations:

- *Spin injection*: How can a nonequilibrium density of spin-polarized carriers, that is, electrons with spin up $n_\uparrow$ (or down $n_\downarrow$) or holes with spin up $p_\uparrow$ (or down $p_\downarrow$) be created? Optical injection might be viewed as the spin extension of the Haynes–Shockley experiment. Spin injection can also be done with carbon nanotubes since they do not alter the spin state over large distances (Chap. 49).
- *Spin filtering*: Separating spin-polarized carriers is required in order to avoid spin-flips altering spin states. Thin ferromagnetic layers as in spin valves (Chap. 49) and chiral materials such as monolayers of double-stranded DNA molecules can be used [4.62].
- *Spin manipulation*: How can the spin and charge states of a nonequilibrium density of spin-polarized carriers be altered and controlled?
- *Spin coherence*: How can the spin and charge states of a nonequilibrium density of spin-polarized carriers be maintained over an important propagation length at room (and higher) temperature in the presence of perturbing magnetic stray fields (terrestrial and other)?
- *Spin accumulation*: How can spin-polarized carriers be coherently concentrated without destroying their spin coherence and without inducing spin flips that will destroy the resulting spin value?
- *Spin detection*: How can spin and charge of a nonequilibrium density of spin-polarized carriers be detected nondestructively?
- *Spin pumping*: This is the spin counterpart of charge pumping that occurs by application of a nonlinear ac signal to a semiconductor in order to achieve current rectification [4.63]. In the spin-polarized case, it is translated into the question of spin angular momentum transfer mode from magnetization precession to conduction electrons. It is experimentally feasible [4.64] with a ferromagnet (F) containing a precessing magnetization and a neighboring normal metal (N) into which a spin current is injected through the F–N interface. In summary, the question is related to ways of efficiently generating a continuous spin current from a time-evolving magnetization (viewed as a spin dynamo [4.65]).

In Table 4.9, we display the spin extension of several physical effects that are familiar in conventional electronics.

Many of the effects are simply the extension from a standard semiconductor or a metal to a ferromagnet. For instance, in the Seebeck effect, a thermal gradient ∇T applied to a metal induces an electric voltage along the direction of ∇T. In order to extend this phenomenon to the spin-Seebeck case, a thermal gradient ∇T is applied to a ferromagnet and a spin voltage is generated from thermal excitation of the magnetic moments in the ferromagnet. Those excitations are used to pump a spin current into a neighboring paramagnetic metal. Spin currents can be used for cooling or heating, and the spin-Peltier effect is based on the ability of the spin-up and spin-down currents to transport heat independently [4.57].

SMR (spin-Hall magnetoresistance) is even more interesting since it is related to conversion of charge into spin current (called interconversion) via SHE (spin-Hall effect). SHE implies that a current can be transferred between a noble metal (such as Pt or Au) and a magnetic insulator (such as yttrium iron garnet (YIG)) [4.22] making it the spin counterpart of the displacement current [4.22]. SHE leads to spin accumulations with opposite magnetizations at the edges of a magnetic material perpendicularly to the charge current and derives from the AHE (anomalous Hall effect) [4.22]. The latter is derived as follows. In an ordinary metal, a galvanomagnetic effect is written in terms of resistivity tensor ρ_{ij} that depends on the applied magnetic field $\boldsymbol{H}$ and contains two contribu-

tions: symmetric and antisymmetric like any rank-2 tensor.

Symmetric terms contain even powers of the applied magnetic field $\boldsymbol{H}$, whereas antisymmetric terms contain odd powers of $\boldsymbol{H}$. The lowest order antisymmetric terms are linear in $\boldsymbol{H}$ components and are termed Hall effect terms, whereas the lowest order symmetric terms are quadratic in $\boldsymbol{H}$ components and are termed magnetoresistive.

Let us suppose that ρ_{ij} in terms of field $\boldsymbol{H}$ in an ordinary metal has an expansion to lowest order (in $\boldsymbol{H}$ components) given by

$$\rho_{ij} = a_{ij} + b_{ijk}H_k + \cdots . \tag{4.8}$$

In contrast, when we have a ferromagnet, an additional term due to magnetization $\boldsymbol{M}$ should be added with respect to field $\boldsymbol{H}$ contribution. Thus one may extend the above expansion to

$$\rho_{ij} = a_{ij} + b_{ijk}H_k + c_{ijk}M_k + \cdots . \tag{4.9}$$

The ordinary Hall ($\boldsymbol{H}$ dependent) effect is thus extended to the AHE ($\boldsymbol{M}$ dependent) present in ferromagnets.

Hall terms are given by cross products $\boldsymbol{H} \times \boldsymbol{J}$ (ordinary Hall effect) or $\boldsymbol{M} \times \boldsymbol{J}$ (AHE), whereas magnetoresistance is related to scalar products between $\boldsymbol{M}$ and current density $\boldsymbol{J}$.

4.3.1 Magnetic Signal Processing Devices

Magnetic signal processing devices extend electronic signal processing to include magnetic field action, meaning that arithmetic, logic, memory, and other operations that traditionally belong to the realm of electronics will be performed by devices containing components that are sensitive to magnetic excitation.

An example given in Table 4.10 is an oscillator whose frequency is controlled by a magnetic field (this might be called magnetic agility) instead of an electrical voltage (as in ordinary electronic agility) despite the fact electrical control is generally preferred.

Fabrication of practical magnetic signal processing devices requires replacing magnetic field control by an electrical one (whenever possible) since stray magnetic fields might interfere with device operation. In addition, it is technically difficult to produce large magnetic fields over a nanometer length at present.

The simplest way of replacing magnetic field control is to trigger coupling between electrical and magnetic degrees of freedom by growing composite materials containing both ferromagnetic and ferroelectric elements. Coupling between magnetic and electrical degrees of freedom effects or magnetoelectric coupling is generated, for instance, through elastic interaction between a piezoelectric and a magnetostrictive substance [4.72].

Structures allowing possible magnetoelectric coupling are magnetic superlattices (also called magnonic or metamaterials), the magnetic analog of semiconducting superlattices or photonic structures (with varying dielectric constants) with modulated magnetization, anisotropy, or type (para, dia, ferro, antiferro, ferri).

Multiferroics are single-phase materials that respond to electrical, magnetic, and elastic excitations. They are more versatile than magnetoelectric composites; however their stoichiometry and growth critically affect the value substantially [4.72] of their magnetoelectric coefficient.

Dilute magnetic semiconductors (DMS) contain magnetic elements that are sensitive to magnetic fields and therefore make good candidates for magnetoelectric coupling (Table 4.10).

Substituting magnetic field control by an electrical one is possible in semiconductors through spin–orbit interaction since it allows generation and manipulation of carrier spins by an electric field [4.73].

Spin–orbit interaction has also been shown to allow electric control of spin-waves in single-crystal YIG

Table 4.10 Magnetic logic and signal-processing devices along with microwave ferrite devices covering applications ranging from L-band (0.5−1 GHz) to W-band (75−100 GHz)

Magnetic signal/device operation	Comments	Reference
Random access memory	Electrical read–write operations in a (Ga,Mn)As (DMS) memory cell	*Mark* et al. [4.66]
Spin-wave guiding	Signal processing with spin-waves in magnetic nanowires	*Choi* et al. [4.67]
Spin-based logic circuits	Design of reconfigurable logic circuits based on spin such as to be compatible with traditional electronic devices	*Khitun* et al. [4.68]
Spin oscillator	Microwave oscillation driven by STT current	*Kiselev* et al. [4.69]
Nonlinear parametric generation of subharmonics	Spectral manipulation of an external microwave signal	*Kobljanskyj* et al. [4.70]
Time-reversed wave-packet generation	Magnetic components allow tweaking of time-reversal symmetry	*Chumak* et al. [4.71]
Microwave ferrite device	Antennas, circulators, isolators, and inductors	*Wu* et al. [4.22]

waveguides, which paves the way to developing electrically tunable magnonic devices [4.74, 75].

Instead of coupling magnetic and electrical degrees of freedom or substituting magnetic field control by an electrical one, interconverting charge and spin degrees of freedom [4.22, 23] through spin currents, such as SHE described above, will ultimately lead to electric-field control (Tables 4.9 and 4.10).

Recently, voltage-controlled magnetism of Co thin films adjacent to Gd_2O_3 gate oxides was demonstrated [4.76]. Using gate voltages of a few volts, the anisotropy field as well as the saturation magnetization were altered, producing a large magnetic anisotropy energy change of up to $0.73\,erg/cm^2$. Voltage-induced reversible oxidation of the Co layer resulted from a large interfacial electric field and a high oxygen ion (O^{2-}) mobility in Gd_2O_3.

4.3.2 Magnetic Quantum Dot Arrays

Submicron quantum dots based on traditional semiconductor, insulator, or metallic materials are well developed [4.77] with many applications. They can be tailored with lithographical techniques and behave in a fashion similar to synthetic atoms. When arranged in arrays, they can be x–y addressed like random access memories or used in applications such as displays, cameras, sensors, etc.

The magnetic counterparts of quantum dot arrays are being developed with magnetic materials that possess a significantly larger number of possibilities stemming from the wide panel of magnetic materials (see *Hurd* [4.26] for a classification of magnetic materials prior to 1982 and *Kronmueller* and *Parkin* [4.78] for more recent developments).

Magnetic elements are specially adapted to scaling and scale integration since submicron-ordered magnetic dot arrays possess properties such as moment, susceptibility, coercivity, and remanence that are significantly enhanced with respect to their bulk counterparts, given the fact that they are closer to quantum atomic structures with augmented magnetic properties when fabricated on the nanometer scale.

In addition, present quantum dot array fabrication methods are fast and inexpensive with electrochemical template deposition methods, without the need for time-consuming processes based on sophisticated lift-off and deposition techniques borrowed from microelectronics operations, adapted and scaled down to tackle magnetic nanostructures.

In a disk-shaped magnetic quantum dot, several ordered structures can occur such as single domain (with in-plane or out-of-plane magnetization), multiple domains, and vortex excitations that are akin to solitons

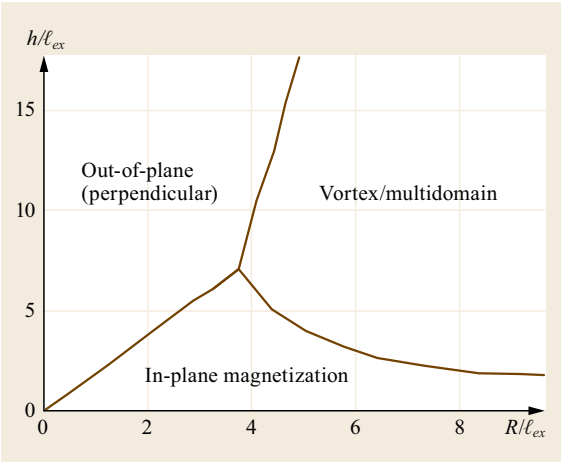


Fig. 4.17 Phase diagram h/ℓ_{ex} versus R/ℓ_{ex} displaying values for which vortex excitations appear in a disk-shaped quantum dot of thickness h and radius R. (After [4.79]) Occurrence of single domains with in-disk plane orientation or perpendicular to it as well as multiple domain regions are indicated

in higher dimension $D \geq 2$ (Fig. 4.17). A cylindrical magnetic quantum dot structure depends on two ratios: h/ℓ_{ex} and R/ℓ_{ex}, where h is the disk thickness and R the radius; ℓ_{ex} is the exchange length (Sect. 4.1).

The phase diagram displaying the regions of occurrence of single domains (with in-plane or perpendicular to plane orientation), multiple domains, and vortex excitations is depicted in Fig. 4.17.

Vortex structures are stable in a disk possessing a radius of several hundred nanometers and a thickness of a few tens of nanometers.

4.3.3 Magnetic Vortex Properties and Applications

A magnetic vortex possesses a central core of about 10 nm diameter with magnetization perpendicular to the x–y plane of the disk. The core magnetization is described by a polarization $p = \pm 1$ that can point in the $\pm z$ direction. Thus it can sense a vertical magnetic field. The core is surrounded by an x–y planar spin flow pattern making a flux-closure magnetization distribution in the plane of the disk. It is characterized by a chirality $C = \pm 1$ that can sense an in-plane magnetic field. Thus a single vortex, being able to sense all spatial components of a magnetic field, is the perfect ingredient of sensitive 3-D magnetometers.

The attractivity of vortex-based magnetic sensors operating at room temperature contrasts sharply with vortex-based superconducting magnetometers that must be cooled at low temperature (helium-4 for traditional

Table 4.11 Magnetic vortex applications in storage, memory, logic, and oscillator arrays

Application	Comments	Reference
Proposal for magnetic logic	Charge and spin operate simultaneously	*Cowburn* [4.80]
Nonvolatile RAM called V-MRAM and AV-MRAM	Not to be confused with vertical MRAM based on stacked ring-shaped GMR cells [4.20, 81]	*Bohlens* et al. [4.82], *VanWaeyenberge* [4.83], *Cowburn* [4.84], *Pigeau* et al. [4.85]
Data storage and memory	A vortex can store 2 bit simultaneously with its polarization and chirality	*Bohlens* et al. [4.82, 86]
Oscillator arrays	Coupled vortex oscillators in arrays of Permalloy disks	*Hanze* et al. [4.87]
Electrical switching of a vortex core in a magnetic disk	Read–write magnetic logic operations	*Yamada* et al. [4.88]
Vortex transistor	A new type of magnetic transistor	*Kumar* et al. [4.89]

superconductors and nitrogen for high-T_c superconductors). Moreover, *Yamada* et al. [4.88] were recently able to perform electrical switching of a vortex core (Table 4.11).

In many areas, magnetometry is needed, for instance, in biology, biomedicine, and geophysics where tensor gradiometry is required in order to determine all 3-D field components and their variation along different spatial directions with respect to Earth coordinates.

Detection of vortices is a very versatile field of research since electron holography [4.90] has been used, as well as magnetoresistance measurement in layered dots with a current traversing the structure along the axial direction (perpendicular to the vortex).

Some vortex detection techniques that perform identification of distinctive magnetization states, such as vortex nucleation and excitation (or switching [4.91]), domain wall motion, ferromagnetic resonance, and spin-wave excitations both in the quasistatic and dynamic regimes, are:

- Differential resistance dV/dI measurement [4.92, 93]
- Spin rectification effect [4.65, 94]
- Inductive methods [4.95, 96]
- RF/microwave transmission measurements [4.97–99]

Wang et al. [4.100] have shown that a resonant magnetic field with 30 mT amplitude stimulates strong axially symmetric magnetization oscillation and forces the vortex core to stay at the center of the disk. The compression of the vortex core by spin-wave excitation leads to fast core reversal [4.100] at 602×10^{-12} s rate. More recently, *Dong* et al. [4.101] managed to confine vortex cores in spin-wave potential wells and obtained core switching times well below 200×10^{-12} s.

The detection range of vortex-sensing magnetometers, being in the kOe range, is too strong for fields usually encountered in biology, biomedicine, and geophysics. Thus we need a conditioning stage for the sensor in order to match its own detection range to another one usually encountered in the above-mentioned areas. The conditioning is made with IMC (integrated magnetic concentrators) that at present possess a gain of 10−100 order of magnitude [4.102].

This gain, defined as the ratio between detectable field in the sensor (kOe) and an external one originating from biological or geophysical sources, is currently being increased steadily with the use of softer materials and better geometrical design to attain at least 1000 [4.103].

In micro-Hall magnetometry where magnetic nanodisks are placed on top of Greek cross-shaped [4.104] Hall plate sensors, vortex motion is detected from their stray-field sensed by Hall voltage measurements.

Another route to vortex detection is through the magnetoelectric effect in layered structures consisting of multiferroic elements or composite elements consisting of alternating piezoelectric layers and giant magnetostrictive elements.

A host of applications exist for vortices (and antivortices their magnetic mirror counterpart) ranging from data storage to magnetic RAM, logic, and information processing devices. More recently a new type of transistor [4.89] based on vortices has been introduced (Table 4.11).

References

4.1 R. Boll, H. Warlimont: IEEE Trans. Magn. **MAG-17**, 3053 (1981)
4.2 E.C. Stoner, E.P. Wohlfarth: Phil. Trans. Roy. Soc. Lond. A **240**, 599 (1948)
4.3 C.L. Platt, M.R. McCartney, F.T. Parker, A.E. Berkowitz: Phys. Rev. B **61**, 9633 (2000)
4.4 J.C. Slonczewski: Phys. Rev. B **39**, 6995 (1989)
4.5 N.C. Koon: Phys. Rev. Lett. **78**, 4865 (1997)
4.6 C. Tannous, J. Gieraltowski: J. Mat. Sci. Mat. Electron. **15**, 125 (2004)
4.7 R. White: IEEE Trans. Magn. **MAG-28**, 2482 (1992)
4.8 F. Schatz, M. Hirscher, M. Schnell, G. Flik, H. Kronmueller: J. App. Phys. **76**, 5380 (1994)
4.9 M.J. Dapino, R.C. Smith, F.T. Calkins, A.B. Flatau: J. Intell. Mater. Syst. Struct. **13**, 737 (2002)
4.10 F.J. Himpsel, J.E. Ortega, G.J. Mankey, R.F. Willis: Adv. Phys. **47**, 511 (1998)
4.11 H.F. Jansen: Phys. Today, Magnetoelectronics **48**, 50 (1995), Special April Issue
4.12 J.B. Goodenough: IEEE Trans. Magn. **MAG-38**, 3398 (2002)
4.13 H.J. Richter: J. Phys. D: App. Phys. **32**, 147 (1999)
4.14 P. Farber, M. Hörmann, M. Bischoff, H. Kronmueller: J. App. Phys. **85**, 7828 (1999)
4.15 P. Coeure: J. Phys. Coll. C-6 **46**, 61 (1985)
4.16 P.E. Wigen: Thin Solid Films **114**, 135 (1984)
4.17 O. Gutfleisch: J. Phys. D **33**, R157 (2000)
4.18 G. Burkard, D. Loss: Europhys. News **33**, 166 (2002)
4.19 H. Hauser, L. Kraus, P. Ripka: IEEE Inst. Meas. Mag. **4**, 28 (2001)
4.20 J. Zhu, Y. Zheng, G. Prinz: J. Appl. Phys. **87**, 6668 (2000)
4.21 I. Zutic, J. Fabian, S. Das Sarma: Rev. Mod. Phys. **76**, 323 (2004)
4.22 M. Wu, A. Hoffmann: Solid state physics. In: *Recent Advances in Magnetic Insulators – From Spintronics to Microwave Applications*, Vol. 64, ed. by M. Wu, A. Hoffmann (Academic Elsevier, New York 2013)
4.23 T. Dietl, D.D. Awschalom, M. Kaminska, H. Ohno: Spintronics. In: *Semiconductors and Semimetals*, Vol. 82, ed. by E.R. Weber (Academic, London 2008)
4.24 H. Kronmueller: J. Magn. Magn. Mat. **140–144**, 25 (1995)
4.25 J.L. Simonds: Phys. Today **48**, 26 (1995)
4.26 C.M. Hurd: Contemporary Phys **23**, 469 (1982)
4.27 C. Kittel: *Introduction to Solid State Physics*, 6th edn. (Wiley, Hoboken 1986)
4.28 M.A. Nielsen, I.L. Chuang: *Quantum Computation and Quantum Information* (Cambridge Univ. Press, New York 2000)
4.29 M.R. Fitzsimmons, S.D. Bader, J.A. Borchers, G.P. Felcher, J.K. Furdyna, A. Hoffmann, J.B. Kortright, I.K. Schuller, T.C. Schulthess, S.K. Sinha, M.F. Toney, D. Weller, S. Wolf: J. Magn. Magn. Mater. **271**, 103 (2004)
4.30 S. Chikazumi: *Physics of Ferromagnetism*, International Series of Monographs on Physics, 2nd edn. (Oxford Univ. Press, Oxford 1997)
4.31 S.O. Kasap: *Principles of Electronic Materials and Devices*, 2nd edn. (McGraw-Hill, New York 2001)
4.32 L.O. Chua: *Introduction to Non-Linear Network Theory* (McGraw-Hill, New York 1969)
4.33 B.K. Chakrabarti, M. Acharyya: Rev. Mod. Phys. **71**, 847 (1999)
4.34 A. Hubert, R. Schäfer: *Magnetic Domains*, 1st edn. (Springer, Berlin Heidelberg 1998)
4.35 A.P. Malozemoff, J.C. Slonczewski: *Magnetic Domains in Bubble-like Materials* (Academic, London 1979)
4.36 D. Buntinx: *Influence of Shape Auisotropy and Exchange Bias on the Domain Structure and Magnetoresistance of Mesoscopic Ferromagnets*, Ph.D. Thesis (Katholieke Univ., Leuwen 2003)
4.37 D. Jiles: *Introduction to Magnetism and Magnetic Materials*, 2nd edn. (Chapman and Hall, London 1991)
4.38 W.A. Harrison: *Electronic Structure and the Properties of Solids* (Freeman, San Francisco 1980)
4.39 P. Bruno: Phys. Rev. B **52**, 411 (1995)
4.40 T.C. Schulthess, W.H. Butler: J. App. Phys. **87**, 5759 (2000)
4.41 A.I. Lichtenstein, M.I. Katsnelson, G. Kotliar: Phys. Rev. Lett. **87**, 067205 (2001)
4.42 A. Barthelemy: GDR Pommes CNRS (Aspet, France) (2001)
4.43 R. Meservey, P.M. Tedrow: Phys. Rep. **238**, 173 (1994)
4.44 J. Moodera, G. Mathon: J. Magn. Magn. Mat. **200**, 248 (1999)
4.45 D.J. Monsma, S.S.P. Parkin: Appl. Phys. Lett. **77**, 720 (2000)
4.46 K.N. Altmann, N. Gilman, J. Hayoz, R.F. Willis, F.J. Himpsel: Phys. Rev. Lett. **87**, 137201 (2001)
4.47 C.A. Merchant, N. Markovic: Phys. Rev. Lett. **100**, 156601 (2008)
4.48 N. Rangaraju, J.A. Peters, B.W. Wessels: Phys. Rev. Lett. **105**, 117202 (2010)
4.49 J. Inoue, H. Ohno: Science **309**, 2004 (2005)
4.50 M. Holub, P. Bhattacharya: J. Phys. D **40**, R179 (2007)
4.51 P. Olsson, C. Domain, J.-F. Guillemoles: Phys. Rev. Lett. **102**, 227204 (2009)
4.52 S.D. Ganichev, E.L. Ivchenko, S.N. Danilov, J. Eroms, W. Wegscheider, D. Weiss, W. Prettl: Phys. Rev. Lett. **86**, 4358 (2001)
4.53 C.L. Yang, H.T. He, L. Ding, L.J. Cui, Y.P. Zeng, J.N. Wang, W.K. Ge: Phys. Rev. Lett. **96**, 186605 (2006)
4.54 A. Nogaret: Phys. Rev. Lett. **94**, 147207 (2005)
4.55 C.W. Sandweg, Y. Kajiwara, A.V. Chumak, A.A. Serga, V.I. Vasyuchka, M.B. Jungfleisch, E. Saitoh, B. Hillebrands: Phys. Rev. Lett. **106**, 216601 (2011)
4.56 G.E.W. Bauer, E. Saitoh, B.J. van Wees: Nat. Mater. **11**, 391 (2012)
4.57 J. Flipse, F.L. Bakker, A. Slachter, F.K. Dejene, B.J. van Wees: Nat. Nanotechnol. **7**, 166 (2012)
4.58 Y. Qi, Z.-G. Yu, M.E. Flatté: Phys. Rev. Lett. **96**, 026602 (2006)
4.59 E. Johnston-Halperin, D. Lofgreen, R.K. Kawakami, D.K. Young, L. Coldren, A.C. Gossard, D.D. Awschalom: Phys. Rev. B **65**, 041306 (2002)

4.60 J.C. Slonczewski: J. Magn. Magn. Mat. **195**, L261 (1999)
4.61 S.M. Young, F. Zheng, A.M. Rappe: Phys. Rev. Lett. **110**, 057201 (2013)
4.62 B. Gohler, V. Hamelbeck, T.Z. Markus, M. Kettner, G.F. Hanne, Z. Vager, R. Naa-man, H. Zacharias: Science **331**, 894 (2011)
4.63 G. Platero, R. Aguado: Phys. Rep. **395**, 1 (2004)
4.64 Y. Tserkovnyak, A. Brataas, G. Bauer: Phys. Rev. Lett. **88**, 117601 (2002)
4.65 Y.S. Gui, N. Mecking, X. Zhou, G. Williams, C.-M. Hu: Phys. Rev. Lett. **98**, 107602 (2007)
4.66 S. Mark, P. Durrenfeld, K. Pappert, L. Ebel, K. Brunner, C. Gould, L.W. Molenkamp: Phys. Rev. Lett. **106**, 057204 (2011)
4.67 S. Choi, K.-S. Lee, K.Y. Guslienko, S.-K. Kim: Phys. Rev. Lett. **98**, 087205 (2007)
4.68 A. Khitun, M. Bao, K.L. Wang: IEEE Trans. Mag. **44**, 2141 (2008)
4.69 S.I. Kiselev, J.C. Sankey, I.N. Krivorotov, N.C. Emley, R.J. Schoelkopf, R.A. Buhrman, D.C. Ralph: Nature **425**, 380 (2003)
4.70 Y. Kobljanskyj, G. Melkov, K. Guslienko, V. Novosad, S.D. Bader, M. Kostylev, A. Slavin: Sci. Rep. (2012) doi:10.1038/srep00478
4.71 A.V. Chumak, V.S. Tiberkevich, A.D. Karenowska, A.A. Serga, J.F. Gregg, A.N. Slavin, B. Hillebrands: Nat. Comm. (2010) doi:10.1038/ncomms1142
4.72 C.-W. Nana, M.I. Bichurin, S. Dong, D. Viehland, G. Srinivasan: J. Appl. Phys. Appl. Phys. Rev. **103**, 031101 (2008)
4.73 D. Awschalom, N. Samarth: APS Phys. **2**, 50 (2009)
4.74 T. Liu, G. Vignale: Phys. Rev. Lett. **106**, 247203 (2011)
4.75 X. Zhang, T. Liu, M.E. Flatté, H.X. Tang: Phys. Rev. Lett. **113**, 037202 (2014)
4.76 C. Bi, Y. Liu, T. Newhouse-Illige, M. Xu, M. Rosales, J.W. Freeland, O. Mryasov, S. Zhang, S.G.E. te Velthuis, W.G. Wang: Phys. Rev. Lett. **113**, 267202 (2014)
4.77 Y. Alhassid: Rev. Mod. Phys. **72**, 895 (2000)
4.78 H. Kronmueller, S. Parkin: *Handbook of Magnetism and Advanced Magnetic Materials* (Wiley, New-York 2007)
4.79 W. Scholz, K.Y. Guslienko, V. Novosad, D. Suess, T. Schrefl, R.W. Chantrell, J. Fidler: J. Mag. Mag. Mat. **266**, 155 (2003)
4.80 R. Cowburn: J. Magn. Magn. Mater. **242–245**, 505 (2002)
4.81 G.A. Prinz: Science **282**, 1660 (1998)
4.82 S. Bohlens, B. Krüger, A. Drews, M. Bolte, G. Meier, D. Pfannkuche: Appl. Phys. Lett. **93**, 142508 (2008)
4.83 B. VanWaeyenberge, A. Pazic, H. Stoll, K.W. Chou, T. Tyliszczak, R. Hertcl, M. Fáhnle, H. Brùckl, K. Rolt, G. Reiss, I. Neudeker, D. Weiss, C.H. Back, G. Schútz: Nature **444**, 461 (2006)
4.84 R.P. Cowburn: Nat. Mater. **6**, 255 (2007)
4.85 B. Pigeau, G. de Loubens, O. Klein, A. Riegler, F. Lochner, G. Schmidt, L.W. Molenkamp, V.S. Tiberkevich, A.N. Slavin: Appl. Phys. Lett. **96**, 132506 (2010)
4.86 J. Thomas: Nat. Nanotech. **2**, 206 (2007)
4.87 M. Hanze, C.F. Adolff, M. Weigand, G. Meier: Appl. Phys. Lett. **104**, 182405 (2014)
4.88 K. Yamada, S. Kasai, Y. Nakatani, K. Kobayashi, H. Kohno, A. Thiaville, T. Ono: Nat. Mater. **6**, 270 (2007)
4.89 D. Kumar, S. Barman, A. Barman: Nat. Sci. Rep. **4**, 4108 (2014)
4.90 M. Heumann, T. Uhlig, J. Zweck: Phys. Rev. Lett. **94**, 077202 (2005)
4.91 Y. Liu, S. Gliga, R. Hertel, C.M. Schneider: Appl. Phys. Lett. **91**, 112501 (2007)
4.92 S. Kasai, Y. Nakatani, K. Kobayashi, H. Kohno, T. Ono: Phys. Rev. Lett. **97**, 107204 (2006)
4.93 W. Lin, J. Cucchiara, C. Berthelot, T. Hauet, Y. Henry, J.A. Katine, E.E. Fullerton, S. Mangin: Appl. Phys. Lett. **96**, 252503 (2010)
4.94 N. Mecking, Y.S. Gui, C.-M. Hu: Phys. Rev. B **76**, 224430 (2007)
4.95 U. Queitsch, J. McCord, A. Neudert, R. Schafer, L. Schultz, K. Rott, H. Bruckl: J. Appl. Phys. **100**, 093911 (2006)
4.96 I. Neudecker, K. Perzlmaier, F. Hoffman, G. Woltersdorf, M. Buess, D. Weiss, C.H. Back: Phys. Rev. B **73**, 134426 (2006)
4.97 H. Zhang, A. Hoffmann, R. Divan, P. Wang: Appl. Phys. Lett. **95**, 232503 (2009)
4.98 V. Novosad, F.Y. Fradin, P.E. Roy, K.S. Buchanan, K.Y. Guslienko, S.D. Bader: Phys. Rev. B **72**, 024455 (2005)
4.99 A. Vogel, A. Drews, T. Kamionka, M. Bolte, G. Meier: Phys. Rev. Lett. **105**, 037201 (2010)
4.100 R. Wang, X. Dong: Appl. Phys. Lett. **100**, 082402 (2012)
4.101 X. Dong, Z. Wang, W. Wang: Appl. Phys. Lett. **104**, 112413 (2014)
4.102 W.C. Griffith, R. Jimenez-Martinez, V. Shah, S. Knappe, J. Kitching: Appl. Phys. Lett. **94**, 023502 (2009)
4.103 C. Schott, R. Racz, A. Manco, N. Simonne: IEEE J. Solid-State Circuits **42**, 2923 (2007)
4.104 M. Rahm, J. Biberger, D. Weiss: Micro-hall-magnetometry. In: *Nano-Physics and Bio-Electronics: A New Odyssey*, ed. by T. Chakraborty, F. Peeters, U. Sivan (Elsevier, New York 2002) p. 257

5. Defects in Monocrystalline Silicon

Wilfried von Ammon, Andreas Sattler, Gudrun Kissinger

The aggregation of instrinsic point defects (vacancies and Si interstitials) in monocrystalline silicon has a major impact on the functioning of electronic devices. While agglomeration of vacancies results in the formation of tiny holes (so-called *voids*, around 100 nm in size, which have almost no stress field), the aggregation of Si interstitials exerts considerable stress on the Si matrix, which, beyond a critical size, generates a network of dislocation loops around the original defect. These dislocation loops are typically microns in size. Consequently, they are much more harmful to device functioning than vacancy clusters. However, the feature size in electronic devices has now shrunk below the 100 nm scale, meaning that vacancy aggregates are also no longer acceptable to many device manufacturers.

This chapter is intended to give an introduction to the properties of intrinsic point defects in silicon and the nucleation and growth of their aggregates. Knowledge in this field has grown further over the last decade. It is now possible to accurately simulate the aggregation process so that the defect behavior of semiconductor silicon can be precisely tailored to the needs of the device manufacturer. Additionally, the impact of various impurities on the aggregation process is elucidated.

5.1 **Technological Impact of Intrinsic Point Defects Aggregates** 112
5.2 **Thermophysical Properties of Intrinsic Point Defects** 113
5.3 **Aggregates of Intrinsic Point Defects** ... 115
5.3.1 Experimental Observations 115
5.3.2 Theoretical Model: Incorporation of Intrinsic Point Defects . 118
5.3.3 Theoretical Model: Aggregation of Intrinsic Point Defects ... 120
5.3.4 Effect of Impurities on Intrinsic Point Defect Aggregation 123
5.4 **Formation of OSF Ring** 127
References 129

As the feature size continues to shrink in device industry and now approaches a design rule of 7 nm, a thorough understanding of defect behavior in bulk silicon is of utmost importance. Three major defect types relevant to device performance have been identified:

1. Vacancy aggregates (known as *voids*, which usually have a size of less than 150 nm)
2. Si interstitial clusters embedded in a network of dislocation loops, each of which extend over several microns (L-pits)
3. Large grown-in oxygen precipitates.

The latter generate stacking faults (OSF) during wafer oxidation. The voids form in the center of the crystal, while L-pits are observed in the outer region. The two concentric defect regions are usually separated by a small OSF ring. The type of defect that develops in the growing crystal is determined by a simple parameter: the ratio of the pull rate to the temperature gradient at the growth interface. In industry, crystals with only one type of defect – voids – are produced almost exclusively. The formation and behavior of voids has been studied intensively and is accurately described by current theoretical models. As the feature size is now smaller than the void size in standard silicon crystals, the growth of so-called *perfect silicon* with almost no detectable defects may be adopted. Furthermore, the doping of crystals with impurities like nitrogen or carbon is being widely investigated. These impurities can significantly reduce the defect size, but they may also have harmful effects, such as enhancing the generation of OSFs. Some models have recently been proposed which may allow us to predict some of the effects of impurities.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_5

5.1 Technological Impact of Intrinsic Point Defects Aggregates

Single intrinsic point defects in silicon – vacancies and interstitials – have not been found to have any negative impact on device performance so far. However, if they aggregate into clusters they can be even detrimental to device functionality. This is also true of extended defects like dislocations. When silicon wafer technology was in its formative years, much of the work devoted to improve wafer quality focused on controlling dislocation density in the silicon crystals, as it was not possible to grow dislocation-free crystals. However, with the introduction of dislocation-free crystal growth processes into mass production, the issue of extended dislocations in relation to bulk silicon quality vanished. As the feature size decreased and the demand for higher device performance increased, it soon became apparent that intrinsic point defects and their aggregation during the cool-down phase of the crystal growth process were having an increasingly negative impact on device performance and yield. Historically, one of the first serious challenges in this regard was the aggregation of Si interstitials in floating zone (FZ)-grown crystals, which results in a local network of dislocation loops (secondary defects), so-called *A-swirls* [5.1] or L-pits [5.2]. Although the diameters of these dislocation loops are only a few microns, they are large enough to generate hot spots in the space charge regions of high-power devices [5.3, 4]. In the second half of the 1980s, the industry began to encounter problems with the early breakdown of the gate oxide in memory devices based on Czochralski (CZ)-grown silicon [5.5]. After intensive gate oxide integrity (GOI) investigations, it was found that the root cause of the gate oxide degradation was tiny micro holes – voids – which were formed by vacancy aggregation during crystal growth [5.6, 7]. Each void is thermally stabilized by an oxide layer present on its inner surface. After wafer polishing, the voids show up as dimples or laser light scattering (LLS) defects on the wafer surface, causing a local thinning of the gate oxide [5.8, 9]. Voids are considerably smaller (less than 150 nm) than A-swirl defects, and so their impact on device performance is only apparent if the location of a void coincides with that of an active element, such as a transistor. In addition, most of these defective transistors can be repaired due to the built-in redundancy of memory chips. Consequently, vacancy aggregates are tolerable for many devices, so long as their density is insignificant compared to those of device process-induced defects. This should be contrasted with A-swirls or L-pits, which always result in permanent device damage due to their large sizes.

Empirically, it has been found that gate oxides 40–50 nm thick are most susceptible to void defects [5.9]. Thinner oxides show higher GOI yields, and when the thickness drops below 5 nm the influence of voids on the GOI yield disappears [5.10, 11]. However, as the feature size continues to shrink, additional adverse effects have been identified, such as shorts between trench capacitors and lack of device reliability [5.12, 13]. As the design rule becomes equal to or less than the void size, these problems are expected to aggravate and device manufacturers may have to switch to materials with extremely small defect sizes or those that contain virtually no defects.

There are three main ways to achieve silicon with no harmful intrinsic point defect aggregates. The first is to grow silicon crystals in a regime where Si interstitials and vacancies are incorporated in equal concentrations (Sect. 5.3.2), resulting in almost complete mutual annihilation of point defects (so-called *perfect* or *pure* silicon) [5.14–17]. The method inevitably involves lower pull rates and very tight control over crystal growth parameters, which yields considerably lower throughput and higher costs, in particular for 300 mm crystal growth.

The second approach is the growth of nitrogen-doped crystals with very fast pull rates (high cooling rates) and subsequent high-temperature ($\approx$1200 °C) wafer annealing [5.18]. Nitrogen doping in conjunction with a fast pull rate decreases the vacancy aggregate size (Sect. 5.3.3 and Sect. 5.3.4) [5.18, 19], meaning that they are easy to dissolve using a high-temperature wafer treatment. Void annihilation first requires the dissolution of the inner oxide layer, which in turn, necessitates the outdiffusion of oxygen. Thus, annealed wafers only exhibit a near-surface defect-free region $\approx$10 µm in depth, which is, however, sufficient for device manufacturing. Annealed wafers also take advantage of the notable mechanical strengthening effects of nitrogen doping [5.20–22], which helps to suppress slippage generation during high-temperature treatment. One very recent development is rapid thermal wafer annealing at a temperature of around 1300 °C. At this temperature, outdiffusion of oxygen is not necessary because the oxygen concentration is usually below that required for oxygen solubility, so the inner oxide layer dissolves throughout the bulk and the voids collapse. This process yields silicon of a similar quality to that resulting from perfect silicon crystal growth.

The third method is the well-known epi wafer approach.

All of these methods require rather precise defect engineering in order to obtain the properties demanded by the device industry, except in the case of pp + epi wafers. Here, the high boron concentration of the substrate suppresses intrinsic point defect aggregation (Sect. 5.3.4) and enhances oxygen precipitation in the bulk. Therefore, this wafer type not only provides a defect-free epi layer, but also provides high slip resistances and superior internal gettering (impurity removal) abilities of metallic contaminants.

5.2 Thermophysical Properties of Intrinsic Point Defects

Understanding intrinsic point defect aggregation undoubtedly requires rather exact knowledge of their respective thermophysical properties.

The intrinsic point defects – vacancies and Si interstitials – can exist in different configurations. Generally, six localized point defect configurations of high symmetry are considered: the vacancy and the split-vacancy on the one hand, and the tetrahedral, the hexagonal, the bond-centered and the [100] split or dumbbell Si interstitial on the other [5.23]. While the localized configuration works rather well for vacancies, theoretical calculations strongly favor an extended configuration of lower symmetry for Si interstitials [5.24, 25]. The extended self-interstitial model was originally proposed to explain the high pre-exponential factor in the coefficient of self-diffusion, and this model now has support from theoretical calculations [5.23, 26, 27]. According to theory, the high value of the pre-exponential factor results from the multitude of self-interstitial configurations with similar energies and the significant lattice relaxations that accompany some of these configurations.

Vacancies and Si interstitials can also exist in various charged states (such as $V^{2+}, V^{+}, V^{0}, V^{-}, V^{2-}$), and at the high temperatures ($> 1000\,^{\circ}\mathrm{C}$) where point defects start to aggregate all states should be present in dynamic equilibrium [5.28]. Due to this equilibration, it is not meaningful to assign a specific charge to vacancies and Si-interstitials, respectively. However, atomistic calculations show that the charged states are much higher in energy and so their populations should be negligible. So far, there is no indication that charged states have any impact on defect aggregation and so they are not considered in current defect nucleation models.

Unfortunately, it is generally not possible to observe intrinsic point defects directly and, consequently, their thermophysical properties cannot be measured directly either. Thus, indirect approaches must be used that involve fitting defect concentrations along with many other parameters. Various experimental systems have been used to infer the thermophysical properties of point defects. The most common of these are metal diffusion experiments where a metallic contaminant such as zinc, gold or platinum is introduced into the bulk via high-temperature drive-in diffusion [5.29–31]. The diffusion rate of the metallic impurity, which is easily detectable using standard methods, is related to the mobility and the concentration of intrinsic point defects (kick-out and Frank–Turnbull mechanism), which provides a way to indirectly probe the behavior of the point defects. These experiments yield good estimates for the product $C_{\mathrm{eq}}D$ of the equilibrium concentration C_{eq} and the diffusivity D for self-interstitials I and vacancies V, respectively. The following values are derived from zinc diffusion results [5.32]

$$C_{\mathrm{I}}^{\mathrm{eq}} D_{\mathrm{I}} = 1.5 \times 10^{26} \exp\left(\frac{-4.95\,\mathrm{eV}}{k_{\mathrm{B}} T}\right) \mathrm{cm}^{-1}\,\mathrm{s}^{-1}$$

and

$$C_{\mathrm{V}}^{\mathrm{eq}} D_{\mathrm{V}} = 1.3 \times 10^{23} \exp\left(\frac{-4.24\,\mathrm{eV}}{k_{\mathrm{B}} T}\right) \mathrm{cm}^{-1}\,\mathrm{s}^{-1}\,.$$

Another frequently used experimental method is the defect analysis of CZ crystals grown with varying pull rates. In this case, the observables are the dynamic responses of the oxidation-induced stacking fault (OSF) ring and the interstitial–vacancy boundary as a function of changes in crystal growth process conditions [5.2, 33–35] (Sect. 5.4). These observables have been quantitatively correlated to intrinsic point defect distributions in crystals and, therefore, can be used to derive thermophysical properties [5.36]. Of particular importance is the complementary nature of crystal growth and metal diffusion experiments with regard to parametric sensitivity. Its sensitivity to self-interstitial and vacancy competition implies that these experiments are suitable for determining some pre-exponential coefficients. The metal diffusion experiments, which can be carried out over a wide range of temperatures, are particularly useful for probing activation energies.

Rapid thermal annealing (RTA) experiments with fast cooling rates have been carried out with silicon wafers to investigate the influence of the free vacancy concentration on oxygen precipitation [5.17, 37]. By heating the wafers to above $1150\,^{\circ}\mathrm{C}$, detectable concentrations of Frenkel pairs are created in the wafer bulk. The wafer is then held at this temperature for around 10 s to allow the self-interstitials to diffuse

out, resulting in a near-equilibration of both interstitials $[C_I^{eq}(T)]$ and vacancies $[C_V^{eq}(T)]$. Following this rapid equilibration, the wafer is rapidly cooled (at rates of $\approx 100\,^{\circ}C/s$) to freeze in the vacancy excess which can subsequently be probed via its impact on oxygen precipitation. This technique is especially useful for gaining information on $[C_V^{eq}(T) - C_I^{eq}(T)]$ at various temperatures [5.38]. Quenching experiments at different holding temperatures are also reported for hydrogen-doped samples, which allow us to quantitatively measure the VH_4 complex [5.39]. From these results, the vacancy formation energy was deduced as $\approx 4\,eV$ with relatively good accuracy.

Experimental data that relate to the recombination coefficient of vacancies and Si interstitials are scarce [5.40, 41]. Nevertheless, there is general agreement that recombination between vacancies and Si interstitials is so fast that the product of the concentrations of both species is always in thermal equilibrium. The exact value is therefore not relevant to subsequent discussion here.

There have been numerous attempts to compute the thermophysical properties of point defects directly by atomistic simulations. The most important aspect of these atomistic calculations is the accuracy of the assumed interatomic interaction. As the ab initio approach, which explicitly considers electronic interactions, is extremely computationally expensive and therefore limited to small systems and zero temperature, most simulations are carried out using empirical potentials, such as the Stillinger–Weber or Tersoff potential functions, discarding explicit representations of electronic interactions. An excellent overview of this field is given by *Sinno* [5.42]. A parameter set that yields rather good results for defect aggregation is shown in Fig. 5.1. It is clear that the difference between the vacancy and Si interstitial concentrations at the melting point is only around 30% in favor of vacancies. On the other hand, vacancies diffuse more slowly than Si interstitials. These two properties are of great relevance to the peculiar defect behavior of silicon.

In the past, fitted values for equilibrium concentrations and diffusivities inferred from various experimental data often spread over many orders of magnitude, meaning that they were of little help. However, in recent years, theoretical methods and the accuracy of experimental data have significantly improved and the predicted values are now in much better agreement with those gleaned from experiments.

Recently, another impact factor became obvious which may have led to different fit results of the thermophysical data of the past. As described in detail later in Sect. 5.3.2, the solubility of the intrinsic point defects is obviously dependent on the thermal stress which is

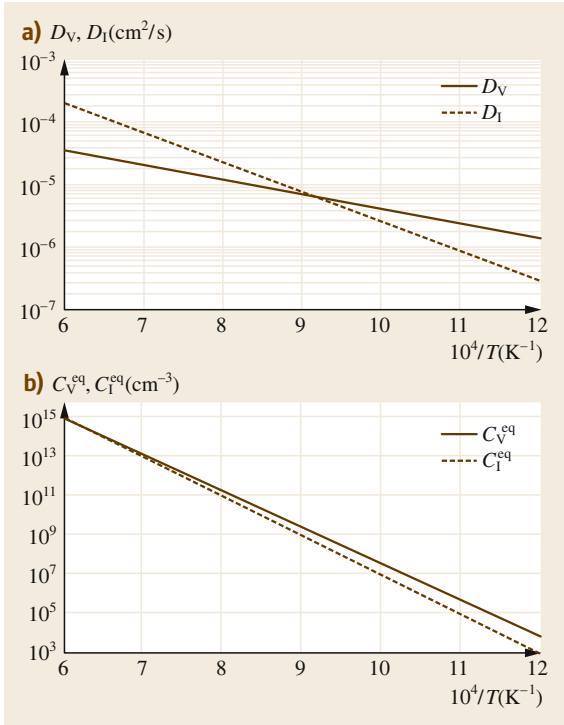


Fig. 5.1a,b Diffusivities and equilibrium concentrations of vacancies and Si interstitials, respectively. (After [5.43])

imposed on the crystal during growth [5.44]. In addition, the diffusivities of the intrinsic point defects are believed to be affected by stress, too. *Sueoka* et al. used density functional theory (DFT) calculations to determine the stress dependence of the formation enthalpies of vacancies and interstitials [5.45]. The best agreement between experimental and simulated values was obtained for plane thermal stress. Sueoka's ab initio calculations lead to the following dependencies of the formation enthalpies of vacancies (H_f^V) and interstitials (H_f^I) on the average plane stress (σ_{ave})

$$H_f^V\,(eV) = 0.308 \times \sigma_{ave}$$

and

$$H_f^I\,(eV) = -0.069 \times \sigma_{ave}$$

with σ_{ave} given in GPa. The relationship is valid in the range $-1\,GPa < \sigma_{ave} < 0\,GPa$, and the average plane stress is defined as $\sigma_{ave} = (\sigma_X + \sigma_Y)/3$. The effect of stress on the formation energies seems small, but these small differences have a notable influence on ξ_{tr} (see (5.1) in Sect. 5.3.2), the parameter which determines the incorporation of the predominant defect type in a growing crystal.

In another work, *Sueoka* et al. [5.46] used DFT calculations to obtain the formation energies of vacancies and interstitials at all sites within a sphere around a dopant atom in silicon [5.46]. This was done for substitutional p-type (B and Ga), neutral (C, Ge, and Sn), and n-type (P, As, Sb, and Bi) dopants. In this way, thermal equilibrium concentrations of vacancies and interstitials were obtained for various dopants and doping levels. For details see [5.46].

5.3 Aggregates of Intrinsic Point Defects

5.3.1 Experimental Observations

As pointed out in Sect. 5.1, the dominant intrinsic point defect aggregates in silicon single crystals are voids and L-pits/A-swirls. Figure 5.2 shows a transmission electron microscope (TEM) image of a void with a 5 nm oxide layer on the inner surface. Oxidation of the inner surface is not observed in FZ crystals due to their very low oxygen contents. The thickness of the oxide layer decreases with increasing cooling rate of the growing crystal. A strain field is generally not observed around the void. Usually, voids exhibit octahedral morphologies, but they can change to platelet or even rodlike shapes if the crystal is doped with nitrogen (Sect. 5.3.4). Under standard growth conditions, twin voids consisting of two partial octahedral voids are predominantly observed [5.47, 48]. At lower oxygen contents and higher cooling rates of the growing crystals, single octahedral voids are preferentially formed [5.49, 50]. Typical sizes range from 70–200 nm depending on the crystal growth conditions. Historically, different notations were introduced for this type of defect depending on their delineation or the detection technique used:

- Crystal originated particle (COP) [5.51]
- Laser light scattering tomography defect (LSTD) [5.2]
- Flow pattern defect (FPD) [5.52]
- D-defects [5.53].

Today, it is generally accepted that they all denote the same defect type: vacancy aggregates.

The TEM image of an L-pit is depicted in Fig. 5.3. The network of perfect dislocation loops characteristic of this defect type is clearly visible. The core of the defect is a self-interstitial aggregate which forms an extrinsic stacking fault [5.55, 56]. When the stacking fault reaches a critical size, the strain exerted on the crystal lattice is relaxed by the generation of dislocation loops. It is believed that the B-swirl observed in FZ crystals [5.57, 58] is most likely related to self-interstitial aggregates with sizes below the critical limit [5.59, 60]. While the B-swirl can be annihilated by appropriate high-temperature treatment, the L-pit/A-swirl is extremely stable.

In the case of epi wafers, voids at the substrate surface are covered by the epilayer and no defects are generated in the epilayer. However, if the layer thickness is very thin (< 1 µm), a dimple remains at the epi-

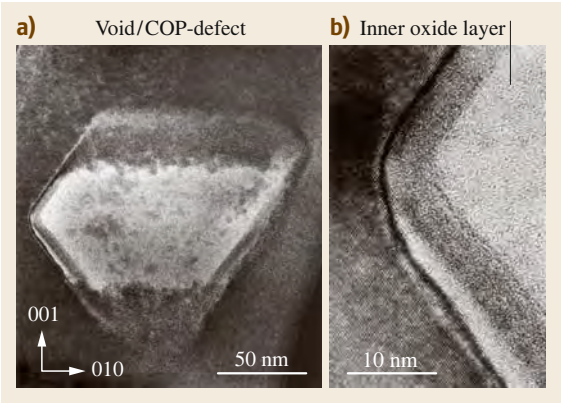


Fig. 5.2a,b TEM picture of a void/COP **(a)**. The inner surface of the void is cover by an oxide layer **(b)**. (After [5.54])

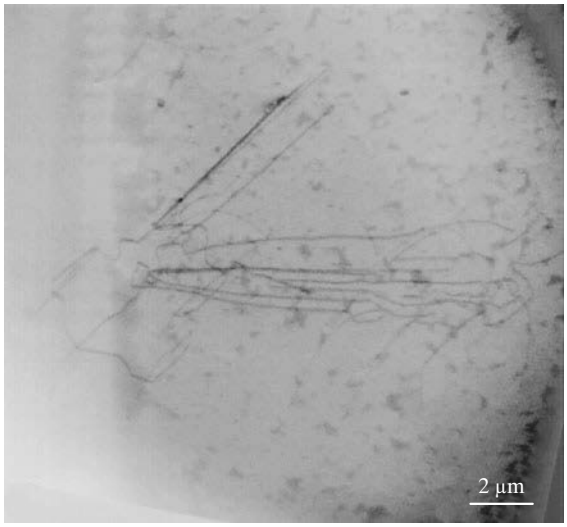


Fig. 5.3 TEM picture of an L-pit. (After [5.55])

layer surface which is detectable by laser light scattering [5.61]. In contrast to voids, L-pits always generate clearly visible and harmful defects in the epilayer [5.62] and are therefore unacceptable to device manufacturers. It is well-established that the pull rate V and the axial temperature gradient G at the growth interface of the growing crystal have the biggest influence on the defect types that develop in the growing crystal [5.33, 34, 63] and their spatial [5.35] as well as their density/size distributions [5.64]. At high pull rates, vacancy-related defects (voids) are observed over the entire crystal volume (Fig. 5.4). When the pull rate is reduced, oxidation-induced stacking faults (OSFs) develop in a small ring-like region near the crystal rim [5.65]. If the pull rate is decreased further, the ring diameter shrinks and L-pits are detected in the area outside the OSF ring [5.2]. No voids are found in this outer region. Thus, the OSF ring obviously represents the spatial boundary between vacancy- and self-interstitial-type defects. At a critical pull rate, the void region and the OSF ring disappear completely and only L-pits are detected. In FZ crystals with their inherently low oxygen contents, no OSF ring is observed as a boundary; instead a defect free-zone is observed as the boundary [5.66].

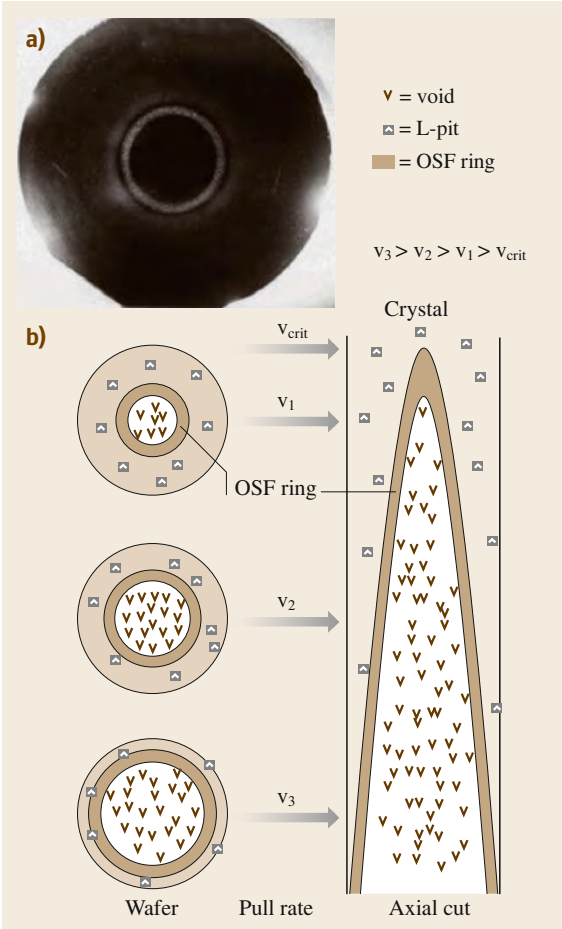


Fig. 5.4a,b Variation in the radial defect behavior of a CZ crystal as a function of pull rate. **(b)** shows a wafer from the crystal part grown with medium pull rate. The defects were delineated by Secco etching

The radial position of the OSF ring can be approximately described by the equation [5.35]

$$\frac{V}{G(r)} = \xi_{tr} = 1.34 \times 10^{-3}\,\mathrm{cm^2\,K^{-1}\,min^{-1}}\,.$$

Here r is the radial position. Recently, somewhat higher values for ξ_{tr} are proposed [5.44]. Thus, the parameter V/G controls the type of grown-in defect: if $V/G > \xi_{tr}$ vacancy aggregates develop, while for $V/G < \xi_{tr}$ Si interstitial-related defects are observed. This was first recognized by *Voronkov* in 1982 [5.63], who derived the following relation between ξ_{tr} and the thermophysical properties of intrinsic point defects

$$\xi_{tr} = \frac{(\Delta H_I + \Delta H_V)\left(D_I C_I^{eq}(T_0) - D_V C_V^{eq}(T_0)\right)}{2k_b T_0^2\left(C_V^{eq}(T_0) - C_I^{eq}(T_0)\right)}\,. \tag{5.1}$$

A further important parameter for the control of grown-in defects is the thermal history of the crystal. While its influence on the type of defect that develops in the growing crystal is negligible, detailed investigations have revealed a strong effect of cooling rate on the defect density/size distribution. For a crystal cooling in the temperature range between roughly 1000 and 1100 °C, there seems to be an exponential relation between the void density N_{void} and the dwell time t (Fig. 5.5) [5.67]. The defect density data in Fig. 5.5 were derived from GOI measurements that can detect much smaller defect sizes than delineation techniques based on etching and/or laser light scattering. The correlation between the various detection methods down to sizes of 0.12 µm was found to be rather good [5.18]. The exact upper value of the above temperature range depends on the crystal growth conditions and will be shown (Sect. 5.3.3) to coincide with the void nucleation temperature. The width of the selected temperature interval, on the other hand, does not have much impact on the above empirical relation. As will be discussed below, the relevant temperature interval is rather small [5.68]. Hence, the dwell time might just as well be replaced by the inverse of the cooling rate at the nucleation temperature.

It was further observed that the average void size increases as dwell time increases (with slower cooling rates) and vice versa (Fig. 5.6) [5.64, 69]. Another

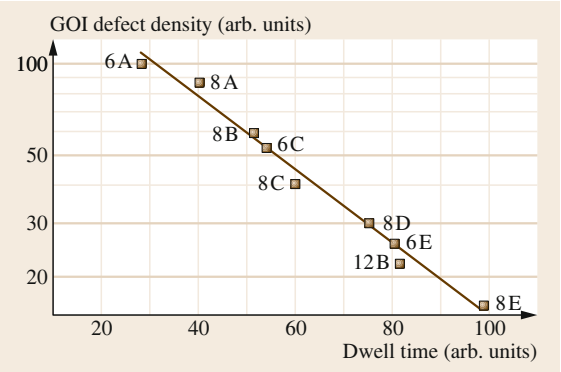


Fig. 5.5 Void density as derived from GOI measurements as a function of the dwell time of the growing crystal at ≈1100 °C. The *straight line* denotes simulation results. The various data points relate to different growth processes. (After [5.64])

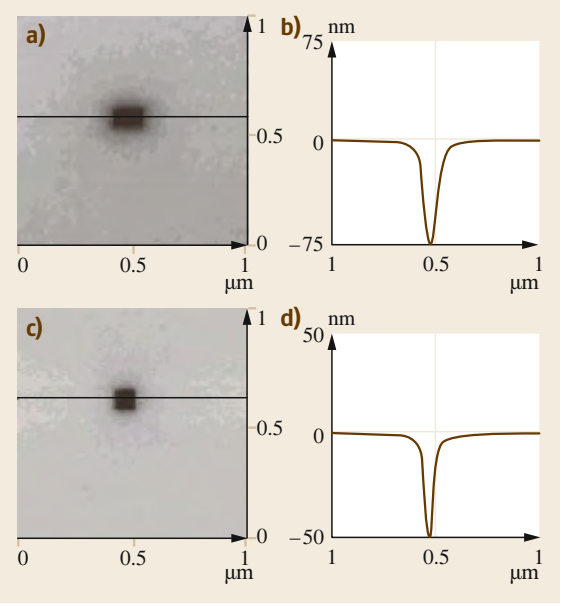


Fig. 5.6a–d Atomic force microscopy measurements of voids at the wafer surface **(a,c)** and their respective depth profiles **(b,d)**. The depth of the surface void is rather close to its original size in the crystal. **(a)** and **(b)** refer to a slowly cooled and **(c)** and **(d)** to a rapidly cooled crystal. (After [5.61])

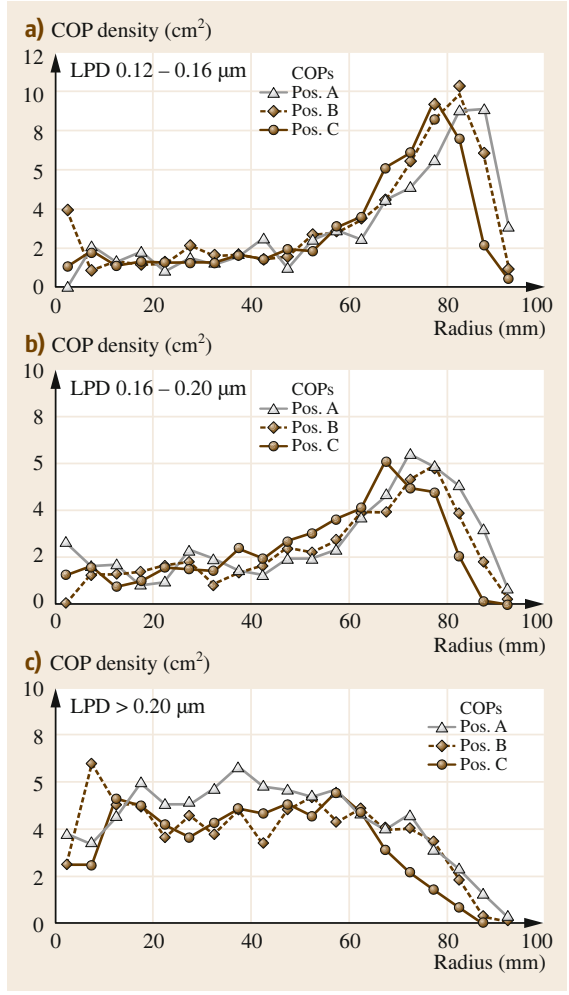


Fig. 5.7a–c Radial variation in the density/size distribution of voids/COPs across a wafer as measured by laser light scattering. The light point defect size (LPD) correlates well to the actual void size. The positions *A*, *B* and *C* relate to different axial crystal positions

striking feature is the characteristic change in the defect density/size distribution across the crystal diameter (Fig. 5.7) [5.70]. While large voids of low density prevail in the crystal center, the distribution gradually shifts to small sizes and high density towards the boundary of the void region. As the cooling rates of the crystal center and rim are almost the same, this remarkable variation in the density/size distribution must be related to the radial inhomogeneity of $V/G(r)$. Unfortunately, similar data are not yet available for L-pits. The behavior of this defect type is more complex, as small interstitial aggregates which do not generate dislocation loops are not detected as L-pits.

Valuable information about the defect nucleation temperature is obtained from transient growth experiments where the pull rate is reduced to zero and the growth process is halted when a certain crystal length has been reached [5.71–73]. After a certain amount of time has elapsed, growth is resumed and the crystal is grown to full length. During the halting period, the various crystal positions are at different temperatures ac-

cording to the axial temperature gradient of the crystal. By analyzing the axial defect behavior of the as-grown crystal, one can identify the temperature ranges where specific defect types nucleate and grow in size. The results of those experiments clearly demonstrated that the nucleation temperature of voids lies between 1000 and 1100 °C and varies with the growth conditions. It was also demonstrated that the nucleation and growth period of the voids occurs over a temperature interval of less than 50 K [5.74]. Void growth is then obviously stopped by the formation of the inner oxide layer. The oxide layer continues to grow down to relatively low temperatures. The scarce data available for L-pits indicate a similar temperature range for nucleation [5.75].

5.3.2 Theoretical Model: Incorporation of Intrinsic Point Defects

The incorporation of intrinsic point defects is the result of two competing fluxes at the growth interface: a vacancy and an interstitial flux which are both directed into the crystal (Fig. 5.8). Each of the two fluxes consists of two components. The first flux component is driven by the advancing growth interface, which generates a convective flux proportional to the pull rate V, and the second is driven by the vacancy/interstitial recombination behind the interface, which creates a concentration gradient and, in turn, a Fickian diffusion flux. The latter is proportional to G. If the pull rate is low, then the Fickian diffusion dominates over the convective flux. As self-interstitials diffuse significantly faster than vacancies, the self-interstitial flux wins over the vacancy flux. During crystal cooling, V–I recombination virtually eliminates the vacancies and the surviving self-interstitials are driven into supersaturation and finally aggregate. On the other hand, at fast pull rates the convective flux dominates and, due to the larger equilibrium concentration of vacancies compared to that of self-interstitials, more vacancies flow into the crystal, resulting in vacancy aggregates. A similar change in the prevailing defect type is obtained when G is varied. The above model, originally proposed by *Voronkov* [5.63], yields the previously mentioned parameter V/G which determines the defect type.

In standard growth processes, G always exhibits a notable increase from the center towards the crystal rim, which (at sufficiently low pull rates) results in a vacancy excess in the center, while Si interstitials dominate at the crystal rim, as indicated in Fig. 5.9. Figure 5.10 illustrates how the radial position of the crossover between ξ_{tr} and $V/G(r)$ and hence the boundary between the void and L-pit regions is shifted when V is varied.

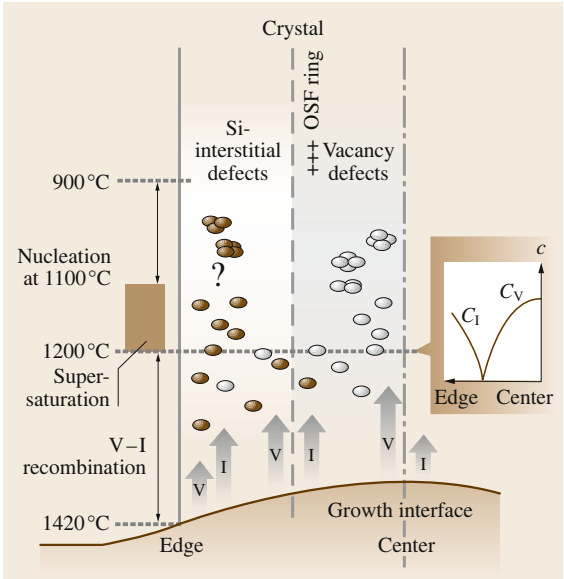


Fig. 5.8 Schematic picture of the incorporation of vacancies (V) and Si interstitials (I), respectively, into the growing crystal. The *insert on the right* shows the radial variation of the remaining species after V–I recombination has ceased and supersaturation starts

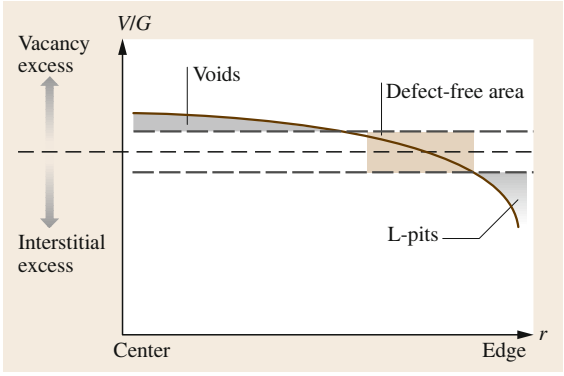


Fig. 5.9 Schematic radial variation in defect types as a function of the radial variation of V/G

The exact value of ξ_{tr} is still under discussion, because G has to be taken directly at the interface where the thermal field in the crystal changes very rapidly [5.35, 76]. Therefore, it has not been possible to accurately measure G at the growth interface [5.77]. Fortunately, computer simulations with sufficiently refined meshes permit a rather accurate calculation of G. Based on these calculations, a value of $0.13\ \mathrm{mm^2\ min^{-1}\ K^{-1}}$ has been determined from growth experiments, with hot zones of different Gs [5.33], which, however, is currently debated again.

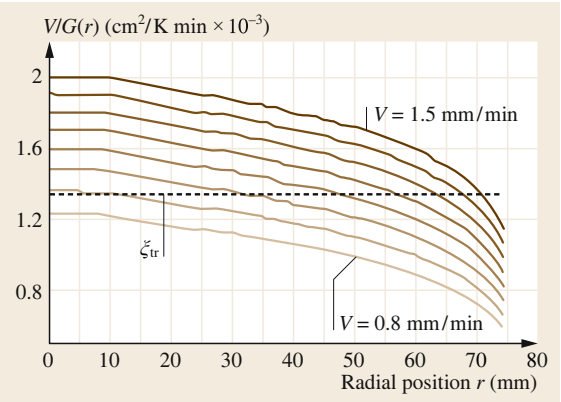


Fig. 5.10 Radial variation in the crossover between the transition value ξ_{tr} and $V/G(r)$ as function of decreasing pull rate V

It should be noted that the above behavior is probably unique to silicon as it stems from the peculiar phenomenon that the vacancy and Si interstitial equilibrium concentrations in crystalline silicon are very similar, but C_v^{eq} is slightly higher than C_I^{eq}, although the diffusivity of Si interstitials exceeds that of vacancies at the melting point. In addition, the quantitative relationship between the relevant parameters which allows the changeover in defect behavior to occur in accessible growth rate regimes is quite astonishing. The time-dependent behaviors of the point defect concentrations are determined by the species conservation equations [5.78]

$$\frac{DC_I}{Dt} = \nabla\left[D_I(T)\nabla C_I - \frac{D_I(T)C_I Q_I^*}{k_b T^2}\nabla T\right] - k_{IV}\left(C_I C_V - C_I^{eq} C_V^{eq}\right) - \frac{\partial}{\partial t}\int_2^\infty n f_I(n,r)\,dn\,, \tag{5.2}$$

where Q_I^* is the reduced heat of transport for Si interstitials, which describes the rate of material flow due to a temperature gradient. A similar expression applies for vacancies (just exchange the I and V notations). The substantial derivative $D/Dt = [\partial/\partial t + V(\partial/\partial z)]$ is defined here as the rate of change of species concentration at a point moving with the velocity (V) of the advancing growth interface. This term relates to the incorporation of species due to convection. The first term on the left-hand side accounts for Fickian diffusion and the second for thermodiffusion. The last term represents the loss of point defects into their aggregates of all possible sizes. This term can be neglected as long as the supersaturation of the surviving species has not increased to the point where cluster nucleation starts. A detailed analysis also reveals that thermodiffusion has no detectable impact. Hence, the simple picture drawn above – which only takes into account convection and Fickian diffusion – is quite reasonable. Based on these simplifications, the incorporated vacancy concentration can be calculated from a simple expression [5.76, 79]

$$C_V(r) \approx \left[C_V^{eq}(T_m) - C_I^{eq}(T_m)\right]\left(1 - \frac{\xi_{tr} G(r)}{V}\right). \tag{5.3}$$

This expression is also applicable to Si interstitials (at $V/G < \xi_{tr}$). It then defines the incorporated Si interstitial concentration C_I with a minus sign.

If V/G approaches ξ_{tr}, then the vacancy and Si interstitial fluxes become equal and both species are annihilated by V–I recombination. Thus, the incorporated V and I concentrations are negligible and no aggregates can form. This particular condition must be met for the growth of so-called perfect and *defect free* silicon, respectively.

Based on the picture outlined above, one would expect to be able to grow crystals which are I-rich in the center and V-rich near the surface by simply engineering the hot-zone and the process conditions in such a way that G is higher in the center than at the rim of the growing crystal (Fig. 5.11). Surprisingly, respective experiments which were carried out some years ago showed that, under such conditions, vacancy aggregates can still prevail in the center and interstitial clusters at the periphery of the crystal [5.80]. The authors explained these findings by the effect of thermo-mechanical stress during growth on the solubility of native point defects by noting that compressive stress increases the solubility of V_s and vice-versa for I_s. As a consequence, the critical value of V/G, where vacancies and interstitials annihilate each other, be-

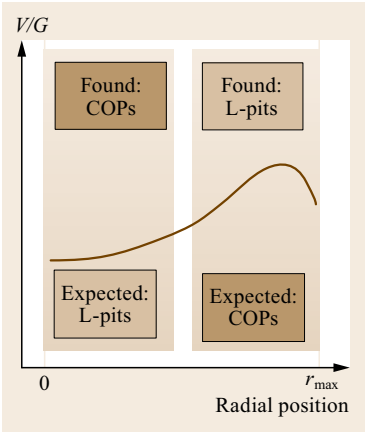


Fig. 5.11 Radial variation of V/G for a special hot zone with compressive stress in the crystal center and tensile stress at the periphery of the growing crystal. Comparison of expected and experimentally found defect behavior

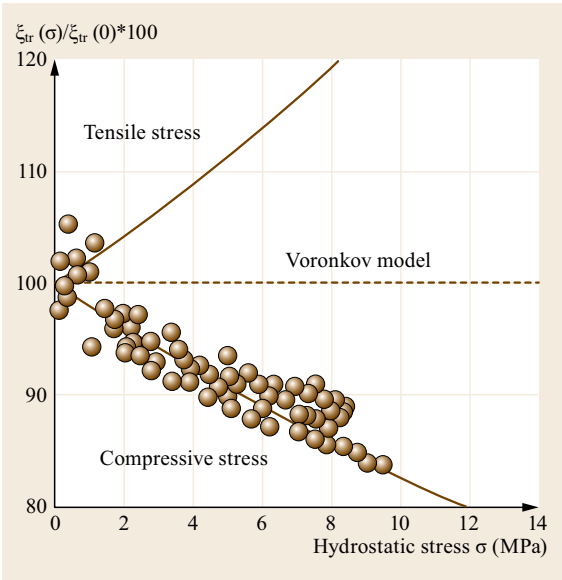


Fig. 5.12 Dependence of $\xi_{tr}(\sigma)/\xi_{tr}(0)$ on tensile and compressive stress in the growing crystal. *Open circles* show the data of *Nakamura* after scaling the hydrostatic stress to account for the differences in elastic parameters used. (After [5.44, Fig. 1], [5.80])

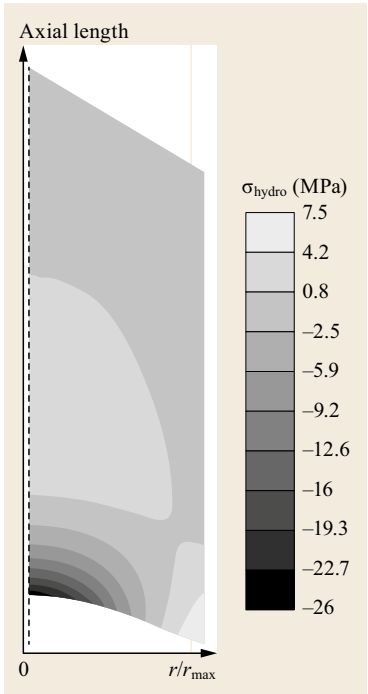


Fig. 5.13 Distribution of stress in a growing crystal which has a similar radial variation of V/G as in Fig. 5.11. However, the stress in the growing crystal is engineered in such a way that *defect free* silicon is obtained

comes a function of the thermo-mechanical stress in the vicinity of the growth interface (5.1). Figure 5.12 shows how $\xi_{tr}(\sigma)/\xi_{tr}(0)$ depends on hydrostatic stress according to [5.80].

The dependence is obtained from the classical expression for $\xi_{tr}(0)$ (5.1) by adding the effect of hydrostatic stress to the formation enthalpy of point defects i. e., by adjusting the volume change due to point defect creation using the same value for both I's and V's for simplicity and neglecting the effect of stress on point defect diffusivity. It should be noted that a more accurate solution is obtained if the diffusion-convection-reaction problem is solved in the full crystal which is especially important for strongly curved interfaces.

Figure 5.13 depicts the distribution of hydrostatic stress and crystal interface shape for a 300 mm crystal where the radial variation of V/G is similar to that of Fig. 5.11, but the thermo-mechanical stress is engineered to balance V/G so that *defect free* silicon can be obtained [5.80]. There is quite large compressive stress in the crystal center and several times smaller tensile stress near the triple point.

The effect of stress on the formation energy of point defects was already discussed by *Tanahashi* et al. [5.81], before the above experiment was carried out, but their estimates were far too low to explain the findings of [5.80]. Recently, *Nakamura* et al. [5.44] published a new set of data showing the dependence of $\xi_{tr}(\sigma)/\xi_{tr}(0)$ on hydrostatic stress. They are in good agreement with Fig. 5.12, if one corrects the effect of the different elastic parameters used in both publications. For a more detailed analysis including extensive atomistic simulations, we refer to the recent literature [5.45, 82, 83].

5.3.3 Theoretical Model: Aggregation of Intrinsic Point Defects

Due to the technological dominance of vacancy-rich silicon, previous theoretical investigations of defect aggregation have mainly focused on the formation of voids. In addition, it was also easier to verify the theoretical results for vacancy aggregation because voids are not obscured by the secondary defect generation that occurs for L-pits.

From Fig. 5.14, it can be seen that, for a standard growth process, the V-supersaturation already begins to build up at around 1200 °C [5.84], which is well above the experimentally determined nucleation temperature of voids (≈ 1100 °C). Thus, the supersaturation C_V/C_V^{eq} necessary to nucleate voids is appreciable, ≈10 [5.78], as estimated from Fig. 5.14. This fact also allows us to decouple the phase where V–I recombination is the major vacancy loss mechanism from the phase where nucleation dominates the vacancy loss, so that one can assume that the cluster formation starts with a fixed concentration which can be cal-

culated from (5.3). For special growth processes that are designed to minimize or even eliminate voids, the starting vacancy concentration after V–I recombination is so small that the beginning of supersaturation (and hence void nucleation) occurs at notably lower temperatures [5.79].

The first vacancy aggregation calculations were performed with a model that was originally developed to simulate oxygen precipitation [5.64, 85]. For the formation of small clusters (less than 20 vacancies), the kinetics are described by chemical rate equations; for larger clusters, the rate equations are expanded into a continuum formulation that yields a single Fokker–Planck equation. The void shape is assumed to be spherical. Input data are the computed temperature field, which can be quite accurately calculated using commercially available standard codes, and the surviving vacancy concentration after V–I recombination has ceased. Besides the physical properties of the vacancies and Si interstitials, the surface energy must also be known, which is estimated to be around 950 erg/cm^2.

As can be seen from Fig. 5.14, the computer simulations yield a nucleation temperature of around 1100 °C, which agrees fairly well with the experimental results. The prediction of the correct aggregation temperature has been a matter of discussion. With the above input data, one ends up with a far lower aggregation temperature than those observed [5.86, 87]. The main reason for this is that the nucleation process does not start as a small octahedral void as previously assumed, but rather as an *amorphous cloud* of vacancies which, after some time, relaxes into an octahedron with (111) facets [5.88]. Proper modeling of this initial nucleation process with atomistic calculations does indeed shift the calculated aggregation temperature into the experimentally observed range.

The results in Fig. 5.14 also demonstrate that a higher cooling rate obviously decreases the nucleation temperature. The origin of this effect is the increased supersaturation, which builds up because there is less time for vacancy consumption. Due to the shift in the nucleation temperature, the residual vacancy concentration after void formation remains higher for fast cooling rates than for slow cooling. This has an important effect on the nucleation of oxygen precipitates, as the latter is strongly enhanced by a higher free vacancy concentration. Figure 5.15 illustrates the evolution of the void density/size distribution during the nucleation period. It is apparent that the maximum of the distribution is shifted to larger void sizes not only due to the growth of the previously nucleated voids, but also due to the fact that the newly nucleated clusters have a larger size. The latter effect results from the decreased supersaturation with higher void density, which, in turn, increases

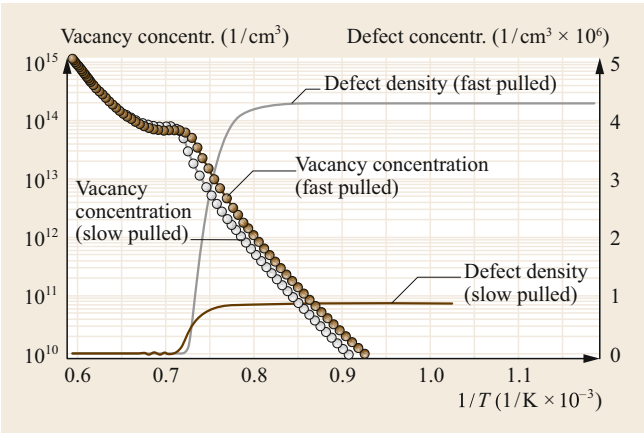


Fig. 5.14 Simulated variation in the vacancy concentration in growing crystals as a function of temperature for rapidly and slowly cooling growth processes. The density of voids nucleated at around 1100 °C and below is shown on the *right scale*

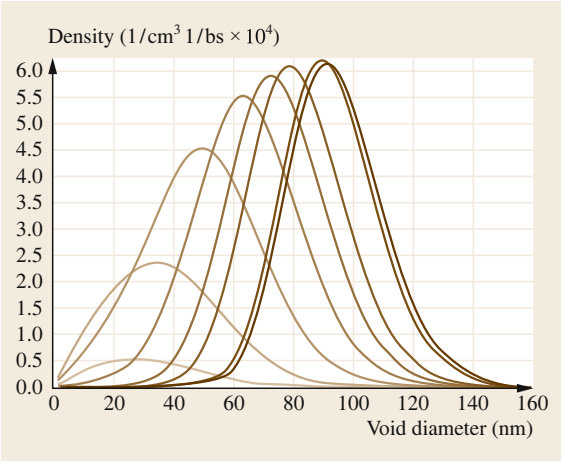


Fig. 5.15 Simulated evolution of the void density–size distribution at various temperatures of the growing crystal. The temperature difference between each curve is 4 K

the critical radius for stable nuclei. Thus, only increasingly large clusters can nucleate towards the end of the nucleation period.

As depicted in Fig. 5.16, a fast cooling rate decreases the cluster size but increases their density and vice versa, which correlates well with the experimental data. The reason for this is the above-mentioned higher supersaturation at fast cooling rates, which entails a larger nucleation rate. The higher cluster density then effectively lowers the residual vacancy concentration. The clusters, however, have no time to grow and so remain small.

A similar effect is obtained when the initial vacancy concentration before the onset of nucleation is

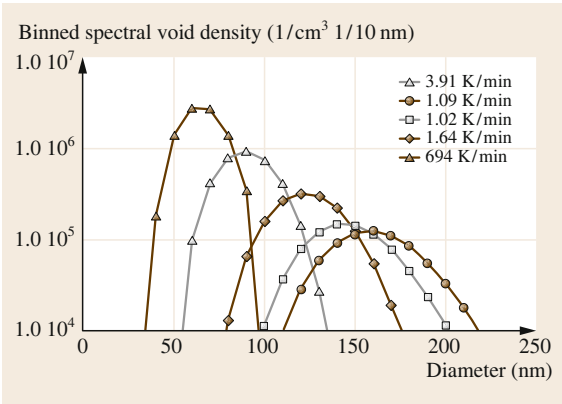


Fig. 5.16 Simulated density/size distributions of voids for growth processes with different cooling rates

reduced [5.89]. A lower vacancy concentration causes a downward shift in the nucleation temperature, where the vacancy diffusivity is lower. As a consequence, they need more time to diffuse to sinks, which in turn drives up supersaturation and so more clusters of a smaller size are formed again. The behavior of the size distribution as a function of the initial vacancy concentration gives a simple explanation for the experimentally found radial variation in the void size distribution in Fig. 5.7. As the initial vacancy concentration decreases towards the crystal rim (Fig. 5.17), the average size of a void shrinks but void density grows. Thus, the density of small voids is low in the center, but increases towards the crystal rim until the vacancy concentration drops below a critical value, where the average void size is below the detection limit. On the other hand, the number of large voids shrinks with decreasing vacancy concentration. Therefore, the density of large COPs is highest in the center. This behavior is also fairly well reproduced by simulation (Fig. 5.18).

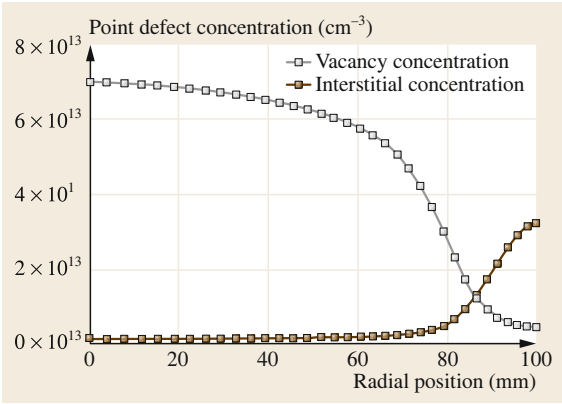


Fig. 5.17 Simulated radial variations in the vacancy and Si interstitial concentrations, respectively. (After [5.70])

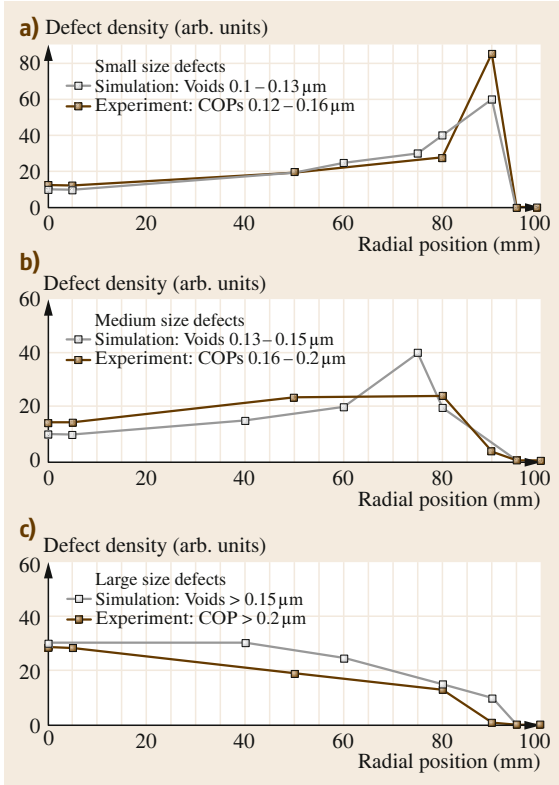


Fig. 5.18a–c Simulated radial variation in the density/size distribution of voids and comparison with experimental data. (After [5.70])

A simple expression that predicts a void density proportional to $q^{3/2}C_V^{-1/2}$ ($q =$ cooling rate, $C_V =$ initial vacancy concentration before nucleation starts) has been derived by *Voronkov* et al. [5.68]. This formula is obviously in excellent agreement with the experimental results (Fig. 5.19). A deviation from the above expression was reported for detached crystals, where significantly higher cooling rates can be achieved (2–70 K/min) [5.90]. In the latter case, this discrepancy was tentatively explained by a suppression of voids in favor of enhanced formation of oxygen precipitates at high cooling rates (> 40 K/min) [5.91]; in other words, so-called *oxide particles* nucleate at around 1100 °C rather than voids.

The $C^{-1/2}$ dependence is more difficult to verify experimentally because a change in the initial vacancy concentration entails a change in the nucleation temperature and so temperature-dependent parameters like diffusivities and equilibrium concentrations must be adjusted. Most of the current aggregation models do not consider the oxidation of the inner void surface, which stops void growth before the vacancies are essentially depleted. This has less impact on the void size

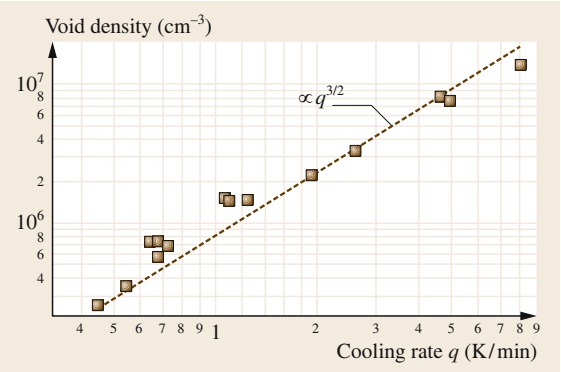


Fig. 5.19 Void density as derived from GOI measurements as a function of cooling rate of the growing crystal. The *straight line* relates to predictions of the $q^{3/2}$ law

distribution but considerably affects the residual vacancy concentration and therefore oxygen precipitation (Sect. 5.4).

The occurrence of double voids is still under discussion. One attempt to explain this phenomenon is based on a partial oxidation of the (111) facets during void growth, which allows the incorporation of further vacancies at the unoxidized residual facet surface [5.92].

5.3.4 Effect of Impurities on Intrinsic Point Defect Aggregation

The influence of impurities on the aggregation of intrinsic point defects has been widely investigated. It was found that many of them can notably modify the defect density and size which is of great technological importance. Many attempts have been made to explain the specific mechanisms. However, a unified theoretical approach has not been achieved. This is also due to the fact that the observed effects occur at concentrations which vary over many orders of magnitude.

Nitrogen

Codoping of silicon crystals with nitrogen is of considerable relevance in technical applications, as nitrogen already has a massive impact on point defect aggregation at very low concentrations. It has long been known to simultaneously suppress interstitial- and vacancy-related defects – A-swirls and D-defects/voids, respectively [5.93] – in floating zone (FZ)-grown crystals. The nitrogen concentrations required are very low ($< 2\times10^{14}$ at/cm^3) and were found to be in the same range as the surviving residual intrinsic point defect concentrations. Owing to the low nitrogen concentration, adverse effects of nitrogen codoping are technically insignificant. Well known is the formation of nitrogen related thermal donors after neutron transmutation doping which, however, can be handled by properly annealing the neutron irradiated ingots [5.94]. Consequently, all FZ grown silicon crystals are nitrogen codoped, today.

In the case of Czochralski (CZ)-grown silicon, nitrogen has been an undesirable dopant in the past, because it enhances the formation of oxidation-induced stacking faults (OSF) [5.95]. As the nitrogen concentration increases, the OSF ring widens [5.96] until it extends over the entire crystal diameter. In addition, due to the comparatively high oxygen concentration (3–8×10^{17} at/cm^3), the desirable defect suppressing effects of nitrogen are largely offset by the interaction of nitrogen and oxygen [5.97]. However, nitrogen still has a notable effect on the defect size: the higher the nitrogen concentration, the lower the defect size [5.18, 98]. This effect is important, because a small defect size is highly favorable for defect annealing [5.16]. It was further found that nitrogen not only reduces the defect size but also systematically changes the void morphology from octahedron to platelet and, finally, to rods or even rod clusters with increasing concentration (Fig. 5.20) [5.99]. Nitrogen doping has gained additional interest as it also strongly enhances oxygen precipitation [5.100–102] which can be used to improve the efficiency of gettering of metallic contaminants in low thermal budget device processes [5.103, 104].

Unfortunately, quantitative data on the chemical and physical properties of nitrogen in silicon are rare. Experimentally, it was found that nitrogen exists as N–N dimers in as-grown silicon [5.105], which has a diffusivity three orders of magnitude higher than oxygen at 1100 °C [5.106]. These diffusion experiments indicate that the N–N dimer is stable up to at least 1270 °C [5.107], which is in line with first-principles

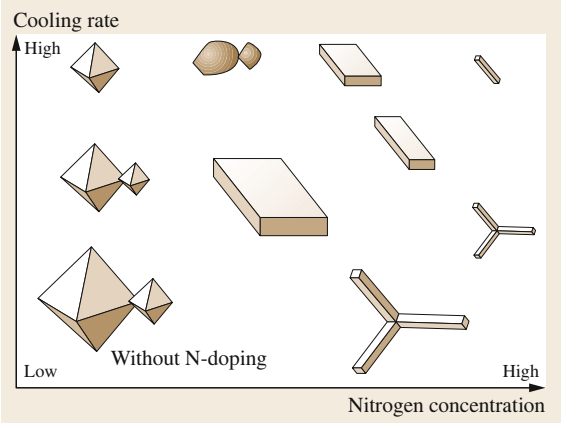


Fig. 5.20 Variation in void morphology as a function of nitrogen concentration and cooling rate of the growing crystal. (After [5.99])

calculations of the N–N dimer configuration [5.108–110]. In contrast with the nitrogen molecule, the N–N dimer in silicon does not have any direct bonds between the two neighboring nitrogen atoms. Experiments with nitrogen-implanted silicon samples heated to melt temperature by a laser pulse and subsequently quenched to room temperature proved that roughly 10% of the nitrogen can be frozen on substitutional sites as single nitrogen atoms [5.111, 112]. The fact that substitutional nitrogen is generally not observed in regularly cooled samples suggests that substitutional nitrogen is not stable in silicon at lower temperatures [5.113].

As the nitrogen concentration needed to suppress A-swirl and D-defects is in the same range as the intrinsic point defect concentration before aggregation, it was suggested that nitrogen directly interacts with vacancies (V) and Si interstitials (I) and thus prevents their aggregation [5.114, 115]. One model proposes that the storage of vacancies in impurity complexes would gradually shift the transition value ξ_{tr} of V/G to higher values which, in turn, would explain the shrinkage of the void region with increasing nitrogen content [5.116]. On the other hand, the model also seems to imply that the A-swirl region simultaneously expands toward the crystal center, which is in conflict with experimental results. In particular, the Li-drift experiments of *Knowlton* et al. [5.117] proved that, despite the suppression of detectable COPs, the extent of the vacancy-rich region does not change due to nitrogen doping. Other attempts are based on the reduction of vacancy diffusivity due to the vacancy–nitrogen interaction [5.118] or on increasing the reaction barrier to vacancy absorption at the void interface [5.119]. However, the fact that nitrogen doping has no impact on the inner oxide layer growth of the voids [5.99] does not support the notion that nitrogen modifies the properties of the void interface. None of these proposals gives any explanation for influence of nitrogen on Si interstitial aggregation.

First-principles calculations by *Sawada* et al. [5.108] and *Kageshima* et al. [5.109] have shown that the N–N dimer can form stable N_2V and N_2V_2 complexes, the latter being more stable due to a larger enthalpy of formation. Based on these results, it has been proposed that the vacancy aggregation is suppressed by the reaction

$$\text{I)} \quad N_2 + V \leftrightarrow N_2V \,,$$

whereas the formation of Si interstitial agglomerates is prevented by the reaction

$$\text{II)} \quad N_2V + I \leftrightarrow N_2 \,.$$

Nitrogen is not likely to be incorporated as a N–N dimer into the crystal. This is concluded from the experimental fact that nitrogen does not evaporate from the silicon melt and so it must be strongly bound to silicon atoms, which is not possible for a nitrogen molecule. It was therefore proposed that single nitrogen atoms are incorporated on interstitial (N_I) as well as substitutional sites (N_s) at the growth interface [5.120]. Behind the growth interface, the highly mobile interstitial nitrogen atoms then quickly form N–N dimers via $N_I + N_I \leftrightarrow N_2$ and nitrogen-vacancy complexes via $N_I + N_s \leftrightarrow N_2V$ inside the crystal. A ratio of $[N_I] : [N_S] = 7 : 1$, which corresponds to 12% of the nitrogen at substitutional sites, has been determined from FZ crystals grown with various nitrogen contents. This is in good agreement with the above-mentioned experimental findings for substitutional nitrogen. As vacancies and Si interstitials have very similar concentrations at the melting point, reactions (I) and (II) mainly act as an additional V–I recombination path during crystal cooling, with only marginal impact on the levels of N_2 and N_2V. However, when recombination is complete and the vacancies or the Si interstitials have been wiped out, reaction (I) or (II) prevents the aggregation of the surviving species. Using this model, it has been possible to determine the radial variations of vacancies and Si interstitials in FZ crystals before the onset of defect aggregation [5.120]. This was achieved by measuring the radial extension of the COP and A-swirl region as a function of the nitrogen concentration. The result is in excellent agreement with theoretical predictions (Fig. 5.21). Surprisingly, the experimental data could not be fitted by assuming that the formation of N_2V_2 complexes prevails, as predicted by theoretical calculations. It is not clear whether the reaction to N_2V_2 is hampered by a high-energy barrier.

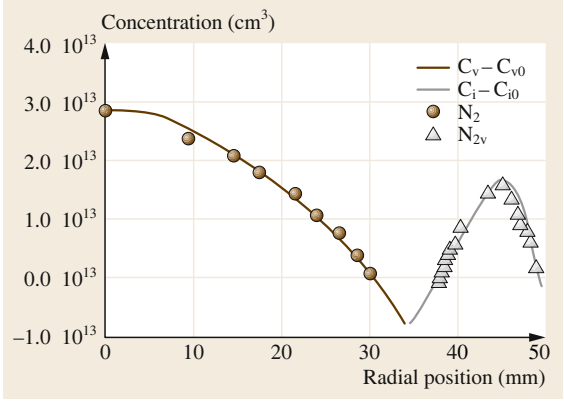


Fig. 5.21 Radial vacancy and Si interstitial distributions as derived from a FZ crystal grown with an axially varying nitrogen concentration. $C_{V0} = 8.8 \times 10^{12}$ at/cm^3 and $C_{I0} = 7.7 \times 10^{12}$ at/cm^3. The *solid* and *dashed lines* refer to simulation results. (After [5.120])

None of the above mentioned models can explain why the effect of nitrogen on defect aggregation is strongly diminished in CZ crystals. It is clear that the well-known interaction between nitrogen and oxygen, which results in the formation of N_xO_y complexes [5.121], is probably the origin of the vanishing nitrogen effect. However, it is also known that, despite oxygen levels of 5-6 $\times 10^{17}$ at/cm^3, ample unreacted nitrogen is left in oxygen- and nitrogen-doped as-grown crystals to suppress defect aggregation [5.122]. It is therefore necessary to look for a mechanism that significantly reduces the concentration of free nitrogen dimers at the nucleation temperature of voids and L-pits, but reestablishes their concentrations at lower temperatures. As the experimentally confirmed N_2O complex is only stable up to about 700 °C [5.121, 123], it has been proposed [5.120] that nitrogen reacts with oxygen at high temperatures according to

$$\text{III)} \quad 2\text{NO} \leftrightarrow \text{N}_2 + 2\text{O}_\text{i} \; .$$

The existence of the NO complex has been theoretically predicted by *Gali* et al. [5.124] and, more recently, has also been experimentally confirmed [5.125, 126]. It has a similar structure to the N_2 complex but with a smaller binding energy. Due to the large excess of oxygen over nitrogen, it has been assumed that the equilibrium of reaction (III) is on the left-hand side near the melting point and so very little N_2 is available. As the temperature decreases, the equilibrium gradually shifts to the right-hand side and more and more N_2 is generated. If the temperature at which N_2 starts to develop in sufficient quantities is below the vacancy aggregation temperature, the formation of voids is not suppressed. If the N_2 formation temperature is above the vacancy aggregation temperature, void formation depends on how much N_2 is available in relation to the vacancy concentration at the void nucleation temperature. The higher the N_2 formation temperature, the more N_2 is available to reduce the number of vacancies by reaction (I). According to reaction (III), the formation temperature of N_2 should increase with lower oxygen and higher nitrogen concentrations. Lower free vacancy concentrations and therefore lower aggregation temperatures result in voids of smaller size but higher density, in agreement with the experimental findings.

After void nucleation, the free vacancy concentration drops rapidly and the above reaction (I) reverses, giving $N_2V \rightarrow N_2 + V$, which slows down the decay of free vacancies. Thus, the N_2V-complex acts as an interim reservoir for vacancies – its formation before void formation removes vacancies and its dissociation after void formation releases them again [5.96]. In conjunction with the lower vacancy aggregation temperature, this results in a notable increase in free vacancy concentration during this cool-down phase compared to nitrogen-free material (Fig. 5.22). As free vacancies are known to strongly favor the nucleation of oxygen precipitates, the latter effect gives a simple explanation for the observed strong enhancement of oxygen precipitation in nitrogen-doped CZ crystals (Sect. 5.4).

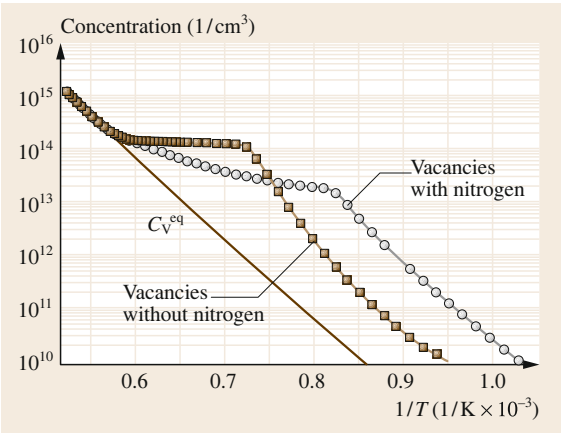


Fig. 5.22 Simulated variations in the vacancy concentration as a function of temperature for a nitrogen-doped and an undoped crystal

The remarkable variation in the void morphology as a function of the nitrogen content is still under discussion. A similar change in morphology is also known for oxygen precipitates. In this case, the effect can be interpreted in terms of the balance between the relaxation energy of the lattice strain and the interfacial energy. However, TEM investigations do not reveal any lattice strain around voids, and so another mechanism is required to explain the varying void morphology.

Boron

Boron doping affects the intrinsic point defect aggregation in a similar way to nitrogen doping, but much higher concentrations are required [5.127]. These high concentration are technologically relevant as highly boron doped wafers are widely used as substrates for epitaxial layers in the electronic industry. It was found that the COP region as well as the diameter of the OSF ring starts to shrink when the boron content exceeds roughly 5 $\times 10^{18}$ at/cm^3 (Fig. 5.23). In contrast with nitrogen, OSF formation is not enhanced and the width of the OSF ring remains unchanged with increasing boron concentration. At around 10^{19} at/cm^3, the COPs and the OSF ring disappear in the center of the crystal. More recently, it was reported that L-pits are also suppressed in the same concentration range [5.128] which is in line with the well known defect-free quality of pp + epi wafers [5.62] (Sect. 5.3.1).

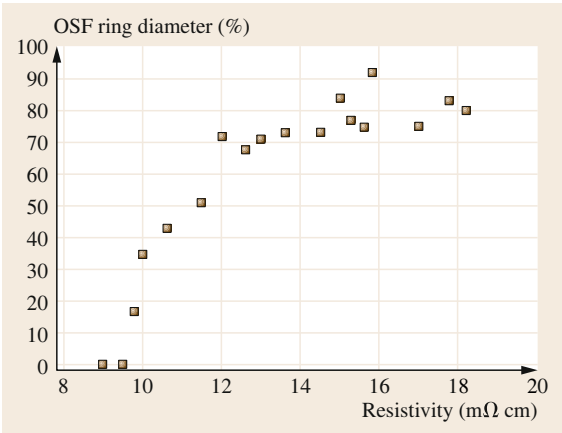


Fig. 5.23 Variation in the OSF ring diameter as a function of the boron concentration. (After [5.129])

Dornberger et al. [5.69] suspected that changes in the equilibrium concentrations and diffusivities of intrinsic point defects in the presence of high boron concentrations would modify the transition value ξ_{tr} of the parameter V/G as a function of the boron content. *Voronkov* and *Falster* [5.87] also proposed a modification of equilibrium concentrations as a result of the shift in the Fermi level due to the high boron concentration. Based on this approach, the latter authors calculated a shift in the transition value ξ_{tr} which would account for the observed shrinkage of the void region. However, *Vanhellemont* et al. recently carried out ab initio calculations which demonstrated that vacancies at the melting temperature of silicon are neutral independent of the position of the Fermi level [5.130]. As with nitrogen doping, the ξ_{tr} shift entails the simultaneous appearance of L-pits in the outer crystal region and a shrinking void region, which is in conflict with experimental results of boron doped crystals, as well. Another attempt by *Sinno* et al. [5.131] considers reversible reactions between boron and intrinsic point defects, in particular self-interstitials. The formation of BI and B_2I complexes consumes self-interstitials which would otherwise be annihilated by recombination. As recombination consumes the remaining self-interstitials, BI and B_2I start to dissociate again according to le Chatelier's principle. As a result, the self-interstitial concentration increases, which shifts the point defect balance towards the intersitial-rich side, causing the void region to shrink. No comment is made about whether or not the model is also able to account for the suppression of L-pit formation by storing supersaturated self-interstitials in boron complexes. Recently, *Sueoka* et al. presented a model which considers the variation of formation energies of intrinsic point defect around dopant atoms and, thus, the change of equilibrium concentrations of vacancies and Si interstitials in dependence on the dopant concentration [5.46]. In case of boron, their results correctly predicted the behaviour of the void region as a function of the boron concentration, but the simultaneous suppression of L-pits was not reproduced. They suspected that, in contrast with vacancies, the Fermi level effect cannot be neglected for self interstitials which would turn them into positively charged species [5.132]. The suppression of interstitial clusters could then be explained by the Coulomb repulsive force. However, the debate on whether or not Si interstitials are charged or neutral is not yet settled.

Carbon

Carbon doping is also known to impact intrinsic point defect aggregation. It has been reported that the void region in the crystal center disappears upon carbon doping [5.59, 133], while the region of Si interstitial aggregates is widened; in other words, carbon does not inhibit the formation of L-pits/A-swirl. This behavior appears to be in line with the model put forward in [5.116], which predicts an upward shift of ξ_{tr} for carbon doping and, in turn, an enlargement of the interstitial-rich region.

It was also found that carbon doping reduces the grown-in defect size [5.99, 134]. Although the size reduction is appreciable, the morphology of the voids is not changed, in contrast to the consequences of nitrogen doping. Only a tendency towards multiple void formation was observed. The thickness of the inner oxide layer was found to be similar to that of undoped crystals, which indicates that carbon doping has no influence on the growth of the inner oxide layer. In more recent studies, it was reported that the morphology and size of Si interstitial aggregates changes with increasing carbon concentration [5.128, 135]. The L-pit, which is a local network of dislocation loops, turns into a globular Si interstitial aggregate, probably a stacking fault of Si interstitials. In addition, the density of these defects is notably increased with the carbon content. As with boron, the effect on defect aggregation is only seen at significantly higher concentrations ($\approx 1\times10^{17}\,\mathrm{at/cm^3}$) than for nitrogen. The above model of *Sueoka* et al. was also applied to carbon and showed qualitative agreement with the experimental findings [5.46].

At higher concentrations, carbon is known to enhance oxygen precipitation too [5.136, 137]. As carbon predominantly resides on substitutional sites, it is very unlikely that, as in the case of nitrogen doping, a higher residual vacancy concentration is responsible for the stronger oxygen precipitation. On the other hand, the small carbon atom exerts a local tensile strain on the surrounding lattice and attracts Si interstitials to form C_xI_y complexes. Thus, Si interstitials ejected into the

lattice by growing oxygen precipitates are effectively removed and, in turn, further precipitate growth is not retarded by a build-up of Si interstitial supersaturation. The enhanced oxygen precipitation may also be related to heterogeneous nucleation at small carbon aggregates [5.138].

Antimony, Arsenic and Phosphorus

In recent years, the effect of n-type dopants in high concentrations like antimony, arsenic and phosphorus have been been investigated in detail [5.128, 139–142]. Similar to boron, highly n-doped wafers are used as substrates for epitaxial layers which are mainly needed for high power devices. In contrast with highly boron doped crystals, all of these n-dopants expand the void region with increasing concentration. e.g., crystals, which are grown under conditions that make them Si interstitial rich (L-pits), are transformed into vacancy rich with strong void formation upon a Sb doping level of $3\times10^{17}\,\mathrm{at/cm^3}$ [5.139]. This is also in line with older measurements on Sb doped floating zone grown crystals where the defect behaviour was shifted from Si interstitial rich (A-swirl) to vacancy rich (D-defects) by Sb doping [5.143]. The void density increases with higher dopant concentrations. However, beyond a certain critical concentration, the void density sharply drops and no defects are observed anymore. The critical level obviously depends on the dopant and is highest for phosphorus and lowest for antimony (Fig. 5.24). The void density is rather uniform across the crystal diameter and also disappears uniformly beyond the critical dopant concentration. The morphology of the voids changes with the n-dopant [5.142]. For Sb-doping, single voids were observed, only, while, in case of P-doping, one quarter of the voids was detected as double voids. More than 50% double voids were found in As-doped crystals. Unlike nitrogen, carbon and boron, oxygen precipitation is not notably affected by the n-dopants, but is rather similar to the behaviour of lightly doped wafers.

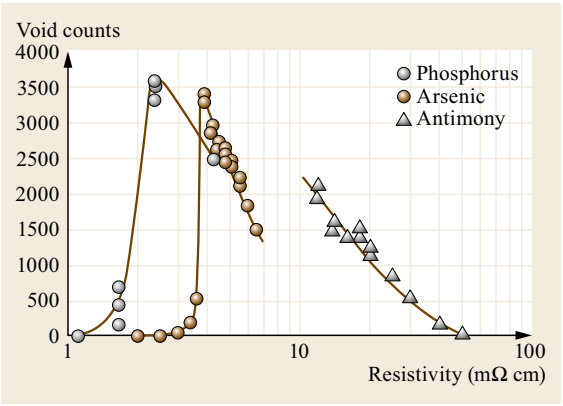


Fig. 5.24 Void counts in dependence of the resistivity for P-doped (*grey circles*), As-doped (*brown circles*) and Sb-doped crystals (*grey triangles*). (After [5.142])

All proposed models explain the observed effects by assuming a shift of ξ_{tr} to lower values. While *Voronkov* et al. [5.144, 145] attribute the shift of ξ_{tr} to vacancy and Si interstitial trapping by the dopants, respectiviely, and can also reproduce the sharp drop of the void density beyond the critical dopant concentration [5.141], *Sueoka* et al. again apply their above mentioned model predicting a change of the formation energy of vacancies and self interstitials in the presence of n-dopants which results in a similar shift of ξ_{tr} [5.46].

5.4 Formation of OSF Ring

The ring-like distributed OSFs are oxygen precipitates with platelet shapes that grow particularly large at the edge of the void region, and exceed a critical size necessary to create stacking faults during subsequent wafer oxidation there [5.146]. The critical radius of the grown-in platelets is $\approx 70\,\mathrm{nm}$. The formation of OSFs can be suppressed if the cooling rate of the growing crystal is increased (this means, for example, that OSFs are not found in oxygen-doped FZ crystals [5.129] which have very high cooling rates) or if the oxygen content is decreased.

The peculiar ring-like distribution is a consequence of the well-known strong enhancement of oxygen precipitation by vacancies [5.38, 147]. The reason for this is that the absorption of vacancies allows the oxygen precipitate, which occupies twice as much volume as the corresponding silicon lattice, to nucleate and grow without building up notable strain energy. As will be described below, the radially inhomogeneous oxygen precipitation is a consequence of a substantial radial inhomgeneity of the free vacancy concentration in contrast to the rather flat radial profile of oxygen.

The vacancy concentration, as pointed out previously, has its maximum at the center of the growing crystal after V–I recombination has ceased (Fig. 5.25). Therefore, the critical supersaturation for void formation is first reached at the crystal center at relatively high temperatures. Hence, the free vacancies are quickly consumed in this area. As the crystal cools, voids are also nucleated in the regions of lower initial vacancy

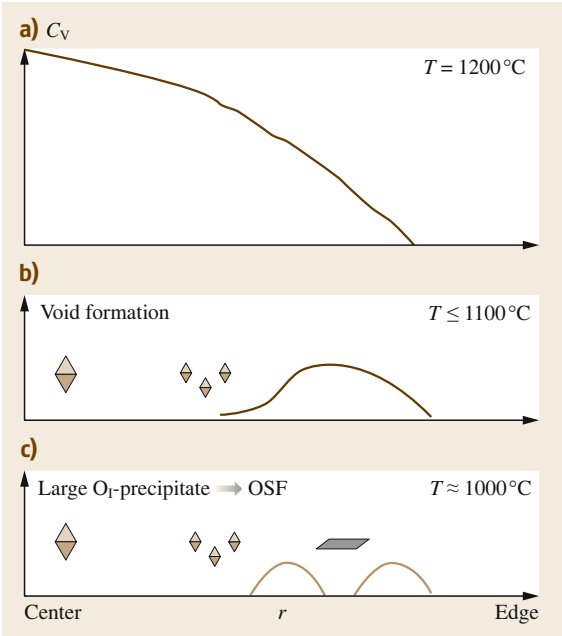


Fig. 5.25a–c Schematic evolution of the radial vacancy profile and accompanying void and large O_I-precipitate formation at decreasing temperatures

concentration, meaning that vacancy consumption then also occurs further away from the crystal center. As the removal of vacancies is enhanced at higher temperatures, where the diffusivity is large, the radial vacancy distribution finally drops to a minimum in the crystal center, and the maximum of the residual vacancies gradually moves toward the V/I boundary upon further cooling. Because of this shifted vacancy maximum, the first oxygen aggregates nucleate close to the V/I boundary on the vacancy-rich side. The relatively high nucleation temperature at this position results in larger sizes but lower densities for the corresponding grown-in oxygen precipitates compared to those nucleated at lower temperatures away from the V/I boundary. Consequently, they preferentially reach the critical size for OSF formation during subsequent wafer oxidation. At sufficiently fast crystal cooling rates, the oxygen precipitates are nucleated at higher densities but smaller sizes and cannot grow to the critical size. Thus, OSF formation is prevented, in agreement with experimental findings. The same density/size effect is obtained if the oxygen content is lowered – similar to the behavior of voids – and so OSFs are also suppressed.

The nucleation of the first oxygen precipitates quickly results in the depletion of residual vacancies in their neighborhood. Thus, the previous vacancy maximum at the V–I boundary is then converted into a pronounced minimum which suppresses further nucleation in this area. As a result, two vacancy peaks with a region of large oxygen precipitates in-between is frozen in as the crystal cools down to room temperature (Fig. 5.25). Later on, when the wafers are subjected to further heat treatments, such as a nucleation step at 780 °C followed by a growth step at 1000 °C, new oxygen precipitates with high density are preferentially nucleated at the radial positions of the two remaining vacancy maxima. This results in a distinct profile for the interstitial oxygen concentration after these heat treatments. In the radial band where the large precipitates and the OSF ring are located, the interstitial oxygen concentration is only slightly reduced because further nucleation of precipitates during wafer heat treatment is suppressed there and the comparatively low density of precipitates does not consume much of the interstitial oxygen. On the other hand, the high density bands at the locations

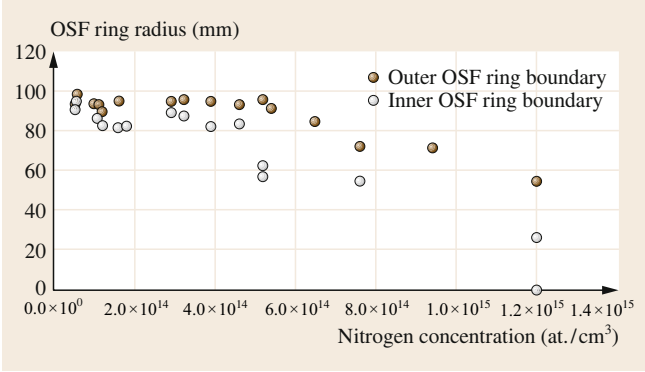


Fig. 5.26 Variations of the outer and inner OSF ring boundaries as a function of the nitrogen concentration. (After [5.96])

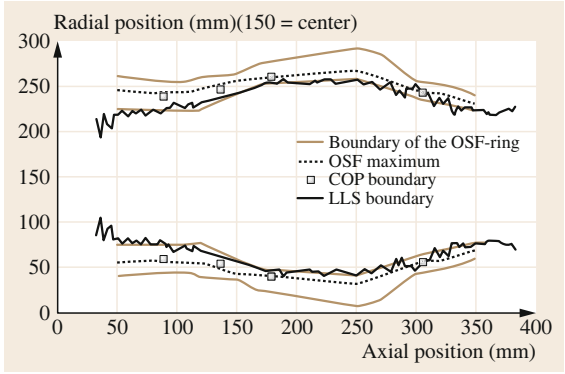


Fig. 5.27 Variations in the OSF ring width and its radial location in comparison to the variations of the COP and LLS boundaries along the crystal axis. The LLS defects were detected on the as-polished wafer surface, which detects only the larger voids, while the COP boundary was determined by LLS measurements after repeated SC1 cleaning, which enlarges the voids to above the LLS detection limit

adjacent to the former vacancy peaks coincide with pronounced minima in the interstitial oxygen content.

Considerable changes in the widths and the radial positions of OSF rings are observed if the crystals are nitrogen-doped. With increasing nitrogen content, the outer and inner OSF ring boundaries are shifted towards the crystal center [5.96]. However, the shift is much larger for the inner OSF ring boundary, resulting in a significant widening of the OSF ring width (Fig. 5.26) until it extends over the entire crystal volume. This behavior was tentatively explained by a schematic model that takes into account the slower build-up of vacancy supersaturation due to the formation of N_2V complexes and the resulting lower void nucleation temperature. The model also predicts that the void and the OSF region overlap, roughly in the area between the inner boundary and the middle of the OSF ring, which is in agreement with experimental data (Fig. 5.27).

In contrast to nitrogen, carbon tends to suppress OSF formation. This is again an indication that the mechanism of the effect of carbon is different. The above-mentioned idea of heterogeneous nucleation due to small carbon clusters may also explain this behavior. As heterogeneous nucleation needs less supersaturation, it prevails over homogeneous nucleation. Furthermore, the density of carbon clusters is much higher than that of homogeneously nucleated oxygen precipitates. Thus, the heterogeneously nucleated oxygen precipitates cannot grow to the critical size due to their large densities.

References

5.1 P.M. Petroff, A.J.R. de Kock: J. Cryst. Growth **30**, 117 (1975)
5.2 S. Sadamitsu, S. Umeno, Y. Koike, M. Hourai, S. Sumita, T. Shigematsu: Jpn. J. Appl. Phys. **32**, 3675 (1993)
5.3 L.I. Bernewitz, B.O. Kolbesen, K.R. Mayer, G.E. Schuh: Appl. Phys. Lett. **25**, 277 (1975)
5.4 B.O. Kolbesen, A. Mühlbauer: Solid State Electron. **25**, 759 (1982)
5.5 W. Bergholz, W. Mohr, W. Drewes: Mater. Sci. Eng. **B4**, 359 (1989)
5.6 M. Itsumi, H. Akiya, T. Ueki, M. Tomita, M. Yamawaki: J. Appl. Phys. **78**(10), 5984 (1995)
5.7 M. Miyazaki, S. Miyazaki, Y. Yanase, T. Ochiai, T. Shigematsu: Jpn. J. Appl. Phys. **34**, 6303 (1995)
5.8 J.G. Park, J.M. Park, K.C. Cho, G.S. Lee, H.K. Chung: Effect of crystal defects in device characteristics, Proc. 2nd Int. Symp. Adv. Sci. Technol. Silicon Materials, Kona-Hawaii, ed. by M. Umeno (1996) p. 2519
5.9 M. Itsumi: Mater. Sci. Eng. **B73**, 184 (2000)
5.10 T. Bearda, M. Houssa, P. Mertens, J. Vanhellemont, M. Heyns: Appl. Phys. Lett. **75**(9), 1255 (1999)
5.11 U. Lambert, A. Huber, J. Grabmeier, J. Vanhellemont, R. Wahlich, G. Kissinger: Microelectron. Eng. **48**, 127 (1999)
5.12 E. Dornberger, D. Temmler, W. v. Ammon: J. Electrochem. Soc. **149**(4), G226 (2002)
5.13 C. Kupfer, H. Roth, H. Dietrich: Mat. Sci. Semicon. Proc. **5**, 381 (2003)
5.14 A.M. Eidenzon, N.I. Puzanov: Inorg. Mater. **33**(3), 272 (1997)
5.15 J.G. Park, H.K. Chung: Wafer requirements: Memory devices, Proc. Silicon Wafer Symp., Portland (1999) p. D1
5.16 J.G. Park: J. Jpn. Assoc. Cryst. Growth **27**(2), 14 (2000)
5.17 R. Falster: Advances of the defect engineering of polished silicon wafers: Perfect silicon and magic denuded zones, Proc. Silicon Wafer Symp., Portland (1999) p. E13
5.18 D. Gräf, M. Suhren, U. Lambert, R. Schmolke, A. Ehlert, W. v. Ammon, P. Wagner: J. Electrochem. Soc. **145**(1), 275 (1998)
5.19 X. Yu, D. Yang, X. Ma, L. Li, D. Que: Semicond. Sci. Technol. **18**, 399 (2003)
5.20 K. Sumino, I. Yonenaga, M. Imai, T. Abe: J. Appl. Phys. **54**(9), 5016 (1983)
5.21 L. Jastrzebski, G.W. Cullen, R. Soydan, G. Harbeke, J. Lagowski, S. Vecrumba, W.N. Henry: J. Electrochem. Soc. **134**(2), 466 (1987)
5.22 G. Wang, D. Yang, D. Li, Q. Shui, J. Yang, D. Que: Physica B **308–310**, 450 (2001)
5.23 D. Maroudas, R. Brown: Appl. Phys. Lett. **62**(2), 172 (1993)
5.24 A. Seeger, K.P. Chik: Phys. Status Solidi A **29**, 455 (1968)
5.25 H.R. Schober: Phys. Rev. B **39**, 13013 (1989)
5.26 R. Car, P.J. Kelly, A. Oshiyama, S.T. Pantelides: Phys. Rev. Lett. **52**, 1814 (1984)
5.27 D. Maroudas, R. Brown: Phys. Rev. B **47**(23), 15562 (1993)
5.28 G.D. Watkins: Mater. Sci. Semicond. Process. **3**, 227 (2000)
5.29 N.A. Stolwijk, J. Holzl, W. Frank, E.R. Weber, H. Mehrer: Appl. Phys. A **39**, 37 (1986)
5.30 H. Bracht, N.A. Stolwijk, H. Mehrer: Phys. Rev. B **52**, 16542 (1995)
5.31 H. Zimmermann, H. Ryssel: Appl.Phys. A **55**, 121 (1992)
5.32 H. Bracht: Native point defects in silicon, Proc. 3rd Int. Symp. Defects Silicon III, Seattle, ed. by T. Abe, W.M. Bullis, S. Kobayashi, W. Lin, P. Wagner (1999) p. 357
5.33 W. v. Ammon, E. Dornberger, H. Oelkrug, H. Weidner: J. Cryst. Growth **151**, 273 (1995)
5.34 M. Hourai, E. Kajita, T. Nagashima, H. Fujiwara, S. Umeno, S. Sadamitsu, S. Miki, T. Shigematsu: Mater. Sci. Forum **196–201**, 1713 (1995)
5.35 E. Dornberger, W. v. Ammon: J. Electrochem. Soc **143**(5), 1648 (1996)

5.36 T. Sinno, R.A. Brown, W. v. Ammon, E. Dornberger: Appl. Phys. Lett. **70**(17), 2250 (1997)
5.37 M. Akatsuka, M. Okui, N. Morimoto, K. Sueoka: Jpn. J. Appl. Phys. **40**, 3055 (2001)
5.38 R. Falster, V.V. Voronkov, F. Quast: Phys. Status Solidi B **222**, 219 (2000)
5.39 N. Fukata, A. Kasuya, M. Suezawa: Jpn. J. Appl. Phys. **40**, L854 (2001)
5.40 D.A. Antoniadis, I. Moskowitz: J. Appl. Phys. **53**(10), 6780 (1982)
5.41 H.J. Gossmann, C.S. Rafferty, A.M. Vredenberg, H.S. Luftman, F.C. Unterwald, D.J. Eaglesham, D.C. Jacobson, T. Boone, J.M. Poate: Appl. Phys. Lett. **64**(3), 312 (1994)
5.42 T. Sinno: Thermophysical properties of intrinsic point defects in crystalline silicon, Proc. 9th Int. Symp. Silicon Mater. Sci. Technol. Semicond. Silicon, Philadelphia, ed. by H.R. Huff, L. Fabry, S. Kishino (The Electrochemical Society, Pennington 2002) p. 212
5.43 T. Frewen, T. Sinno, E. Dornberger, R. Hoelzl, W. v. Ammon, H. Bracht: J. Electrochem. Soc. **150**(11), G673 (2003)
5.44 K. Nakamura, R. Suewaka, B. Ko: ECS Solid State Lett. **3**(3), N5 (2014)
5.45 K. Sueoka, E. Kamiyama, J. Vanhellemont, K. Nakamura: ECS Solid State Lett. **3**(6), P69 (2014)
5.46 K. Sueoka, E. Kamiyama, J. Vanhellemont: J. Appl. Phys. **114**, 153510 (2013)
5.47 T. Ueki, M. Itsumi, T. Takeda: Jpn. J. Appl. Phys. **37**, 1669 (1998)
5.48 M. Itsumi: J. Cryst. Growth **237–239**, 1773 (2002)
5.49 S. Umeno, Y. Yanase, M. Hourai, M. Sano, Y. Shida, H. Tsuya: Jpn. J. Appl. Phys. **38**, 5725 (1999)
5.50 M. Nishimura, Y. Yamaguchi, K. Nakamura, J. Jablonski, M. Watanabe: Electrochem. Soc. Symp. Proc. **13**, 188 (1998)
5.51 J. Ryuta, E. Morita, T. Tanaka, Y. Shimanuki: Jpn. Appl. Phys. **29**, L1947 (1990)
5.52 H. Yamagishi, I. Fusegawa, N. Fujimaki, M. Katayama: Semicond. Sci. Techn. **7**, A135 (1992)
5.53 P.J. Roksnoer, M.M.B. Van de Boom: J. Cryst. Growth **53**, 563 (1981)
5.54 H. Bender, J. Vanhellemont, R. Schmolke: Jpn. J. Appl. Phys. **36**, L1217 (1997)
5.55 R. Schmolke, W. Angelberger, W. v. Ammon, H. Bender: Solid State Phenom. **82–84**, 231 (2002)
5.56 K. Nakai, M. Hasebe, K. Ohta, W. Ohashi: J. Cryst. Growth **210**, 20 (2000)
5.57 H. Föll, B.O. Kolbesen: Appl. Phys. **8**, 319 (1975)
5.58 P.M. Petroff, A.J.R. de Kock: J. Cryst. Growth **36**, 1822 (1976)
5.59 J. Chikawa, T. Abe, H. Harada: Impurity effect on the formation of microdefects during silicon crystal growth. In: *Semiconductor Silicon*, ed. by H.R. Huff, T. Abe, B. Kolbesen (The Electrochemical Society, Pennington 1986) p. 61
5.60 H. Föll, U. Gösele, B.O. Kolbesen: J. Cryst. Growth **40**, 90 (1977)
5.61 R. Schmolke, M. Blietz, R. Schauer, D. Zemke, H. Oelkrug, W.V. Ammon, U. Lambert, D. Gräf: In: *Advanced Silicon Wafers for 0.18 μmDesign Rule and Beyond: Epi and fLASH!, High Purity Silicon VI, Phoenix 2000*, ed. by C.L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H.J. Dawson (The Electrochemical Society, Pennington 2000)
5.62 W. v. Ammon, E. Dornberger, P.O. Hansson: J. Cryst. Growth **198/199**, 390 (1999)
5.63 V.V. Voronkov: J. Cryst. Growth **59**, 625 (1982)
5.64 E. Dornberger, J. Esfandyari, D. Gräf, J. Vanhellemont, U. Lambert, F. Dupret, W. v. Ammon: In: *Simulation of Grown-in Voids in Czochralski Silicon Crystals, Crystalline Defects and Contamination Control: Their Impact and Control in Device Manufacturing II, Nürnberg 1997*, ed. by B.O. Kolbesen, P. Stallhofer, C. Claeys, F. Tardiff (The Electrochemical Society, Pennington 1997)
5.65 M. Hasebe, Y. Takeoka, S. Shinoyama, S. Naito: Ring-like distributed stacking faults in CZ-Si wafers. In: *Defect Control in Semiconductors*, ed. by K. Sumino (Elsevier, Amsterdam 1990) p. 157
5.66 H. Yamagishi, I. Fusegawa, K. Takano, E. Iino, N. Fujimaki, T. Ohta, M. Sakurada: In: *Evaluation of FDPs and COPs in Silicon Single-Crystals, Semiconductor Silicon, San Francisco 1994*, ed. by H.R. Huff, W. Bergholz, K. Sumino (The Electrochemical Society, Pennington 1994)
5.67 W. v. Ammon, E. Dornberger: In: *Properties of Crystalline Silicon*, EMIS Datareviews, Vol. 20, ed. by R. Hull (INSPEC, London 1999)
5.68 V.V. Voronkov, R. Falster: J. Cryst. Growth **194**, 76 (1998)
5.69 E. Dornberger, D. Gräf, M. Suhren, U. Lambert, P. Wagner, F. Dupret, W. v. Ammon: J. Cryst. Growth **180**, 343 (1997)
5.70 E. Dornberger, J. Esfandyari, J. Vanhellemont, D. Gräf, U. Lambert, F. Dupret, W. v. Ammon: In: *Simulation of Non-Uniform Grown-in Void Distributions in Czochralski Crystal Growth, Semiconductor Silicon, San Francisco 1998*, ed. by H.R. Huff, U. Gösele, H. Tsuya (The Electrochemical Society, Pennington 1998)
5.71 M. Hourai, T. Nagashima, E. Kajita, S. Miki: In: *Oxygen Precipation Behavior in Silicon During Czochralski Crystal Growth, Semiconductor Silicon, San Francisco 1994*, ed. by H.R. Huff, W. Bergholz, K. Sumino (The Electrochemical Society, Pennington 1994) p. 156
5.72 T. Iwasaki, A. Tomiura, K. Nakai, H. Haga, K. Kojima, T. Nakashizu: In: *Influence of Cooling Condition During Crystal Growth of CZ-Si on Oxide Breakdown Property, Semiconductor Silicon, San Francisco 1994*, ed. by H.R. Huff, W. Bergholz, K. Sumino (The Electrochemical Society, Pennington 1994) p. 744
5.73 K. Takano, K. Kitagawa, E. Iino, M. Kimura, H. Yamagishi: Mater. Sci. Forum **196–201**, 1707 (1995)
5.74 M. Akatsuka, M. Okui, S. Umeno, K. Sueoka: J. Electrochem. Soc. **150**(9), G587 (2003)
5.75 J. Furukawa, H. Tanaka, Y. Nakada, N. Ono, H. Shiraki: J. Cryst. Growth **210**, 26 (2000)
5.76 V.V. Voronkov, R. Falster: J. Appl. Phys. **86**(11), 5975 (1999)

5.77 A. Natsume, N. Inoue, K. Tanahashi, A. Mori: J. Cryst. Growth **225**, 221 (2001)
5.78 T. Sinno, E. Dornberger, W. v. Ammon, R.A. Brown, F. Dupret: Mater. Sci. Eng. **28**, 149 (2000)
5.79 Z. Wang, R. Brown: J.Crystal Growth **231**, 442 (2001)
5.80 A. Sattler, W. von Ammon, M. Weber, W. Haeckl, H. Schmidt: Semiconductor Wafers of Silicon and Method for Their Production, US Patent Application 8 043 427B2 (2007)
5.81 K. Tanahashi, N. Inoue: J. Mat. Sci. Mater. Electron. **10**, 359 (1999)
5.82 J. Vanhellemont, E. Kamiyama, K. Sueoka: ECS Solid State Lett. **3**(5), X3–X4 (2014)
5.83 K. Sueoka, E. Kamiyama, J. Vanhellemont, K. Nakamura: Phys. Stat. sol. B **251**(11), 2159 (2014)
5.84 E. Dornberger, W. v. Ammon, D. Gräf, U. Lambert, A. Miller, H. Oelkrug, A. Ehlert: The impact of dwell time above 900 °C during crystal growth on the gate oxide integrity of silicon wafers, Proc. 4th Int. Symp. High Purity Silicon, San Antonio, ed. by C.L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H.J. Dawson (1996) p. 140
5.85 J. Esfandyari, G. Hobler, S. Senkader, H. Pötzl, B. Murphy: J. Electrochem. Soc. **143**, 995 (1996)
5.86 V.V. Voronkov, R. Falster: J. Cryst. Growth **198/199**, 399 (1999)
5.87 V.V. Voronkov, R. Falster: J. Appl. Phys. **87**(9), 4126 (2000)
5.88 T.A. Frewen, S.S. Kapur, W. Haeckl, W. v. Ammon, T. Sinno: J. Cryst. Growth **279**, 258 (2005)
5.89 N.I. Puzanov, A.M. Eidenzon: Semicond. Sci. Technol. **7**, 406 (1992)
5.90 K. Nakamura, T. Saishoji, J. Tomioka: J. Cryst. Growth **237–239**, 1678 (2002)
5.91 V.V. Voronkov: Mater. Sci. Eng. **B73**, 69 (2000)
5.92 V.V. Voronkov, R. Falster: J. Cryst. Growth **226**, 192 (2001)
5.93 T. Abe, M. Kimura: In: *Semiconductor Silicon, 1990*, ed. by H.R. Huff, K. Barraclough, J. Chikawa (The Electrochemical Society, Pennington 1990)
5.94 W. v. Ammon, P. Dreier: Silicon bulk technology for power devices, Proc. Int. Symp. Power Semicond. Devices, Tokyo (1988) p. 134
5.95 D.-R. Yang, Y.-W. Wang, H.-N. Yao, D.-L. Que: Prog. Nat. Sci. **3**(2), 176 (1993)
5.96 W. v. Ammon, R. Hoelzl, T. Wetzel, D. Zemke, G. Raming, M. Blietz: Microelectron. Eng. **66**, 234 (2003)
5.97 W.V. Ammon, A. Ehlert, U. Lambert, D. Gräf, M. Brohl, P. Wagner: In: *Gate Oxide Related Bulk Properties of Oxygen Doped Floating Zone and Czochralski Silicon, Semiconductor Silicon, San Francisco 1994*, ed. by H.R. Huff, W. Bergholz, K. Sumino (The Electrochemical Society, Pennington 1994)
5.98 K. Nakai, Y. Inoue, H. Yokota, A. Ikari, J. Takahashi, A. Tachikawa, K. Kitahara, Y. Ohta, W. Ohashi: J. Appl. Phys. **85**(8), 4301 (2001)
5.99 J. Takahashi, K. Nakai, K. Kawakami, Y. Inoue, H. Yokota, A. Tachikawa, A. Ikari, W. Ohashi: Jpn. J. Appl. Phys. **42**, 363 (2003)
5.100 F. Shimura, R.S. Hockett: Appl. Phys. Lett. **48**, 224 (1986)
5.101 Q. Sun, K.H. Yao, H.C. Gatos, J. Lagowski: J. Appl. Phys. **71**(8), 3760 (1992)
5.102 K. Aihara, H. Takeno, Y. Hayamizu, M. Tamatsuka, T. Masui: J. Appl. Phys. **88**(6), 3705 (2000)
5.103 K. Nakai, Y. Inoue, H. Yokota, A. Ikari, J. Takahashi, W. Ohashi: Formation of grown-in defects in nitrogen doped CZ-Si crystals, Proc 3rd Int. Symp. Adv. Sci. Technol. Silicon Materials, Kona, ed. by M. Umeno (2000) p. 88
5.104 D. Gräf, U. Lambert, R. Schmolke, R. Wahlich, W. Siebert, E. Daub, W. v. Ammon: 300 mm Epi pp-wafer: Is there sufficient gettering?, Proc. 6th Int. Symp. High Purity Silicon, Seattle, ed. by C.L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H.J. Dawson (2000) p. 319
5.105 H.J. Stein: Nitrogen in crystalline silicon, Proc. Int. Symp. Oxyg. Carbon, Hydrog. Nitrogen Cryst. Silicon, Boston, ed. by J.C. Mikkelsen Jr., S.J. Pearton, J.W. Corbett, S.J. Pennycook (1986) p. 523
5.106 Y. Itoh, T. Abe: Appl. Phys. Lett. **53**(1), 39 (1988)
5.107 A. Hara, A. Ohsawa: Interaction of oxygen and other point defects in silicon crystals, Proc. Int. Symp. Adv. Sci. Technol. Silicon Materials, Kona, ed. by K. Kohra (1991) p. 47
5.108 H. Sawada, K. Kawakami: Phys. Rev. B **62**(3), 1851 (2000)
5.109 H. Kageshima, A. Taguchi, K. Wada: Appl. Phys. Lett. **76**(25), 3718 (2000)
5.110 R. Jones, S. Öberg, F.B. Rasmussen, B.B. Nielson: Phys. Rev. Lett. **72**, 1882 (1994)
5.111 K.L. Brower: Phys. Rev. B **26**, 6040 (1982)
5.112 H.J. Stein: Appl. Phys. Lett. **47**(12), 1339 (1985)
5.113 K. Murakami, H. Itoh, K. Takita, K. Masuda: Appl. Phys. Lett. **45**(2), 176 (1984)
5.114 W.V. Ammon, D. Gräf, W. Zulehner, R. Schmolke, E. Dornberger, U. Lambert, J. Vanhellemont, W. Hensel: In: *Suppression of Point Defect Aggregation in FZ Silicon Single Crystals by Nitrogen Doping; Extendend Abstracts, Semiconductor Silicon, San Diego 1998*, ed. by H.R. Huff, U. Gösele, H. Tsuya (The Electrochemical Society, Pennington 1998)
5.115 K. Nakamura, T. Saishoji, S. Togawa, J. Tomioka: The effect of nitrogen on the grown-in defect formation in CZ silicon crystals. In: *Proceedings of the Kazusa Akademia Park Forum on the Science and Technology of Silicon Materials*, ed. by K. Sumino (Kazusa Akademia Park, Chiba 1999) p. 116
5.116 V.V. Voronkov, R. Falster: J. Electrochem. Soc. **149**(3), G167 (2002)
5.117 W.B. Knowlton, J.T. Walton, J.S. Lee, Y.K. Wong, E.E. Haller, W. v. Ammon, W. Zulehner: Mater. Sci. Forum **196–201**, 1761 (1995)
5.118 T. Ono, S. Umeno, T. Tanaka, E. Asayama, M. Hourai: Behavior of defects in nitrogen doped CZ-Si crystals, Proc. Int. Symp. Forum Sci. Technol. Silicon Materials, Shonan Village Center, Kanagawa, ed. by H. Yamata-Kaneta, K. Sumino (Japan Technical Information Service, Tokyo 2001) p. 95
5.119 K. Nakamura, T. Saishoji, S. Togawa, J. Tomioka: Influence of nitrogen on the pont defect reaction in

silicon, Proc. Int. Symp. Forum Sci. Technol. Silicon Materials, Shonan Village Cent., ed. by H. Yamata-Kaneta, K. Sumino (Japan Technical Information Service, Tokyo 2001) p. 109
5.120 W. v. Ammon, R. Hölzl, J. Virbulis, E. Dornberger, R. Schmolke, D. Gräf: J. Cryst. Growth **226**(1), 19 (2001)
5.121 P. Wagner, R. Oeder, W. Zulehner: Appl. Phys. A **46**, 73 (1988)
5.122 W. v. Ammon, P. Dreier, W. Hensel, U. Lambert, L. Köster: Mater. Sci. Eng. **B36**, 33 (1996)
5.123 M.W. Qi, S.S. Tan, B. Zhu, P.X. Cai, W.F. Gu, M. Xu, T.S. Shi, D.L. Que, L.B. Li: J. Appl. Phys. **69**, 3775 (1991)
5.124 A. Gali, J. Miro, P. Deak, C. Ewels, R. Jones: J. Phys. Condens. Mat. **8**, 7711 (1996)
5.125 X. Yue, J. Chen, X. Ma, D. Yang: Mater. Sci. Eng. **R74**, 1 (2013)
5.126 H.Ch. Alt, H.E. Wagner: J.Appl.Phys. **106**, 103511 (2009)
5.127 M. Suhren, D. Gräf, U. Lambert, P. Wagner: Crystal defects in highly boron doped silicon, Proc. 4th Int. Symp. High Purity Silicon, San Antonio, ed. by C.L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H.J. Dawson (The Electrochemical Society, Pennington 1996) p. 132
5.128 K. Nakamura, R. Suewaka, T. Saishoji, J. Tomioka: The effect of impurities on the grown-in defects in CZ-Si crystals (B, C, N, O, Sb, As, P), Proc. Forum Sci. Technol. Silicon Mater., ed. by H. Yamad-Kaneta, K. Sumino (2003) p. 161
5.129 W. v. Ammon: Crystal growth of large diameter CZ Si crystals, Proc 2nd Int. Symp. Adv. Sci. Technol. Silicon Materials, Kona, ed. by M. Umeno (1996) p. 233
5.130 J. Vanhellemont, E. Kamiyama, K. Sueoka: ECS J. Solid State Sci. Technol. **2**, 166 (2013)
5.131 T. Sinno, H. Susanto, R. Brown, W. v. Ammon, E. Dornberger: Appl. Phys. Lett. **75**, 1544 (1999)
5.132 S. Ma, S. Wang: Phys.Rev. B **81**, 193203 (2010)
5.133 T. Abe, T. Masui, H. Harada, J. Chikawa: In: *VLSI Science and Technology, 1985*, ed. by W.M. Bullis, S. Broyda (The Electrochemical Society, Pennington 1985)
5.134 R. Takeda, T. Minami, H. Saito, Y. Hirano, H. Fujimori, K. Kashima, Y. Matsushita: Influence of LSTD size on the formation of denuded zone in hydrogen-annealed CZ silicon wafers, Proc. 6th Int. Symp. High Purity Silicon, Phoenix, ed. by C.L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H.J. Dawson (The Electrochemical Society, Pennington 2000) p. 331
5.135 M. Porrini, V.V. Voronkov, R. Falster: Mater. Sci. Eng. B **134**, 185 (2006)
5.136 S. Kishino, M. Kanamori, N. Yoshihizo, M. Tajima, T. Iizuka: J. Appl. Phys. **50**, 8240 (1978)
5.137 T. Fukuda: Appl. Phys. Lett. **65**(11), 1376 (1994)
5.138 F. Shimura: J. Appl. Phys. **59**, 3251 (1986)
5.139 M. Porrini: Cryst. Res. Technol. **40**(10/11), 1054 (2005)
5.140 W. Sugimura, T. Ono, S. Umeno, M. Hourai, K. Sueoka: ECS Transactions **2**(2), 95 (2006)
5.141 V.V. Voronkov, R. Falster, M. Porrini, J. Duchini: Phys. Status Solidi A **209**(10), 1898 (2012)
5.142 M. Porrini, J. Duchini, A. Bazzali: Crystal. Res. Technol. **49**(8), 564 (2014)
5.143 T. Abe, H. Harada, J. Chikawa: Mat. Res. Soc. Symp. Proc. **14**, 1 (1983)
5.144 V.V. Voronkov, R. Falster: J. Electrochem. Soc. **149**, G167 (2002)
5.145 G. Borionetti, D. Gambaro, M. Porrini, V.V. Voronkov: Grown-in microdefect distribution in doped silicon crystals. In: *Semiconductor Silicon 2002–2*, ed. by H.R. Huff, L. Fabry, S. Kishino (Electrochemical Society, Pennington 2002) p. 505
5.146 K. Sueoka, M. Akatsuka, K. Nishihara, T. Yamamoto, S. Kobayashi: Mater. Sci. Forum **196–201**, 1737 (1995)
5.147 J. Vanhellemont, C. Claeys: J. Appl. Phys. **62**(9), 3960 (1987)

6. Diffusion in Semiconductors

Derek Shaw

Atomic diffusion in semiconductors refers to the migration of atoms, including host, dopant and impurities. Diffusion occurs in all thermodynamic phases, but the solid phase is the most important in semiconductors. There are two types of semiconductor solid phase: amorphous (including organic) and crystalline. In this chapter we consider crystalline semiconductors and describe the processes by which atoms and defects move between lattice sites. The emphasis is on describing the various conditions under which diffusion can occur, as well as the atomic mechanisms that are involved, rather than on tabulating data. For brevity's sake, we also focus on the general features found in the principal semiconductors from Groups IV, III–V and II–VI; IV–VI and oxide semiconductors are excluded from consideration. It is not surprising that most of the data available in this field relate to the semiconductors that are technologically important – they are used to fabricate electronic and optoelectronic devices. One unavoidable consequence of this technological need is that diffusion data tend to be acquired in a piecemeal fashion.

6.1 **Basic Concepts** 134
6.2 **Diffusion Mechanisms** 134
6.2.1 Vacancy and Interstitial Diffusion Mechanisms 135
6.2.2 The Interstitial–Substitutional Mechanism: Dissociative and Kick-Out Mechanisms 135
6.2.3 The Percolation Mechanism 135
6.3 **Diffusion Regimes** 136
6.3.1 Chemical Equilibrium: Self- and Isoconcentration Diffusion 136
6.3.2 Chemical Diffusion (or Diffusion Under Nonequilibrium Conditions) 136
6.3.3 Recombination-Enhanced Diffusion 138
6.3.4 Surface Processing Effects 138
6.3.5 Short-Circuit Paths 138
6.4 **Internal Electric Fields** 138
6.5 **Measurement of Diffusion Coefficients** 139
6.5.1 Anneal Conditions 139
6.5.2 Diffusion Sources 139
6.5.3 Profiling Techniques 139
6.5.4 Calculating the Diffusion Coefficient 140
6.6 **Hydrogen in Semiconductors** 140
6.7 **Diffusion in Group IV Semiconductors** 141
6.7.1 Germanium 141
6.7.2 Silicon 141
6.7.3 $Si_{1-x}Ge_x$ Alloys 141
6.7.4 Silicon Carbide 142
6.8 **Diffusion in III–V Compounds** 143
6.8.1 Self-Diffusion 143
6.8.2 Dopant Diffusion 143
6.8.3 Compositional Interdiffusion 144
6.9 **Diffusion in II–VI Compounds** 144
6.9.1 Self-Diffusion 145
6.9.2 Chemical Self-Diffusion 145
6.9.3 Dopant Diffusion 145
6.9.4 Compositional Interdiffusion 146
6.10 **Nano Volume Diffusion** 146
6.11 **Diffusion in Molten Semiconductors** 146
6.12 **The Meyer–Neldel Rule** 146
6.13 **Conclusions** 147
6.14 **General Reading and References** 147
References 147

Diffusion describes the movement of atoms through space, primarily due to thermal motion, and it occurs in all forms of matter. This chapter focuses on atom diffusion in crystalline semiconductors, where diffusing atoms migrate from one lattice site to adjacent sites in the semiconductor crystal. The diffusion of atoms and defects (both native and dopant) is at the heart of semiconductor processing. This includes the growth, post-growth and device construction stages. To achieve required goals, it is essential to control the diffusion aspects which calls for an understanding of what is actually diffusing in a given situation. The needs of device technology have provided the main impetus for investigating atomic diffusion in a semiconductor lattice. As the physical dimensions of devices have shrunk, the problems relating to the understanding of diffusion features in complex structures have multiplied. Common examples relate to degra-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_6

dation of a doped structure, such as a p-n junction or super lattice, and stability of a metal contact or diffusion barrier. *Knez* [6.1] identifies four separate diffusion situations that can arise in the postprocessing of a surface layer located on a substrate (which may itself be a thin layer, for example, HgTe on CdTe):

1. Lateral diffusion of components in the surface layer
2. Diffusion of surface components into the substrate
3. Diffusion of substrate components into the surface layer
4. A diffusion barrier between the surface layer and substrate.

Example 4 is provided by a Cu contact to a Si substrate in which graphene [6.2], Ta_2O_5 and TiO_2 [6.3] films provide diffusion barriers. The graphene films are approximately or equal to a few μm thick in contrast to the metal films which lie between 1 and 4 nm in thickness. The investigation of diffusion phenomena is pursued by both experimental means and by ab initio calculations: the latter are increasingly important in identifying which native defects may be involved.

6.1 Basic Concepts

The mobility of an atom or defect in the crystal lattice is characterized by its diffusivity or diffusion coefficient, D. Diffusion of a species is observed when a concentration gradient exists so that there is a flow down the concentration gradient. For a one-dimensional situation with flow parallel to the x-axis Fick [6.4] gave

$$J = -D\frac{\partial c}{\partial x}\,, \tag{6.1}$$

where J is the flux of atoms or defects per unit area per second and c is the concentration of the diffusing species. Equation (6.1) leads to [6.4]

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial^2 x}\,. \tag{6.2}$$

If the lattice is semi-infinite beyond $x = 0$ and the surface concentration at $x = 0$ is maintained at a constant value of c_0 (6.2) can be solved to give [6.2]

$$c = c_0\,\mathrm{erfc}\left[\left(x/2\sqrt{(Dt)}\right)\right], \tag{6.3}$$

where erfc is the complementary error function. Alternatively if only a fixed amount, M, of the diffusant is deposited on the surface at $x = 0$ and $t = 0$, then (6.2) has the solution [6.4]

$$c = \frac{M}{2\sqrt{\pi Dt}}\exp(-x/4Dt)\,. \tag{6.4}$$

As the kinetic energy, and therefore mobility, of a species increases with temperature, it is found that quite generally D obeys an Arrhenius relation

$$D = D_0\exp(-Q/k_\mathrm{B}T)\,, \tag{6.5}$$

where k_B and T are Boltzman's constant and the temperature, respectively: D_0 and Q are the parameters determined by the diffusion mechanism of the particular atom or defect. Q is known as the activation enthalpy or energy for the diffusion process.

At the atomic level an atom or defect migrates by successive jumps between neighboring lattice sites. In a three-dimensional isotropic lattice consider a species that migrates by jumping from one lattice site to a nearest neighbor site. If the distance between nearest neighbor sites is a and the atom or defect makes n random jumps in time t then the mean-square displacement $\langle R^2\rangle = na^2$ so using Einstein's relation [6.5]

$$D = \frac{na^2}{6t}\,. \tag{6.6}$$

Equation (6.6) applies to random jumps (i. e., successive jumps are completely independent) whereas depending on the diffusion mechanism some correlation between successive jumps can occur. This is accounted for by introducing a correlation factor, f, into the right-hand side of (6.6) ($f \le 1$) [6.5]. Some idea of the magnitudes involved is provided by taking $D = 10^{-12}\,\mathrm{cm}^2/\mathrm{s}$, $t = 10^4\,\mathrm{s}$, $a = 2\times10^{-8}\,\mathrm{cm}$, and $f = 1$ so that the diffusion length $2\sqrt{(Dt)} = 2\,\mu\mathrm{m}$ whereas the total distance traveled $na = 3\,\mathrm{cm}$.

6.2 Diffusion Mechanisms

Two categories of diffusion mechanisms are recognized: defect and nondefect. A simple example of the latter class is the simultaneous jumps of two adjacent atoms in order to exchange sites. There is a general consensus, however, that nondefect mechanisms do not play any significant role in semiconductor diffusion, although nondefect contributions have recently been proposed for self- and dopant diffusion in Si [6.6]. In the elemental semiconductors Si and Ge, vacancies and interstitials are the primary defects. Historically a vacancy was

known as a Schottky defect. In binary compound semiconductors (such as GaAs, GaN, CdTe, ZnO, HgCdTe) there are two sublattices, the anion and cation, so there are vacancies and interstitials for each sublattice together with antisite defects on each sublattice. There is also the possibility of complexes forming between native defects such as the Frenkel pair [6.5] (an interstitial–vacancy complex) and di-vacancies as well as between native defects and dopant species. Further complexity arises due to the various states of ionization of the defects: the defect mobility will depend on its charge state. The relative concentrations of the different charge states will be determined by the position of the Fermi level. It is also important to know in a given situation whether the crystal is electrically intrinsic or extrinsic.

6.2.1 Vacancy and Interstitial Diffusion Mechanisms

In the Si lattice, a vacancy V_{Si} can migrate by a nearest neighbor Si atom jumping into the vacancy (in other words the Si atom and V_{Si} have exchanged sites so that the Si atom has also migrated). Equally a substitutional dopant atom can migrate by jumping into a V_{Si} at a nearest neighbor site. Similarly, in a binary semiconductor such as GaAs, Ga atoms can migrate over the Ga sublattice via jumps into nearest neighbor Ga vacancies, as can dopant atoms substituting on Ga sites. Antisite defects can diffuse by jumping into vacancies in the same sublattice, such as the As antisite in GaAs, As_{Ga}, diffusing by jumps into an adjacent Ga vacancy. For self-interstitials, such as Si_i or Ga_i, their concentrations are sufficiently small for neighboring interstitial sites to be always empty, which means that the occupancy of nearest neighbor sites is not a factor when determining jump rates. If self-interstitial, such as Si_i, pushes an Si atom on a normal lattice site into an interstitial site instead of jumping into a neighboring interstitial site, thereby replacing the displaced Si atom, the process is the *interstitialcy mechanism*. This concept extends to a substitutional dopant atom forming a pair with a self-interstitial, which then migrates with the dopant atom alternating between substitutional and interstitial sites. The *dissociative mechanism* refers to the case of a dopant occupying both interstitial and substitutional sites with migration via interstitial sites. It is also possible for point defects to form complexes which can diffuse as a single entity. Examples are the Frenkel pair $V_{Si}Si_i$, the defect pair formed between a substitutional dopant atom and an adjacent vacancy, and di-vacancies such as $V_{Ga}V_{Ga}$, $V_{Ga}V_{As}$ or $V_{As}V_{As}$. In the case of a mobile (cation–anion) vacancy pair it is feasible that a link can exist between the cation and anion self-diffusivities [6.7].

6.2.2 The Interstitial–Substitutional Mechanism: Dissociative and Kick-Out Mechanisms

The interstitial–substitutional diffusion mechanism arises when a dopant species Z occupies both interstitial and substitutional sites, represented by Z_i and Z_s, respectively, and diffusion is restricted to jumps of Z_i. In this case, we may ask how the Z_s concentration, $[Z_s]$, is linked to the Z_i migration. Consider the diffusion of Z in Si. The dissociative mechanism (also known as the Frank–Turnbull mechanism) is based on the defect interaction

$$Z_i + V_{Si} \rightleftharpoons Z_s \tag{6.7}$$

and application of the law of mass action (LMA) leads to $[Z_i][V_{Si}] \propto [Z_s]$. In the kick-out mechanism, the defect interaction is

$$Z_i + Si_s \rightleftharpoons Z_s + Si_i \tag{6.8}$$

and therefore $[Z_i] \propto [Z_s][Si_i]$. $[Si_s]$ is omitted from equation (6.8) as it is effectively constant because $[Z_i]$, $[Z_s]$, and $[Si_i]$ are generally many orders of magnitude smaller. In order to sustain growth of $[Z_s]$ by either mechanism, it is clearly necessary to have either a supply of V_{Si} or a means of removing Si_i. For simplicity, neutral charge states have been assigned to all the defects in (6.7) and (6.8). A detailed treatment of the kick-out mechanism has been given by *Frank* et al. [6.8]. For in-diffusion of Z, the Frank–Turnbull mechanism consumes vacancies and will therefore tend to reduce the local vacancy concentration, whereas the local self-interstitial concentration will be enhanced by the kick-out mechanism. Out-diffusion of Z reverses the effects on the native defect concentrations.

6.2.3 The Percolation Mechanism

The percolation mechanism [6.9] was proposed to explain group V dopant diffusion in Si at high dopant concentrations ($> \approx 10^{20}\,cm^{-3}$). At low concentrations, diffusion is via dopant–V_{Si} pairs. As the dopant concentration increases, regions occur in the Si lattice where the proximity of the dopant atoms enhances the mobility and concentration of the V_{Si}. Within this network, the diffusivity of dopant–V_{Si} pairs is thereby also enhanced so that the dopant diffusivity increases overall. The percolation network only forms once the dopant concentration exceeds a certain critical value. In principle, this mechanism could extend to other highly doped materials.

6.3 Diffusion Regimes

The mobility of a native defect and/or dopant atom reflects the physical and chemical environment under which diffusion is occurring. Two types of environment arise: conditions of chemical equilibrium and those of chemical nonequilibrium. Diffusion in temperature gradients is excluded although this can occur during cooling (heating) from (to) a diffusion anneal – only isothermal conditions are considered.

6.3.1 Chemical Equilibrium: Self- and Isoconcentration Diffusion

Chemical equilibrium means that the concentrations of all chemical components, including native defects, are uniform and stable throughout the semiconductor sample, and where appropriate (such as in a compound material), the solid is in equilibrium with the ambient vapor of the components so that the level of nonstoichiometry is both defined and controlled. Experimentally this requires diffusion to be carried out in a sealed system. Self-diffusion refers to the diffusion of the host atoms, such as Si atoms diffusing in the Si lattice. Isoconcentration diffusion describes the diffusion of dopant atoms when the same dopant concentration is uniform throughout the sample, such as for As diffusion in Si when (As) is constant. In either case, diffusion can only be observed if some of the diffusing atoms are tagged, such as by using a radioisotope or an isotopically enriched diffusion source. The diffusivity of a tagged or tracer species is related to the concentration of the native defect that provides the diffusion path. Several native defects may be involved, together with their various states of ionization (determined by the Fermi level). Specific relations can be found in *Shaw* [6.10]. A diffusion flux of the tracer arises from a concentration gradient of the tracer, which is not to be confused with a chemical concentration gradient.

6.3.2 Chemical Diffusion (or Diffusion Under Nonequilibrium Conditions)

This category contains all of the diffusion phenomena that are of technological interest and importance. In this case, diffusion occurs due to spatial gradients in the concentrations of the chemical components in the material, which in turn are caused by departures from equilibrium: the diffusion processes are attempting to either restore or achieve equilibrium.

Chemical Self-Diffusion

Chemical self-diffusion describes the process whereby a compound semiconductor changes from one level of nonstoichiometry to another through changes in the native defect populations: there is no change in the chemical composition. Changes in nonstoichiometry can arise due to a change in the ambient partial pressure of one the components (e.g., that of As in the case of GaAs), or through a change in temperature under a controlled component partial pressure. Good examples are provided by the II–VI semiconductors [6.11] and in particular $Hg_{0.8}Cd_{0.2}Te$ [6.12], where p- to n-type conversion is used to form p-n junctions by annealing in Hg vapor. High Hg partial vapor pressures give n-type due to cation donor interstitials and low pressures yield p-type arising from cation vacancies. Changes in the native defect concentrations can lead also to either an increase or decrease in a dopant diffusivity. Equally if two dopants are present they can modify each others diffusivity.

Dopant Diffusion

Suppose we have a dopant diffusing into the semiconductor from a source located at an external surface (such as a surface layer) or in an external phase (such as a gas, vapor, or liquid). The existence of the dopant concentration gradient can lead to various effects which influence the dopant flux. For example, if the dopant diffuses via a vacancy mechanism, then at any position in the diffusion region the increase in the dopant concentration requires a supply of vacancies, so that to maintain local defect equilibrium there must also be an associated vacancy flux. If the dopant controls the position of the Fermi level, then the concentration of ionized native defects will increase or decrease, depending on their charge state, relative to their intrinsic concentrations (the Fermi-level effect). This means that the concentrations of defects having opposite (the same) polarity to the dopant will be increased (decreased). Increases in the concentrations of ionized native defects due to Zn, Si, or Te diffusion into GaAs/GaAlAs superlattice structures explain the disordering of such structures [6.13]. Usually a substitutional dopant will have a different size to that of the host atom it replaces. This size difference creates a local mechanical strain in the lattice which in turn can cause changes in the local concentrations of native defects as well as to jump rates and hence dopant diffusivity [6.14]. If the strain is large enough, misfit dislocations will be generated [6.15], otherwise there will be a strain energy gradient matching the dopant concentration gradient which can enhance or retard the dopant flux [6.16]. A particular scenario is dopant diffusion in strained epilayers. Whether diffusion is enhanced or retarded is a complex matter [6.17]. However, significant effects

are found, such as the decrease in the B diffusivity in strained $Si_{1-x}Ge_x$ epilayers by a factor ≈ 10 as the strain increase from 0 to 0.64 [6.18].

Compositional Interdiffusion (CID)

Compositional interdiffusion (CID) describes component diffusion across the initial interface separating two semiconductors of different chemical composition, such as the interface between an epilayer and its substrate. Chemical composition here refers to major components: dopants and deviations from stoichiometry are excluded as their contribution to the concentrations of the primary chemical components is negligible. The result is a mixing of the components to achieve an ultimate uniform composition. One consequence of this intermixing is the movement of lattice planes known as the Kirkendall effect [6.5]. CID can be exploited in the formation of graded band-gap structures and during material preparation, such as in $Hg_{1-x}Cd_xTe$, where an alternating sequence of HgTe and CdTe epilayers of appropriate thickness are first grown and then interdiffused. CID can also pose problems in the fabrication of multiple quantum wells and superlattice structures when sharp boundaries are necessary. In particular, donor or acceptor diffusion into GaAs-based superlattices can cause essentially complete intermixing on the cation sublattice [6.13]. This phenomenon is also known as diffusion-induced disorder. B or As doping also results in rapid intermixing at a Si/Ge interface [6.19].

Transient-Enhanced Diffusion

Ion implantation is a common procedure for doping a semiconductor surface layer. The implantation process, however, does create a significant amount of lattice damage so that a subsequent anneal stage is required in order to a achieve both full electrical activity of the implanted dopant and recovery from the lattice damage. During the implantation process, the implant ions create collision cascades of vacancies and self-interstitials (an excess of native defects). The postimplant anneal serves to remove or reduce this excess. In addition to vacancy–interstitial recombination, the excess native defects can interact to form clusters (which may also contain implant ions as well as residual impurities) and extended defects such as dislocations and dislocation loops. At the start of the postimplant anneal, the local concentrations of vacancies and self-interstitials in the implant region can greatly exceed equilibrium values and therefore enhance the diffusivity of the implant ion in the implant region. As annealing proceeds, the excess concentrations will diminish and will be reflected in a decreasing dopant diffusivity until values appropriate to local equilibrium are reached. This temporary enhancement in the dopant diffusivity is known as transient enhanced diffusion (TED). The topic is a complex one to analyze quantitatively and detailed consideration of the issues involved in the case of B implants in Si can be found in the review by *Jain* et al. [6.20]. In the fabrication of shallow p-n junctions using ion implants and rapid thermal annealing (RTA), TED can determine the lower limit to the junction depth. TED of B in Si can be reduced with coimplants of Si prior to RTA [6.21]. TED of Be and Si in GaAs has also been described [6.22]. Some workers use the term TED to describe the enhanced diffusivity of a dopant, incorporated during growth, which occurs when the structure is annealed at a lower temperature than the growth temperature, with the consequence that there is an initial supersaturation of the relevant native defects at the anneal temperature.

Segregation, Gettering, Precipitation, and Clustering

A variety of important scenarios arise, involving many of the above regimes, during growth and/or subsequent processing stages of materials and structures. The segregation of acceptor dopants in InP [6.23] and in III–V heterostructures [6.24, 25] has been observed and modeled. The segregation (or accumulation) due to diffusion of the dopant is in effect a partitioning to preferred (higher solubility) regions within the layer structure. Segregation to a surface or boundary can also occur. Gettering describes the segregation, or cleaning up, of a fast diffusing impurity from the active regions of a device structure. Such impurities are typically Group 1B and transition metals and are incorporated either during growth or during subsequent processing. Gettering sites in Si are provided by O precipitates, self-ion implant damage, and nanocavities [6.26]. In contrast, Group 1B impurities in $Hg_{1-x}Cd_xTe$ are gettered by regions of high cation vacancy concentrations [6.27].

Precipitation occurs when a species – whether native defect, dopant, or impurity – becomes supersaturated, and in order to achieve its equilibrium concentration the species excess is removed by the formation of precipitates within the host lattice. A self-interstitial or vacancy excess can be removed through the nucleation of dislocation loops, stacking faults, or voids which then provide sites for precipitation of the remaining excess. In the case of a dopant, impurity, or stoichiometric excess, nucleation of a precipitate can be spontaneous (homogenous) or heterogenous. The latter occurs at the site of an impurity atom (e.g., C atoms in Si serve as nucleation centres for the precipitation of O) or at dislocations, giving rise to the term *decoration*. Growth of any precipitate proceeds via diffusion of the particular species from solution in the matrix to the precipitate and is generally diffusion limited. Equations describing the growth

of precipitates [6.28–31] and voids [6.32] are available. Invariably local stress fields will be present which influence the diffusion and, if present initially, they may also play a role in the nucleation stage. The precipitation and out-diffusion of a species during cooling has also been analyzed [6.33]. The precipitation of O impurities in Si presents a unique case study because of the high [O], its technological importance, and its complexity [6.34, 35]. The rather simpler case of B precipitation in Si has been described by *Solmi* et al. [6.36]. A cluster (or agglomerate) refers to a close configuration of at least a few dopant atoms (with or without associated native defects) or, for a compound, host species. *Solmi* and *Nobili* [6.37] have identified $(2As_{Si} - V_{Si})^0$ and $(4As_{Si} - V_{Si})^+$ clusters in heavily As-doped Si. Heavy C doping [6.38] and B implants [6.39] in Si give rise to self-interstitial clusters with C and B, respectively. In Si, according to *Ortiz* et al. [6.40], if the number of self-interstitials in a cluster exceeds ≈ 10, there is a transition to a {113} defect.

6.3.3 Recombination-Enhanced Diffusion

The local energy released in the nonradiative recombination of excess free carriers can help a diffusing species to surmount the energy barrier separating it from an adjacent lattice site – in other words, the energy barrier facing a jump is effectively reduced. This situation is important in the degradation in performance of device structures which utilizes high excess minority carrier concentrations, such as light-emitting and laser diodes. Transient excess free carriers also arise when a surface is exposed to energetic particle beams as in electron beam microscopy and secondary ion mass spectrometry (SIMS) analysis.

6.3.4 Surface Processing Effects

The concentration of native point defects within the interior of a sample can be altered by surface processes. In the case of Si it is well known that during surface oxidation or nitridation there is injection of Si interstitials or vacancies, respectively. All diffusants can therefore be affected during the processing stage. Ion-beam milling injects Hg interstitials into $Hg_{1-x}Cd_xTe$ in sufficient quantities to convert p-type to n-type. The effects of surface layers can also be significant, for example, an n-type layer on a p-type substrate will have an electric field across the junction which will also be the case if there is a difference in band gaps between the surface layer and substrate.

6.3.5 Short-Circuit Paths

The existence of dislocations and subgrain boundaries in single-crystal material generally provides high diffusivity routes for all atomic species relative to the surrounding lattice. Care is always needed when evaluating experimental data to ensure that bulk or volume diffusion is not masked by short-circuit paths [6.40]. In the case of polycrystalline Si, the grain boundaries may provide high diffusivity routes, as in the cases of As and B [6.41], or retard diffusion, as for Au [6.42]. The situation is a complex one, as grain growth can also occur during any anneal. Ge and Si can both be obtained free of dislocations which is not the case for compound semiconductors. Analysis of diffusion profiles in II–VI compounds shows the diffusant diffusivity to be a factor $\approx 10^{-10}$ greater in a dislocation than in the bulk [6.43]. *Kaur* et al. [6.44] provide a comprehensive account of short-circuit path diffusion.

6.4 Internal Electric Fields

If a concentration gradient of a dopant exists and the semiconductor is electrically extrinsic, free carriers from the dopant, due to their much greater mobility, will diffuse ahead of the parent dopant atoms. This separation creates a local electric field whose direction is such as to pull the ionized dopant atoms after the free carriers (and also to pull the free carriers back). There will also be an associated local space charge. However, if the space charge is negligible then [6.45]

$$2n_i \sinh\left(\frac{qV}{k_B T}\right) = [D^+] - [A^-]\,, \tag{6.9}$$

where V is the local electrostatic potential arising from the ionized donor(s), D^+ and acceptor(s), A^-. q, k_B, and T are the electronic charge, Boltzman's constant, and the absolute temperature, respectively. Neglect of the space charge is justified provided that the diffusion length, $2\sqrt{(Dt)}$, greater than equal to six Debye screening lengths (typically $\approx 10^2$ nm) [6.46]. For planar diffusion parallel to the x-axis the internal electric field, E, due to dopant concentration gradients is obtained from (6.9) as $E = -\partial V/\partial x$. E exerts a force on each ionized dopant atom so for an ionized donor E will create an additional flux $\propto -D(D^+)\partial[D^+]/\partial x$ where $D(D^+)$ is the donor diffusivity. E will cause drift of any other ionized species. For ionized acceptors with $[A^-] < [D^+]$ their drift will be toward $x = 0$. Equation (6.9) is based on the Fermi level being determined by

Maxwell–Boltzman statistics. The situation for the degenerate case requiring Fermi–Dirac statistics has been considered by *Nilsson* [6.47] and *Kroemer* [6.48]

Internal electric fields can arise in other circumstances such as in depletion layers where E must be calculated from Poisson's equation, in graded bandgap structures [6.16, 49], and at the interfaces of heterostructures. Cubic II–VI and III–V strained layer heterostructures grown on the {111} direction are piezoelectric and typical strains from lattice mismatch of $\approx$ 1% can give $E \approx 10^5\,\mathrm{V/cm}$ in the absence of free carrier screening [6.50]. In wurtzite heterostructures based on the Ga and In nitrides, even higher fields are found ($E \approx 10^6\,\mathrm{V/cm}$) due to piezoelectric and spontaneous polarization [6.51]. These fields can be important in CID and chemical self-diffusion.

6.5 Measurement of Diffusion Coefficients

6.5.1 Anneal Conditions

Accurate control of the sample temperature and its ambient conditions are essential if reproducible results are to be obtained in a diffusion anneal. Depending on the time spent at the anneal temperature, the warm-up and cool-down times may also be important. In particular, the cool down maybe a quench or a programmed cooling. An appropriate choice of ambient is needed to preserve the sample surface (to avoid evaporation, surface melting or alloying with the diffusant source for example). For compound semiconductors it is necessary to define the level of nonstoichiometry by controlling the ambient partial vapor pressure of one of the components, such as As for GaAs or Hg for $Hg_{1-x}Cd_xTe$, as this will be a factor in determining the dopant concentration. If the dopant is present in an external phase, rather than an ion implant, knowledge of the phase diagram of all of the components is required [6.52]. Control over the partial pressure is best achieved in a sealed system, typically a fused silica ampoule. Annealing in a vertical or horizontal resistance-heated furnace requires a minimum anneal of 30 to 60 min to avoid uncertainties due to warm-up and cool down. The drive to shallow dopant profiles has been facilitated through RTA techniques. These are based on radiant heating of the sample, and linear heating rates of 100–400 °C/s with cooling rates of up to 150 °C/s are available. RTA however precludes the use of a sealed system and, in this case, a popular means of preventing surface deterioration is to seal the sample with an inert, impervious capping layer, such as SiN.

6.5.2 Diffusion Sources

Consideration is restricted to planar samples with diffusion normal to a principal face. This is a common situation and diffusion of a dopant or tracer species can take place from:

1. A surrounding vapor or gas phase.
2. A surface layer, which may be evaporated, chemically deposited (CVD), or a spun-on silicate glass, all incorporating the diffusant.
3. Epilayers containing the diffusant, which may provide the external surface or be buried within the epitaxial structure.
4. Ion implants of a dopant either directly into the sample surface or into a thin surface layer so as to avoid lattice damage.

It is obviously desirable that negligible diffusion occurs prior to reaching the anneal temperature and similarly there is negligible out-diffusion during the cool down. In self-diffusion experiments, the tracer can be a radiotracer or an isotopically enriched species. A key requirement for either form of tracer is availability, and a radiotracer must have a half-life that is long enough for the experiment to be completed.

6.5.3 Profiling Techniques

Determining the spatial distribution of a diffusant for various anneal times is fundamental to obtaining its diffusion coefficient or diffusivity. Most methods are destructive, as they generally require a bevel section through the diffusion zone or the sequential removal of layers which are normal to the diffusion direction. The two broad profiling techniques are electrical and species specific. Electrical methods are primarily the p-n junction method, spreading resistance and capacitance–voltage profiling. Limitations of the electrical methods are: (i) assumptions are needed to link the electrical data to the diffusant concentration (such as, the diffusant is the only electrically active centre involved and that it is fully ionized); (ii) the assumption that the anneal temperature defect situation is *frozen-in* during cool down (there is no out-diffusion, precipitation, or segregation). Electrical methods are the most direct means of measuring chemical self-diffusivities and can readily detect changes in host concentrations of < 1 part in 10^4. Specie-specific (chemical element

or isotope) profiling means that the chemical concentration of the diffusant is determined regardless of its location(s) in the lattice and of its electrical state. Profiling of the diffusant using a radiotracer has been widely used [6.53], but has now been largely supplanted by secondary ion mass spectrometry (SIMS) which is now the standard procedure for diffusant profiling. This is because SIMS can measure atomic concentrations within the range 10^{16}–$10^{22}\,\mathrm{cm}^{-3}$ with spatial resolutions of at least nanometres per decade of concentration. In favorable cases, the detection limit can approach $10^{13}\,\mathrm{cm}^{-3}$ (1 ppb). Primary factors determining the resolution are progressive roughening of the eroded surface and *knock-on* effects due to the probing ion beam displacing the diffusant to greater depths. A further issue arises when the atomic mass of the diffusant is close to that of the host species.

Nondestructive profiling techniques applicable to CID in quantum well and superlattice structures utilize either high-resolution x-ray diffraction (HRXRD) [6.54] or photoluminescence (PL). The detail in HRXRD patterns reflects the CID profiles at the interfaces and can also reveal the presence of strain in the structures. The use of PL requires the presence of optically active centres in the quantum well. CID changes the shape and depth of the quantum well, which in turn changes the photon energies in PL spectra. HRXRD and PL can also be combined. Neutron reflectometry has been used to measure Si self-diffusion in $^{29}\mathrm{Si}/^{28}\mathrm{Si}$ isotope multilayers [6.55]. Diffusion lengths were in the range 1–6 nm. A particularly useful feature of these nondestructive techniques is that they allow successive anneals to be performed on the same sample.

6.5.4 Calculating the Diffusion Coefficient

Once a planar diffusion profile, representing volume or bulk diffusion has been obtained, the first step is to see if the profile can be fitted to either (6.3) or (6.4), so providing a value for D. If, despite fulfilling the boundary conditions necessary for (6.3) or (6.4), a fit to either of these equations is not possible this indicates that D is a function of c. The dependence of D on c can be established by a Boltzman–Matano analysis provided $c_0(t)$ is constant and $c_{x,t}/c_0$ varies as $x/\sqrt{t}$ [6.52]. More complex situations and profiles require numerical integration of the appropriate diffusion equation(s) with matching to the experimental $c(x)$ versus x profile: in other words, a suitable model with adjustable parameters is used to simulate the observed profiles. The interpretation of HRXRD and/or PL data provides a good example of a simulation scene in which an assumed D, either c-dependent or -independent, is used to calculate the resulting CID profile and its effect on the x-ray patterns and/or PL spectra. Whereas SIMS can detect diffusivities as low as $\approx 10^{-19}\,\mathrm{cm}^2/\mathrm{s}$ the HRXRD limit is $\approx 10^{-23}\,\mathrm{cm}^2/\mathrm{s}$. It is not unusual however in the case of complex profiles to fit consecutive sections of the profile to a Gaussian (i. e., $\exp(-x^2/4Dt)$) dependence and so obtain several diffusivities. This procedure is wholly empirical and, although providing diffusivities with the right order of magnitude, provide little insight of the diffusion mechanisms.

6.6 Hydrogen in Semiconductors

Hydrogen is a ubiquitous element in semiconductor materials. It can be incorporated either deliberately, or inadvertently during growth, or in subsequent processing involving surface treatments where organic solvents, acid, or plasma etching are used. H occupies interstitial sites as H^-, H^+, H^0, and H_2. It will also interact with impurities, dopants, and native defects, which can result in passivation of the defect. H also bonds to the dangling bonds at extended defects. H^- is mobile at room temperature in p-type Ge [6.56]. In Si, H forms the complex $n\mathrm{H}-\mathrm{V_{Si}}$ where $n = 1, 2, 3, 4$ with $\mathrm{H}-\mathrm{V_{Si}}$ being mobile at $\approx 100\,^\circ\mathrm{C}$ [6.57]. The presence of H in Si can enhance the diffusivity of O in Si by two to three orders of magnitude [6.58]. *Mathiot* [6.59] has modeled H diffusion in terms of the simultaneous diffusion of H^-, H^+, and H^0 with the formation of immobile neutral complexes. In polycrystalline Si, the grain boundaries retard H diffusion, so H diffuses faster in the surrounding lattice than in the grain boundary. H is an ever present and important donor impurity in ZnO: interstitial H and H on O sites provide shallow donor states [6.60]. *Fan* et al. [6.61] found H concentrations at the 2% level in as grown ZnO. In $Hg_{0.75}Cd_{0.25}Te$, $n\mathrm{H}-\mathrm{V_{Hg}}$ complexes occur where $n = 1, 2, 3, 4$. The complexes $3\mathrm{H}-\mathrm{V_{Hg}}$ and $4\mathrm{H}-\mathrm{V_{Hg}}$ convert p-type $Hg_{0.75}Cd_{0.25}Te$ to n-type [6.62].

6.7 Diffusion in Group IV Semiconductors

The materials of interest are Ge, Si, Si/Ge alloys, and SiC. Diffusants divide into one of five categories: self-, other group IV's, slow diffusers (typically dopants from Groups III and V), intermediate diffusers, and fast diffusers. A particular feature is that self-diffusion is always slower than the diffusion of other diffusants. With the exception of SiC, which has the zinc blende structure, as well as numerous polytypes (the simplest of which is the wurtzite, 2H–SiC, form), the other members of this group have the diamond lattice structure.

6.7.1 Germanium

The dominant native defect in Ge is the vacancy and self-diffusion is accounted for primarily as due to the singly ionized vacancy acceptor, V_{Ge}^{-}, with a smaller contribution from the neutral vacancy, V_{Ge}^{0} [6.63]. *Werner* et al. [6.64] measured the self-diffusivity D(Ge) between 535 and 904 °C in intrinsic Ge and obtained $D(\mathrm{Ge}) = 13.6\exp(-3.09\,\mathrm{eV}/k_\mathrm{B}T)\mathrm{cm}^2/\mathrm{s}$. Relative to its intrinsic value, D(Ge) is reduced by acceptor doping and increased by donor doping, as would be expected from the dependence on V_{Ge}^{-} [6.65]. The diffusivities of donor dopants (P, As, Sb) are very similar to each other, as are the diffusivities for acceptor dopants (Al, In, Ga). The acceptor group diffusivities, however, are very close in magnitude to the intrinsic D(Ge), whereas those for the donor group are 10^2 to 10^3 times larger. *Bracht* [6.65] has given a detailed account of self- and dopant diffusion in Ge. Li is a fast interstitial diffuser with a diffusivity exceeding the donor group diffusivities by factors of 10^7 to 10^5 between 600 and 900 °C, whereas Cu [6.66] and Au [6.67] are intermediate (dissociative) diffusers. Evidence of the effects of the internal electric field is also available [6.60].

6.7.2 Silicon

Si stands alone due to the intensive investigations that have been lavished on it over many decades. In the early days, diffusion data yielded many perplexing features. Today the broad aspects are understood along with considerable detail, depending on the topic. Diffusion in Si covers many more topics than arise in any other semiconductor and it is still a very active area in R & D. It is now recognized that, apart from foreign purely interstitial species, self-interstitials, Si_i, and vacancies, V_{Si}, are involved in all diffusion phenomena. So far the best self-diffusion results obtained in intrinsic Si are $D_0 = 530\,\mathrm{cm}^2/\mathrm{s}$ and $Q = 4.75\,\mathrm{eV}$ in the temperature range 855−1388 °C [6.66]. Two distinct facets of self- and dopant diffusion in Si are: (a) the diffusivity has two or three components, each with differing charge states; (b) the diffusivity reflects contributions from both Si_i, and V_{Si} [6.6, 63, 67–70]. Thus, the Si self-diffusivity is determined by Si_i and V_{Si} mechanisms and by three separate defect charge states: neutral (0), positive (+), and negative (−). For the common dopants (B, P, As, and Sb), B and P diffuse primarily via the Si_i defect, As diffuses via both Si_i, and V_{Si} defects, whereas Sb diffuses primarily via V_{Si} defects. Two defect charge states are involved for B $(0, 1^+)$, As $(0, 1-)$ and Sb $(0, 1^-)$, and three for P $(0, 1^-, 2^-)$. The situation for Al [6.71], Ga [6.13], and In [6.71] has Si_i dominant for Al and In diffusion whereas both Si_i, and V_{Si} are involved for Ga. The associated charge states are Al $(0, 1^+)$, Ga $(0, 1^+)$ and In (0). The diffusivities of the Group V dopants (P, As, Sb) lie close to each other and are up to a factor ≈ 10 greater than the intrinsic self-diffusivity. The acceptor dopants (B, Al, Ga, In) also form a group with diffusivities that are up to a factor $\approx 10^2$ greater than the donor dopants. A relatively recently observed interesting feature is that the diffusivities of B and P in intrinsic Si depend on the length of the anneal time, showing an initial change until reaching a final value [6.72]. This time effect is attributed to the time needed for equilibration of the Si_i, and V_{Si} concentrations at the anneal temperature.

The data presented by *Tan* and *Gösele* [6.71] show that Au, Pt, and Zn are intermediate (kick-out) diffusers and that H, Li, Cu, Ni, and Fe are fast interstitial diffusers. Recent evidence shows that Ir diffusion occurs by both kick-out and dissociative mechanisms [6.73]. To provide some perspective: at 1000 °C the diffusivity of H is $\approx 10^{-4}\,\mathrm{cm}^2/\mathrm{s}$ compared to a self-diffusivity of $8\times10^{-17}\,\mathrm{cm}^2/\mathrm{s}$. C and O are important impurities because they occur in high concentrations and can affect the electrical properties through forming complexes with other ionized defects. Although O also occupies interstitial sites and diffuses interstitially it should be classed as an intermediate diffuser because a diffusion jump entails the breaking of two Si-O bonds [6.74]. O can be present at concentrations of $\approx 10^{18}\,\mathrm{cm}^{-3}$ with a diffusivity of $0.13\exp(-2.53\,\mathrm{eV}/k_\mathrm{B}T)\mathrm{cm}^2/\mathrm{s}$. C has a diffusivity that is a little larger than those of Group III dopants: its mechanism is unresolved between the kick-out mode or a diffusing complex comprising a Si_i and C_{Si}.

6.7.3 $Si_{1-x}Ge_x$ Alloys

Si and Ge form a continuous range of alloys in which there is a random distribution of either element coupled with a continuous bandgap range. The alloys have

attracted considerable interest from a device perspective and are commonly prepared as epilayers on Si substrates so that the epilayer will generally be in a strained state. Diffusivity data are sparse and, in the case of dopants, is limited to B, P, and Sb. One might expect that the diffusivity $D(Z)$ of dopant Z would increase continuously as x goes from 0 to 1 at any temperature below the melting point of Ge. However, in the case of B, D(B) varies little for $x \lesssim 0.4$; even so, D(B) increases by a factor $\approx 10^3$ from $\approx 10^{-15}\,\mathrm{cm}^2/\mathrm{s}$ in traversing the composition range at 900 °C [6.75, 76]. D(P) increases by a factor of ≈ 4 for x between 0 and 0.24, only to show a decrease at $x = 0.40$ [6.77]. Limited data suggest that D(Sb) rises continuously across the composition range, increasing by a factor $\approx 10^6$ at 900 °C [6.77]. Surface oxidation enhances D(B) and D(P), indicating that the diffusivities are dominated by a self-interstitial mechansim, whereas D(Sb) is reduced by surface oxidation, pointing to a vacancy mechanism. Compressive strain retards D(B) whereas tensile strain gives a marginal enhancement [6.75]. Compressive strain enhances both D(P) and D(Sb) [6.78]. In relaxed samples D(As) increases from $x = 0$ to $x = 1$ and for $x < 0.35$ both interstitial and vacancy mechanisms are involved but only the vacancy mechanism for $x > 0.35$ [6.78]. Overall, some disagreement exists between different workers about the behavior of $D(Z)$, which may well stem from difficulties with characterizing the experimental conditions. Ge and Si self-diffusivities have been obtained [6.79]. Compositional interdiffusion has been characterized in SiGe heterostructures with x between 0.36 and 0.75 [6.80].

6.7.4 Silicon Carbide

Its large bandgap, high temperature stability (SiC sublimes at ≈ 2730 °C), and high dielectric breakdown strength make SiC a suitable material for devices intended for operation at high temperatures and high powers. It also has potential optoelectronic applications. Material characterization is complicated, as SiC occurs in a range of polytypes (different stacking sequences of closed packed layers). Common polytypes are the cubic zinc blende phase 3C–SiC and the hexagonal phases 2H–SiC (wurtzite), 4H–SiC, and 6H–SiC. The refractory nature, polytypism, and variations in stoichiometry introduce difficulties in the measurement of diffusivities. Typical diffusion anneal temperatures for acceptor (B, Al, Ga) and donor (N, P) dopants are in the range 1800–2100 °C. Significant reductions in the diffusion temperature can be achieved through using a thin surface layer of the oxide of the dopant together with annealing in an oxidizing atmosphere [6.81]. Ab initio calculations for single vacancies and antisites in 4H–SiC [6.82] found the antisites Si and C to be both neutral and therefore generally both electrically and optically inactive. The C vacancy is amphoteric with charge states ranging from 2^+ to 2^-. The Si vacancy is also amphoteric with charge states ranging from 1^+ to 3^-. Similar calculations for 3C–SiC predict divalent a donor for both Si and C interstitials [6.83]. *Bockstedte* et al. [6.84] have calculated, using ab initio methods, the activation energies, Q, for self-diffusion in 3C–SiC by vacancies and interstitials. Generally Q is smaller for self-interstitials but the defect charge state is also an important factor. The Si vacancy is predicted to be metastable, readily transforming to the stable complex V_c–C_{Si}: the complex Si_c–V_{Si} is unstable, reverting to V_c.

The Si and C self-diffusivities, D(Si) and D(C), respectively, were measured between 1850 and 2300 °C by *Hong* et al. ([6.85] and references therein) in both 3C–SiC and 6H–SiC. The ratio $D(\mathrm{C})/D(\mathrm{Si}) \approx 650$ in 3H–SiC and ≈ 130 in 6H–SiC. N doping increased D(Si) and marginally reduced D(C). This behavior suggests that native acceptors are important for Si self-diffusion and that native donors are only marginally involved in determining D(C). Of particular interest is that, between the two polytypes, the self-diffusivities in 6H–SiC exceeded those in 3H–SiC by less than a factor ≈ 3. This suggests that diffusivities are insensitive to the particular polytype. More recent measurements of D(C), between 2100 and 2350 °C, in 4H–SiC found diffusivities that were $\approx 10^5$ times smaller than the earlier results for 3H–SiC and 6H–SiC, mainly due to differences in D_0 [6.83]. There is currently no explanation for these huge differences and the question of the reliability of the self-diffusivity data must be considered, particularly with respect to residual impurity levels.

Earlier work by *Vodakov* et al. [6.86] found that the diffusivity of B in six different polytypes of SiC, excluding 3C–SiC, varied by ≤ 30%, not only for diffusion along the c-axis but also perpendicular to it. The diffusivities of some common dopants have been summarized by *Vodakov* and *Mokhov* [6.87]. Diffusion mechanisms for B in 4H– and 6H–SiC have been discussed by *Usov* et al. [6.88]. A recent finding is that an SiO_2 layer on the surface of 6H–SiC greatly enhances B diffusion [6.89], yielding a diffusivity of $\approx 6 \times 10^{-16}\,\mathrm{cm}^2/\mathrm{s}$ at 900 °C. This compares to a temperature of ≈ 1400 °C (extrapolated) for the same diffusivity without an SiO_2 surface layer. Electric fields of $\approx 10^6\,\mathrm{V/cm}$ have been found in 4H/3C/4H–SiC quantum wells due to the spontaneous polarization in the 4H–SiC matrix [6.90].

6.8 Diffusion in III–V Compounds

The III–V binary compounds are formed between the cations B, Al, Ga, In and the anions N, P, As, and Sb. Mutual solubility gives rise to the ternaries such as $Al_{1-x}Ga_xAs$, and to the quaternaries, such as $Al_{1-x}Ga_xAs_{1-y}P_y$. The B compounds offer little more than academic interest, whereas the rest of the III–V family are important materials in both electronic and optoelectronic devices. The nitrides all have the wurtzite structure, with the remaining compounds possessing the zinc blende structure. In view of the wide range of binaries, ternaries, and so on, it is not surprising that diffusivity measurements have focussed mainly on those compounds relevant to devices: essentially GaAs and GaAs-based materials. An important characteristic of these compounds is the high vapor pressure of the anion components; it is the variations in a particular vapor pressure that give rise to significant changes in levels of nonstoichiometry. This means that a proper characterization, at a given temperature, of any diffusivity must specify the doping level and the ambient anion vapor pressure during the diffusion anneal: the latter determines the native defect concentrations in intrinsic samples, and both factors have equal importance in controlling the concentrations under extrinsic conditions. On both the anion and the cation sublattices, the possible native point defects are the vacancy, the self-interstitial and the antisite and all can occur in one or more charge states. Complexes formed between native defects and dopants (inclusive of residual impurities) have also to be considered.

6.8.1 Self-Diffusion

Self-diffusivity data are limited to the Ga and In compounds [6.60, 89], and even here systematic measurements are restricted to GaAs [6.71, 91] and GaSb ([6.91] and references therein). For GaAs, early evidence (based largely on CID in AlGaAs structures), concluded that the Ga self-diffusivity D(Ga) was determined by the triply ionized Ga vacancy V_{Ga}^{3-} and doubly ionized Ga interstitial Ga_i^{2+}. More recent and direct measurements of D(Ga) in Ga isotope structures identified the three vacancy charge states V_{Ga}^{2-}, V_{Ga}^{1-} and V_{Ga}^{0} as being responsible for D(Ga) in intrinsic and lightly doped GaAs; the possibility remains that V_{Ga}^{3-} and Ga_i^{2+} could dominate at high doping levels. Between 800 and 1200 °C the Arrhenius parameters for D(Ga) are $D_0 = 0.64\,cm^2/s$ and $Q = 3.71\,eV$ in intrinsic GaAs under a partial As_4 vapor pressure of ≈ 1 atm. The situation for As self-diffusion is less clear, but the evidence points to the dominance (in the diffusion process) of the neutral As interstitial over the As vacancy (the supposedly dominant native defect, the As antisite, is not involved). Data have been obtained for both Ga and Sb self-diffusion in intrinsic GaSb under Ga and Sb rich conditions. There is a conflict between the results obtained with bulk material and those from heterostructures (see [6.91] and references therein). *Shaw* [6.91] concluded that the defects involved in Ga self-diffusion were the Frenkel pair Ga_iV_{Ga} and V_{Ga} even though the Ga antisite Ga appears to be the dominant native defect. Two parallel mechanisms were also identified for Sb self-diffusion, namely one due to the defect pair Sb_iV_{Ga} and the second due to either to the mixed vacancy pair $V_{Ga}V_{Sb}$ or to the triple defect $V_{Ga}Ga_{Sb}V_{Ga}$. Reliable results for D(Ga) in intrinsic GaP under a partial vapor pressure P_4 of ≈ 1 atm are also available [6.92]: between 1000 and 1190 °C the Arrhenius parameters for D(Ga) are $D_0 = 2.0\,cm^2/s$ and $Q = 4.5\,eV$. Data on the effects of doping and changes in levels of nonstoichiometry are lacking. In GaN the self-diffusivity of N has $D_0 = 1.6\times10^2\,cm^2/s$ and $Q = 4.1\,eV$ [6.93]. This value for Q is in good agreement with the 4.3 eV obtained from first principles calculations for the diffusivity of V_N^{1+} [6.94]. Ab initio calculations show that the N vacancy is a major defect in GaN with a lower formation energy than that for the Ga vacancy [6.95]. The Ga self-interstitial in GaN is reported to be mobile at room temperature [6.96].

6.8.2 Dopant Diffusion

Most of the data on dopant diffusion in the III–Vs refer to GaAs [6.50], notably for Be [6.97], Cd [6.93], C, Si, S, Zn, and Cr [6.70]. The singly ionized acceptors Be, Zn, and Cd (which occupy Ga sites) and the singly ionized donors C and S (which occupy As sites) all diffuse by the kick-out mechanism. The native interstitials involved are Ga_i^{2+} and As_i^0, apart from Be where the data are best accounted for in terms of the singly ionized interstitial Ga_i^{1+}. Si is an amphoteric dopant and at low concentrations it predominantly occupies Ga sites as a singly ionized donor Si_{Ga}^{1+}. At high concentrations compensation starts to occur due to increasing occupancy as a singly ionized acceptor on As sites. At low concentrations Si_{Ga}^{1+} diffusion is attributed to a vacancy mechanism (V_{Ga}^{3-}). Cr sits on Ga sites and is a deep-level acceptor dopant, important in the growth of high resistivity GaAs. Depending on circumstances, it can diffuse by either the kick-out or the Frank–Turnbull mechanism. The creation of extended defects in the diffusion zone by Zn in-diffusion is a well established feature. The same feature has been found by *Pöpping* et al. [6.98] for Zn in-diffusion in GaP. They

further concluded that Zn diffuses via the kick-out process in GaP through the involvement of either Ga_i^{1+} or Ga_i^{2+}. The diffusivity of Mn in GaAs has been measured between 680 and 825 °C in an Ar ambient [6.99] and *Bosi* et al. [6.100] have investigated Ge diffusion in Ge/GaAs epilayers. The diffusion behavior of vanadium in GaN between 1000 and 1100 °C has also been reported [6.101]. H has a significant and important impact in GaN due to its passivation of acceptor dopants and data on its drift, diffusion, and trapping are available [6.100, 101].

6.8.3 Compositional Interdiffusion

The III–V binaries, ternaries, and quaternaries are the bases for the fabrication of numerous quantum well and superlattice structures. The general situation in which the cation and anion sublattices in each layer can contain up to four different components with concentrations ranging from 0 to 100% presents an impossibly complex problem for characterizing diffusion behavior with any rigor. The role of strain and internal electric fields must also be considered as parameters. As a consequence, CID studies have been limited to simpler structures, primarily GaAs–AlAs and GaAs–AlGaAs, with interpretations in terms of known diffusion features in GaAs [6.69]. Doping is an important ingredient of these multilayer structures and it was soon discovered that the acceptors Be, C, Mg, Zn and the donors Si, Sn, S, Se, and Te could all cause complete disorder of the structure through enhancement of the CID process on either or both sublattices [6.70, 102]. An interesting exception, however, is found in GaAs–GaAsSb, where either Si or Be reduce CID. Two generally accepted reasons for these dopant effects are: (i) the Fermi-level effect in which the dopant (acceptor/donor) concentration is high enough to make the semiconductor extrinsic so that the concentrations of native (donor/acceptor) defects are increased; (ii) if the dopant diffuses by the kick-out mechanism then in-diffusion will generate a local excess of the native self-interstitial. Clearly (i) operates for dopants incorporated during growth or by subsequent in-diffusion, whereas (ii) is restricted to in-diffusion. Either way the increase in the local native defect concentration(s) leads to a direct enhancement of CID. In the case of GaAs–GaAsSb, cited earlier, Si will also decrease the concentrations of native donors such as anion vacancies, which would have a direct impact on the reduction of CID on the anion sublattice. On the other hand, Be should increase native donor concentrations and therefore give enhanced CID of the anions, contrary to observation. Overall, the general features of the dopant-induced disordering process seem to be understood but problems still remain. *Harrison* [6.103] has commented on the approximations commonly made when extracting quantitative information from CID data. The demands of III–V device technology present increasing complexity when attempting to understand the physical processes involved, so that recourse to empirical recipes is sometimes needed. This is illustrated by structures comprising GaInNAs quantum wells with GaAs barriers, all enclosed within AlAs outer layers, whose optoelectronic properties can be improved by the judicious choice of time/temperature anneals [6.104].

6.9 Diffusion in II–VI Compounds

Interest in II–VI materials predates that in the III–Vs because of their luminescence properties in the visible spectrum which, based on powder technology, resulted in the application of the bigger bandgap materials (such as ZnS) as phosphors in luminescent screens. The development of crystal growth techniques extended interest in the optoelectronic properties of the wider family of II–VI binary compounds formed between the group II anions Zn, Cd, and Hg and the group VI cations O, S, Se, and Te. As with III–Vs, ternary and quaternary compounds are readily formed. The ternary range of compositions $Hg_{1-x}Cd_xTe$ has proved to be a particularly important family member due to their unique properties and consequent extensive exploitation in infrared systems. ZnS, ZnO, CdS, and CdSe crystallize in the wurtzite structure whereas the remaining binaries have the zinc blende structure. ZnS, CdS, and CdSe can also occur in the zinc blende phase. The Hg compounds all have the zinc blende structure. The native point defects that can occur are similar to the III–Vs; namely, vacancies, self-interstitials, and antisites for the cation and anion sublattices. Recent interest has expanded to include the cation dopants Be, Mg, and Mn, usually in ternary or quaternary systems. A distinctive feature of atomic diffusion in the II–VI compounds is their much higher diffusivities relative to those in Group IV and III–V semiconductors. The relative ease of measurement has ensured that much more self- and dopant data are available compared to the III–Vs. A further difference is that both cation and anion vapor pressures are significant compared to the III–Vs, where the cation vapor pressures are negligible. As a result the level of nonstoichiometry can be varied between cation rich to anion rich. Excess concentrations are typically up to

$10^{19}\,cm^{-3}$. This has the consequence that the solubility of dopants can be controlled by using the partial vapor pressure of one of the components. Unless otherwise stated, the material in the following sections is drawn from the reviews by *Shaw* [6.9, 105–107] and by *Capper* et al. [6.108]. Within the last decade or so much interest has developed in the optoelectronic properties of ZnO and accounts of defects [6.106] and diffusion [6.107] in ZnO are available.

6.9.1 Self-Diffusion

Where the anion self-diffusivity D_A has been measured as a function of the ambient anion or cation partial vapor pressure in undoped material (ZnSe, CdS, CdSe, CdTe, and $Hg_{0.8}Cd_{0.2}Te$), a consistent pattern of behavior has emerged: in traversing the composition range from anion-rich to cation-rich, D_A is inversely proportional to the rising cation partial vapor pressure, P_C, until close to cation saturation, when D_A starts to increase with P_C. Strong donor doping in anion-rich CdS and CdSe had no effect on D_A. This evidence points to either a neutral anion interstitial or a neutral anion anti-site/anion vacancy complex as the diffusion mechanism over most of the composition range, changing to an anion vacancy mechanism as the cation-rich limit is approached.

The situation for cation self-diffusion proves to be more complicated due to different patterns of behavior of the cation self-diffusivity, D_C, across the compounds. In undoped ZnSe, ZnTe, CdTe, and $Hg_{0.8}Cd_{0.2}Te$ (above $\approx 300\,°C$), D_C is largely independent of P_C across the composition range. Such an independence excludes native point defect mechanisms and (excluding nondefect mechanisms) indicates self-diffusion via neutral complexes such as a cation interstitial/cation vacancy pair or a cation vacancy/anion vacancy pair. Donor or acceptor doping increases D_C, pointing to the involvement of ionized native defects or complexes. Cd self-diffusion in Cu doped CdTe between 500 and 900 °C identified the doubly ionized Cd interstitial as the defect involved with a diffusivity of $0.23\exp(-0.99\,eV/k_BT)\,cm^2/s$ [6.109]. The Arrhenius parameters for Zn self-diffusion in undoped ZnSe above 760 °C are $D_0 = 9.8\,cm^2/s$ and $Q = 3.0\,eV$ and those for Hg in undoped $Hg_{0.8}Cd_{0.2}Te$ above 250 °C are $D_0 = 3.8\times10^{-3}\,cm^2/s$ and $Q = 1.22\,eV$. In the case of undoped ZnS, CdS, CdSe, and HgTe D_C generally varies with P_C across the composition range. The simplest variations are found in CdSe and HgTe. In CdSe, D_C can be attributed to the parallel diffusion of singly ionized (1^+) and doubly ionized $(2+)$ Cd self-interstitials. D_C in HgTe initially falls with rising P_C and then increases when crossing the anion-rich to cation-rich boundary, corresponding to diffusion by a singly ionized $(1-)$ Hg vacancy and by a singly ionized $(1+)$ Hg interstitial, respectively. The behavior patterns in ZnS and CdS, however, present substantial problems in their interpretation: donor doping can also enhance D_C, pointing to the participation of an ionized native acceptor mechanism.

6.9.2 Chemical Self-Diffusion

Changes in the electrical conductivity or conductivity type by step changes in P_C in sample anneals have been used to characterize the change in level of nonstoichiometry through the chemical self-diffusivity, D_Δ, in CdS, CdTe, and $Hg_{0.8}Cd_{0.2}Te$. D_Δ obviously describes the diffusion of one or more ionized native defects, but in itself does not identify the defect(s).

In CdS and CdTe, D_Δ is attributed to the singly ionized (1^+) and/or doubly ionized $(2+)$ Cd interstitial; in CdTe, depending on the temperature, D_Δ exceeds D_C by a factor 10^5 to 10^6. Modeling based on the simultaneous in-diffusion and out-diffusion of doubly ionized cation interstitials $(2+)$ and vacancies $(2-)$ gives a satisfactory quantitative account of type conversion (p → n) in $Hg_{0.8}Cd_{0.2}Te$ [6.10].

6.9.3 Dopant Diffusion

Although much information is available on dopant diffusion, it is mainly empirical and it is not uncommon for a dopant diffusivity to be independent of the dopant concentration (as revealed by an erfc profile, which indicates a constant diffusivity for a given diffusion profile) under one set of conditions only to give profiles which cannot be characterized by single diffusivities when the conditions are changed. Equally, the variation of dopant diffusivity with P_C may differ at different temperatures. A further difficulty when attempting to identify a diffusion mechanism is that the local electroneutrality condition is usually not known with any certainty due to significant concentrations of various ionized native defects. A good example of the problems encountered is provided by the diffusion of In in $Hg_{0.8}Cd_{0.2}Te$, where diffusion of the singly ionized (1^-) pair $In_{Hg}V_{Hg}$ can account for some of the diffusion features. Some dopants, however, can present clear-cut diffusion properties which permit a well-defined interpretation. The diffusion of As in $Hg_{0.8}Cd_{0.2}Te$ is one such case [6.110]. All of the observed features of D(As) are accounted for on the basis that: (i) As occupies cation and anion sites as singly ionized donors $(1+)$ and acceptors $(1-)$, respectively; (ii) only the ionized donor is mobile and diffuses by a vacancy mechanism on the cation sublattice; (iii) the diffusion sample is electri-

cally intrinsic throughout, so the As concentration is always less than the intrinsic free carrier concentration.

6.9.4 Compositional Interdiffusion

Empirical information, based on bulk material, exists for CID in the following ternaries: (ZnCdHg)Te, (ZnCd)Se, (ZnCdHg)SSe, CdSeTe, ZnCdS, HgCdTe, and CdMnTe. It might be expected that features evident in the binaries, such as donor doping enhancing the cation diffusivity but having no effect on that of the anion, and the anion diffusivity increasing (decreasing) with anion (cation) partial vapor pressure across most of the composition range, would continue to be seen. This means that in a ternary or quaternary system, donor doping will enhance CID on the cation sublattice, but not on the anion sublattice, and annealing under a high (low) anion (cation) vapor pressure will enhance CID on the anion sublattice. This effect of the anion vapor pressure has been confirmed in CdSSe and CdSeTe and more recently in ZnSSe/ZnSe superlattices [6.111]. In (donor) doping has also been found to enhance the CID of cations in CdMnTe [6.112], as has N (acceptor) doping in ZnMgSSe/ZnSSe superlattices [6.113, 114]. The consequences of doping on CID in the II–VIs are obviously very similar to the III–V situation.

6.10 Nano Volume Diffusion

In view of the importance of nanoscale materials and devices it is appropriate to consider if Fickian diffusion (6.1) occurs in nanoscale systems. An immediate problem is evident in (6.1) which shows that a plane of constant c has a shift velocity of $\mathrm{d}x/\mathrm{d}t \propto 1/\sqrt{t}$ which $\to \infty$ as $t \to$ zero. For $0.3 \leq x \leq \approx 1.6\,\mathrm{nm}$ *Balogh* et al. [6.115] found $x \propto t^{0.7}$ for Si diffusion into amorphous Ge, changing to $x \propto \sqrt{t}$ at larger x. The continuum approach clearly fails as $t \to$ zero and recourse to discrete and atomistic models becomes necessary for describing diffusion on the nanoscale. Whereas in Fickian diffusion the parameters can be measured the parameters for discrete models can only be estimated.

6.11 Diffusion in Molten Semiconductors

So far, this chapter has been concerned with atom diffusion in crystalline semiconductors but diffusion in the liquid state must also be recognized and is termed thermodiffusion. Examples are provided by doping during crystal growth in which the dopant has a greater solubility in the solid than in the liquid, so that there is a diffusion flux of the dopant to the growth interface: if the solubility is less, then there will be a flux away from the interface. In the case of growth of the binary MX, away from the eutectic, if the melt is M or X-rich there will be flux of M or X away from the interface. Temperature gradients will also promote thermodiffusion. However, the existence of convection currents in the melt can replace any thermodiffusion that may be present. Convection can be avoided in a microgravity environment, or by employing a vertical temperature gradient in which the temperature increases with height (as in the temperature gradient zone melting technique), or in nanoscale processing. Thermodiffusion is an essential feature in the fabrication of micro and nano metal doped semiconductor devices [6.116]. The redistribution of a dopant ion implant in Si following pulsed laser annealing has been interpreted in terms of thermodiffusion [6.115]. Thermodiffusion has also been invoked to account for the mixing at an Fe–Si interface resulting from exposure to a 230 MeV Au beam [6.117]. A diffusivity of $10^{-4}\,\mathrm{cm}^2/\mathrm{s}$ for As in liquid Si has been reported arising from As ion implants [6.118–120].

6.12 The Meyer–Neldel Rule

The Meyer–Neldel rule is found to apply to most phenomena that can be characterized by an Arrhenius expression. In the case of a diffusion coefficient $D = D_0 \exp(-Q/k_\mathrm{B}T)$, the rule states that a correlation exists between D_0 and Q. A phenomenological basis for the model has been given by *Dyre* [6.121] and the rule has been examined for diffusivities in Si and Ge by *Khait* et al. [6.122]. In $Hg_{0.8}Cd_{0.2}Te$ a good fit to $\ln D_0 = 15\,Q - 23.7$ is found for twelve self- and impurity diffusivities.

6.13 Conclusions

The first step in a diffusion investigation is to collect empirical data, which then leads to the second step where experiments can be designed to study the effects of the Fermi level (through the background doping level), of the ambient atmosphere (such as oxidizing, inert or vapor pressure of a system component) and of the sample structure (such as in a quantum well, superlattice or MBE layer). The third step is to identify the diffusion mechanism and the associated defects using the experimental results in conjunction with the results from first-principles calculations of defect formation energies and their activation energies for diffusion.

It must also be recognized that Fickian diffusion is not applicable at the nanoscale [6.113].

Clarification of the active processes involved can be gained by numerical modeling (see *Noda* [6.123]). These data then provide the basis for the development of process simulators and defect engineering in which the concentrations and spatial distributions of host atoms, dopants and defects are organized according to requirements. Most progress towards achieving this ideal scenario has been made in Si, GaAs and $Hg_{0.8}Cd_{0.2}Te$. The reality elsewhere is that the boundaries between steps are blurred, with the third step often being taken with inadequate experimental information e.g., as in GaN and ZnO. Much work remains to be done in order to achieve a full understanding of diffusion processes in semiconductors but valuable insights are being provided from first-principles calculations of native defect properties: e.g., the diffusion of O in ZnO [6.121].

6.14 General Reading and References

General background material for diffusion in semiconductors can be found in [6.8, 11, 14, 50]. More specific accounts are given in [6.12, 122] for Si and in [6.6] for Si and Ge. Si, Ge, and GaAs are covered by [6.69]. The III–Vs and II–VIs are dealt with in [6.50] and [6.103] respectively. A review of nanoscale volume diffusion is available in [6.123]. The volumes in the EMIS Datareviews Series (IEE, Stevenage, UK) cover all the important semiconductors. The series *Defects and Diffusion in Semiconductors* ed. by D. J. Fisher (Trans Tech. Publications, Brandrain 6, Switzerland) offers an annual and selective retrospective of recent literature.

References

6.1 M. Knez: Semicond. Sci. Technol. **27**, 074001 (2012)
6.2 P. Alen, M. Vehkamaki, M. Ritali, M. Leskela: J. Electrochem. Soc. **153**, G304 (2006)
6.3 J. Hong, S. Lee, S. Lee, H. Han, C. Mahata, H.-W. Yeon, B. Koo, S.-I. Kim, T. Nam, K. Byun, B.-W. Min, Y.-W. Kim, H. Kim, Y.-C. Joo, T. Lee: Nanoscale **6**, 7503 (2014)
6.4 J. Crank: *The Mathematics of Diffusion*, 2nd edn. (Clarendon Press, Oxford 1979)
6.5 M.E. Glicksman: *Diffusion in Solids* (Wiley, New York 2000)
6.6 A. Ural, P.B. Griffin, J.D. Plummer: J. Appl. Phys. **85**, 6440 (1999)
6.7 I.B. Belova, D. Shaw, G.E. Murch: J. Appl. Phys. **106**, 113707 (2009)
6.8 W. Frank, U. Gösele, H. Mehrer, A. Seeger: Diffusion in silicon and germanium. In: *Diffusion in Crystalline Solids*, ed. by G.E. Murch, A.S. Nowick (Academic, Orlando 1984) p. 63
6.9 D. Mathiot, J.C. Pfister: J. Appl. Phys. **66**, 970 (1989)
6.10 D. Shaw: General features of diffusion in semiconductors. In: *Atomic Diffusion in Semiconductors*, ed. by D. Shaw (Plenum, London 1973) p. 1
6.11 D. Shaw: Self- and impurity diffusion processes in widegap II–VI materials. In: *Widegap II–VI Compounds for Opto-electronic Applications*, ed. by H.E. Ruda (Chapman and Hall, London 1992) p. 244
6.12 D. Shaw, P. Capper: J. Mater. Sci. Mater. El **11**, 169 (2000)
6.13 T.Y. Tan, U. Gösele, S. Yu: Crit. Rev. Sol. St. Mater. Sci. **17**, 47 (1991)
6.14 R.B. Fair: Concentration profiles of diffused dopants in silicon. In: *Impurity Doping Processes in Silicon*, ed. by F.F. Wang (North-Holland, Amsterdam 1981) p. 349
6.15 S.M. Hu: J. Appl. Phys. **70**, R53 (1991)
6.16 G.B. Abdullaev, T.D. Dzhafarov: *Atomic Diffusion in Semiconductor Structures* (Harwood, Chur 1987)
6.17 M. Laudon, N.N. Carlson, M.P. Masquelier, M.S. Daw, W. Windl: Appl. Phys. Lett. **78**, 201 (2001)
6.18 K. Rajendran, W. Schoenmaker: J. Appl. Phys. **89**, 980 (2001)
6.19 H. Takeuchi, P. Ranada, V. Subramanian, T.-J. King: Appl. Phys. Lett. **80**, 3706 (2002)
6.20 S.C. Jain, W. Schoenmaker, R. Lindsay, P.A. Stolk, S. Decoutere, M. Willander, H.E. Maes: J. Appl. Phys.

91, 8919 (2002)
6.21 L. Shao, J. Chen, J. Zhang, D. Tang, S. Patel, J. Liu, X. Wang, W.-K. Chu: J. Appl. Phys. **96**, 919 (2004)
6.22 Y.M. Haddara, J.C. Bravman: Ann. Rev. Mater. Sci. **28**, 185 (1998)
6.23 I. Lyubomirsky, V. Lyahovitskaya, D. Cahen: Appl. Phys. Lett. **70**, 613 (1997)
6.24 C.H. Chen, U. Gösele, T.Y. Tan: Appl. Phys. A **68**, 9, 19, 313 (1999)
6.25 P.N. Grillot, S.A. Stockman, J.W. Huang, H. Bracht, Y.L. Chang: J. Appl. Phys. **91**, 4891 (2002)
6.26 E. Chason, S.T. Picraux, J.M. Poate, J.O. Borland, M.I. Current, T. Diaz de la Rubia, D.J. Eaglesham, O.W. Holland, M.E. Law, C.W. Magee, J.W. Mayer, J. Melngailis, A.F. Tasch: J. Appl. Phys. **81**, 6513 (1997)
6.27 J.L. Melendez, J. Tregilgas, J. Dodge, C.R. Helms: J. Electron Mater. **24**, 1219 (1995)
6.28 F.S. Ham: J. Phys. Chem. Solids **6**, 335 (1958)
6.29 F.S. Ham: J. Appl. Phys. **30**, 915 (1959)
6.30 F.S. Ham: J. Appl. Phys. **30**, 1518 (1959)
6.31 S.C. Jain, A.E. Hughes: Proc. R. Soc. Lond. A. **360**, 47 (1978)
6.32 D. Peak, J.W. Corbett: Radiation Effects **36**, 197 (1978)
6.33 D. Shaw: Phys. Stat. Sol. A **60**, 251 (1980)
6.34 A. Borghesi, B. Pivac, A. Sassella, A. Stella: J. Appl. Phys. **77**, 4169 (1995)
6.35 K.F. Kelton, R. Falster, D. Gambaro, M. Olmo, M. Cornaro, P.F. Wei: J. Appl. Phys. **85**, 8097 (1999)
6.36 S. Solmi, E. Landi, F. Baruffaldi: J. Appl. Phys. **68**, 3250 (1990)
6.37 S. Solmi, D. Nobili: J. Appl. Phys. **83**, 2484 (1998)
6.38 B. Columbeau, N.E.B. Cowern: Semicond. Sci. Technol. **19**, 1339 (2004)
6.39 S. Mirabella, E. Bruno, F. Priolo, D. Salvador, E. Napolitani, A.V. Drigo, A. Carnera: Appl. Phys. Lett. **83**, 680 (2003)
6.40 C.J. Ortiz, P. Pilcher, T. Fhner, F. Cristiano, B. Columbeau, N.E.B. Cowern, A. Claverie: J. Appl. Phys. **96**, 4866 (2004)
6.41 H. Puchner, S. Selberherr: IEEE Trans. Electron. Dev. **42**, 1750 (1995)
6.42 C. Poisson, A. Rolland, J. Bernardini, N.A. Stolwijk: J. Appl. Phys. **80**, 6179 (1996)
6.43 D. Shaw: Semicond. Sci. Technol. **12**, 1079 (1997)
6.44 I. Kaur, Y. Mishin, W. Gust: *Fundamentals of Grain and Interphase Boundary Diffusion* (Wiley, Chichester 1995)
6.45 G.J. Phelps: Semicond. Sci. Technol. **27**, 035013 (2012)
6.46 S.M. Hu: J. Appl. Phys. **43**, 2015 (1972)
6.47 N.G. Nilsson: Phys. Stat. Sol. A **50**, K43 (1978)
6.48 H. Kroemer: IEEE Trans. Electron Dev. **ED-25**, 850 (1978)
6.49 A.P. Vasov, B.S. Sokolovskii, L.S. Monastyrskii, O.Yu. Bonchyk, A. Barcz: Thin Solid Films **459**, 28 (2004)
6.50 E.A. Caridi, T.Y. Chang, K.W. Goossen, L.F. Eastman: Appl. Phys. Lett. **56**, 659 (1990)
6.51 A. Hangleiter, F. Hitzel, S. Lafman, H. Rossow: Appl. Phys. Lett. **83**, 1169 (2003)
6.52 B. Tuck: *Atomic Diffusion in III-V Semiconductors* (Adam Hilger, Bristol 1988)
6.53 S.J. Rothman: The measurement of tracer diffusion coefficient in solids. In: *Diffusion in Crystalline Solids*, ed. by G.E. Murch, A.S. Nowick (Academic, Orlando 1984) p. 1
6.54 R.M. Fleming, D.B. McWhan, A.C. Gossard, W. Wiegmann, R.A. Logan: J. Appl. Phys. **51**, 357 (1980)
6.55 E. Hüger, R. Kube, H. Bracht, J. Stahn, T. Geue, H. Schmidt: Phys. Stat. Sol. B **249**, 2108 (2012)
6.56 V. Kolkovsky, S. Klemm, M. Allardt, J. Weber: Semicond. Sci. Technol. **28**, 025007 (2013)
6.57 S.K. Estreicher, A. Docaj, M.B. Bebek, D.J. Backlund, M. Stavola: Phys. Stat. Sol. A **209**, 1872 (2012)
6.58 Y.L. Huang, Y. Ma, R. Job, W.R. Fahmer, E. Simeon, C. Claeys: J. Appl. Phys. **98**, 033511 (2005)
6.59 D. Mathiot: Phys. Rev. B **40**, 5867 (1989)
6.60 E.V. Lavrov: Physica B **404**, 5075 (2009)
6.61 J.C. Fan, G.W. Ding, S. Fung, Z. Xie, Y.C. Zhong, K.S. Wong, G. Brauer, W. Anwand, D. Grambole, C.C. Ling: Semicond. Sci. Technol. **25**, 085009 (2010)
6.62 L. Xue, D.H. Tang, X.D. Qu, L.Z. Sun, W. Lu, J.X. Zhong: J. Appl. Phys. **110**, 053704 (2011)
6.63 E. Simoen, C. Claeys: Diffusion and solubility of dopants in germanium. In: *Germanium-Based Technologies from Materials to Devices*, ed. by C. Claeys, E. Simoen (Elsevier, Amsterdam 2007) p. 67
6.64 M. Werner, H. Mehrer, H.D. Hocheimer: Phys. Rev. B **37**, 3930 (1985)
6.65 H. Bracht: Diffus. Fundam. **8**, 1.1 (2008)
6.66 N.A. Stolwijk, W. Frank, J. Hlzl, S.J. Pearton, E.E. Haller: J. Appl. Phys. **57**, 5211 (1985)
6.67 A. Strohm, S. Matics, W. Frank: Diffus. Defect Forum **194–199**, 629 (2001)
6.68 H. Bracht, E.E. Haller, R. Clark-Phelps: Phys. Rev. Lett. **81**, 393 (1998)
6.69 H. Bracht: Nuclear Instrum. Methods Phys. Res. B **253**, 105 (2006)
6.70 A. Ural, P.B. Griffin, J.D. Plummer: Phys. Rev. Lett. **83**, 3453 (1999)
6.71 T.Y. Tan, U. Gösele: Point defects, diffusion, and precipitation. In: *Handbook of Semiconductor Technolgy*, , ed by K. A. Jackson, W. Schrter, Vol. 1, (Wiley-VCH, Weiheim 2000) p. 231
6.72 O. Krause, H. Ryssel, P. Pichler: J. Appl. Phys. **91**, 5645 (2002)
6.73 S. Solmi, A. Parisini, M. Bersani, D. Giubertoni, V. Soncini, G. Carnevale, A. Benvenuti, A. Marmiroli: J. Appl. Phys. **92**, 1361 (2002)
6.74 J.S. Christensen, H.H. Radamson, A.Yu. Kuznetsov, B.G. Svensson: Appl. Phys. Lett. **82**, 2254 (2003)
6.75 L. Lerner, N.A. Stolwijk: Appl. Phys. Lett. **86**, 011901 (2005)
6.76 R.C. Newman: J. Phys. Condens. Mat. **12**, R335 (2000)
6.77 N.R. Zangenberg, J. Fage-Pedersen, J. Lundsgaard-Hansen, A. Nylandsted-Larsen: Defect Diffus. Forum **194-199**, 703 (2001)
6.78 A.D.N. Paine, A.F.W. Willoughby, M. Morooka, J.M. Bonar, P. Phillips, M.G. Dowsett, G. Cooke: Defect Diffus. Forum **143-147**, 1131 (1997)
6.79 J.S. Christensen, H.H. Radamson, A.Yu. Kuznetsov, B.G. Svensson: J. Appl. Phys. **94**, 6533 (2003)
6.80 P. Laitinen, I. Rihimki, J. Räisänen: Phys. Rev. B **68** 155209 (2003)

6.81 A. Strohm, T. Voss, W. Frank, P. Laitinen, J. Räisänen: Z. Metallkunde **93**, 737 (2002)
6.82 Y. Dong, W. Chern, P.M. Mooney, J.L. Hoyt, G. Xia: Semicond. Sci. Technol. **29**, 015012 (2014)
6.83 J.M. Lento, L. Torpo, T.E.M. Staab, R.M. Nieminen: J. Phys. Condens. Matter **16**, 1053 (2004)
6.84 M. Bockstedte, A. Mattausch, O. Pankratov: Phys. Rev. B **68**, 205201 (2003)
6.85 J.D. Hong, R.F. Davis, D.E. Newbury: J. Mater. Sci. **16**, 2485 (1981)
6.86 Y.A. Vodakov, G.A. Lomakina, E.N. Mokhov, V.G. Oding: Sov. Phys. Solid State **19**, 1647 (1977)
6.87 Y.A. Vodakov, E.N. Mokhov: Diffusion and solubility of impurities. In: *Silicon Carbide – 1973*, ed. by R.C. Marshall, J.W. Faust Jr., C.E. Ryan (Univ. South Carolina Press, Columbia 1973) p. 508
6.88 I.O. Usov, A.A. Suvorova, Y.A. Kudriatsev, A.V. Suvorova: J. Appl. Phys. **96**, 4960 (2004)
6.89 N. Bagraev, A. Bouravleuv, A. Gippius, L. Klyachkin, A. Malyarenko: Defect Diffus. Forum **194–199**, 679 (2001)
6.90 S. Bai, R.P. Devaty, W.J. Choyke, U. Kaiser, G. Wagner, M.F. MacMillan: Appl. Phys. Lett. **83**, 3171 (2003)
6.91 D. Shaw: Semicond. Sci. Technol. **18**, 627 (2003)
6.92 L. Wang, J.A. Wolk, L. Hsu, E.E. Haller, J.W. Erickson, M. Cardona, T. Ruf, J.P. Silveira, F. Brione: Appl. Phys. Lett. **70**, 1831 (1997)
6.93 O. Ambacher, F. Freudenberg, R. Dimitrov, H. Angerer, M. Stutzmann: Jpn. J. Appl. Phys. **37**, 2416 (1998)
6.94 S. Limpijumnong, C.G. Van de Walle: Phys. Rev. B **69**, 035207 (2004)
6.95 M.G. Ganchenkova, R.M. Nieminen: Phys. Rev. Lett. **96**, 196402 (2006)
6.96 U. Gerstmann, A.P. Seitsonen, F. Mauri: Phys. Stat. Sol. B **245**, 924 (2008)
6.97 J.C. Hu, M.D. Deal, J.D. Plummer: J. Appl. Phys. **78**, 1595 (1995)
6.98 J. Pöpping, N.A. Stolwijk, G. Bösker, C. Jäger, W. Jäger, U. Södervall: Defect Diffus. Forum **194–199**, 723 (2001)
6.99 O. Koskelo, J. Risnen, F. Tuomisto, J. Sadovski: Semicond. Sci. Technol. **24**, 045011 (2009)
6.100 M. Bosi, G. Attolini, C. Ferrari, C. Frigeri, M. Calicchio, F. Rossi, K. Vad, A. Csik, Z. Zolnai: J. Crystal Growth **318**, 367 (2011)
6.101 A. Bchetnia, C. Saidia, M. Souissi, T. Boufaden, B. El Jani: Semocond. Sci. Technol. **24**, 095020 (2009)
6.102 A.Y. Polyakov, N.B. Smirnov, S.J. Pearton, F. Ren, B. Theys, F. Jomard, Z. Teukam, V.A. Dmitriev, A.E. Nikolaev, A.S. Usikov, I.P. Nikitina: Appl. Phys. Lett. **79**, 1834 (2001)
6.103 I. Harrison: J. Mater. Sci. Mater. Electron. **4**, 1 (1993)
6.104 S. Govindaaraju, J.M. Reifsnider, M.M. Oye, A.L. Holmes: J. Electron. Mater. **32**, 29 (2003)
6.105 D. Shaw: J. Crystal Growth **86**, 778 (1988)
6.106 D. Shaw: J. Electron. Mater. **24**, 587 (1995)
6.107 D. Shaw: Diffusion in MCT. In: *Mercury Cadmium Telluride*, ed. by P. Capper, J. Garland (Wiley, Chichester 2011)
6.108 P. Capper, C.D. Maxey, C.L. Jones, J.E. Gower, E.S. O'Keefe, D. Shaw: J. Electron. Mater. **28**, 637 (1999)
6.109 M.D. McCluskey, S.J. Jokela: J. Appl. Phys. **106**, 071101 (2009)
6.110 D.J. Fisher (Ed.): *Diffusion and Defects in ZnO, (Trans. Tech. Publications, Pfäffikon* 2013)
6.111 D. Shaw: Semicond. Sci. Technol. **27**, 035003 (2012)
6.112 D. Shaw: Semicond. Sci. Technol. **15**, 911 (2000)
6.113 M. Kuttler, M. Grundmann, R. Heitz, U.W. Pohl, D. Bimberg, H. Stanzel, B. Hahn, W. Gebbhart: J. Crystal Growth **159**, 514 (1996)
6.114 A. Barcz, G. Karczewski, T. Wojtowicz, J. Kossut: J. Crystal Growth **159**, 980 (1996)
6.115 Z. Balogh, Z. Erdélyi, D.L. Beke, G.A. Langer, A. Csik, H.-G. Boyen, U. Wiedwald, P. Ziemann, A. Portavoce, C. Girardeaux: Appl. Phys. Lett. **92**, 143104 (2008)
6.116 M. Strassburg, M. Kuttler, O. Stier, U.W. Pohl, D. Bimburg, M. Behringer, D. Hommel: J. Crystal Growth **184–185**, 465 (1998)
6.117 M. Eslamian, M. Ziad Saghir: Fluid Dyn. Mater. Process. **8**, 353 (2012)
6.118 J.C. Wang, R.F. Wood, P.P. Pronko: Appl. Phys. Lett. **33**, 455 (1978)
6.119 S.K. Srivastava, D.K. Avasthi, W. Assmann, Z.G. Wang, H. Kucal, E. Jacquet, H.D. Carstanjen, M. Toulemonde: Phys. Rev. B **71**, 193405 (2005)
6.120 P. Baeri, S.U. Campisano, G. Foti, E. Rimini: Appl. Phys. Lett. **33**, 137 (1978)
6.121 J.C. Dyre: J. Phys. C **19**, 5655 (1986)
6.122 Y.L. Khait, R. Beserman, D. Shaw, K. Dettmer: Phys. Rev. B **50**, 14893 (1994)
6.123 T. Noda: J. Apl. Phys. **94**, 6396 (2003)

7. Photoconductivity in Materials Research

Stephen Reynolds, Monica Brinza, Mohammed L. Benkhedir, Guy J. Adriaenssens

Photoconductivity is the incremental change in the electrical conductivity of a semiconductor or insulator upon illumination. The behavior of photoconductivity with photon energy, light intensity and temperature, and its time evolution and frequency dependence, can reveal a great deal about carrier generation, transport and recombination processes. Many of these processes now have a sound theoretical basis and so it is possible, with due caution, to use photoconductivity as a diagnostic tool in the study of new electronic materials and devices. This chapter describes the main steady-state and transient photoconductivity techniques applied in the investigation of semiconductors whose performance is limited by the presence of localized electronic states. These materials tend to be disordered, and possess low carrier mobilities and short free-carrier lifetimes in comparison with crystalline silicon. They are often prepared as thin films, and are of interest for large-area applications, for example in solar cells, display backplane transistors, photoemissive devices such as organic light-emitting diodes (OLEDs) and medical imagers. However, examples of where these techniques have been useful in the study of defective crystalline semiconductors are also given. The approach followed here is by way of an introduction to the techniques, the physics supporting them, and their applications, it being understood that readers requiring more detailed information will consult the references provided.

7.1 **Steady-State Photoconductivity (SSPC)** . 153
7.1.1 Definitions and Overview 153
7.1.2 Example Applications in Materials Research 154

7.2 **Constant Photocurrent Method (CPM) and Related Techniques** 157
7.2.1 CPM 157
7.2.2 Dual Beam Photoconductivity (DBP)..... 158
7.2.3 Fourier-Transform Photocurrent Spectroscopy (FTPS) 159

7.3 **Steady-State Photocarrier Grating Method (SSPG)**..... 160

7.4 **Modulated Photocurrent Spectroscopy (MPC)** 161
7.4.1 MPC Background and Experiment..... 161
7.4.2 MPC Density of States Analysis..... 162
7.4.3 MPC Applications 163

7.5 **Switch-on and Switch-off Transients** .. 164
7.5.1 Switch-on Transient..... 164
7.5.2 Switch-off Transient..... 165

7.6 **Transient Photocurrent Spectroscopy (TPC)** 166
7.6.1 TPC Principles 166
7.6.2 TPC Experiment..... 166
7.6.3 TPC Density-of-States Analysis 167
7.6.4 TPC Applications..... 168

7.7 **Time-of-Flight (TOF) and Related Techniques** 168
7.7.1 TOF Mobility and DOS Measurements..... 168
7.7.2 Interrupted Field TOF (IFTOF)..... 170

7.8 **Other Photoconductivity-Related Techniques** 171
7.8.1 Surface Photovoltage (SPV)..... 171
7.8.2 Spin-Dependent Recombination..... 171
7.8.3 Time-Resolved Microwave Conductivity (TRMC)..... 171

References..... 172

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_7

Photoconductivity, defined as the increase in conductivity of a material resulting from the absorption of optical photons, has traditionally played a significant role in materials research, most notably in the study of covalently bonded semiconductors and insulators. The basic processes that govern the magnitude of the photocurrent are the generation of free electrons and holes through the absorption of incident photons, their transport through the material under the influence of an electric field, and their recombination. The study of any of those aspects as a function of illumination, temperature and field strength, and their development over time, will offer insights into the structure and electronic properties of the material under investigation. However, given that several processes may be involved in the production of a specific photocurrent, a sufficiently comprehensive dataset is needed to differentiate between alternative interpretations. For instance, a low photocurrent may be the result of a low optical absorption coefficient at the given photon energy, but it may also be due to geminate recombination of the photogenerated electron–hole pairs, or it may reflect the formation of excitons. Photogenerated charge needs to be transported in a material to realize a current, a process which depends upon the effective carrier mobility. This can vary by up to ten orders of magnitude; from around $10^4\,\mathrm{cm}^2/(\mathrm{V\,s})$ in semiconductors such as crystalline Si and GaAs where band transport dominates, to $10^{-6}\,\mathrm{cm}^2/(\mathrm{V\,s})$ or less in disordered inorganic and organic semiconductors where transport is via low-mobility extended state conduction mediated by carrier traps, or by carrier hopping between localized states, or by a combination of such processes depending on temperature and excitation regimes.

Given the complexity of these processes, use of a range of photoconductivity experiments is often advisable, as is the combination of photoconductivity with complementary techniques such as optical absorption, photoluminescence, dark current activation, charge collection, electron paramagnetic resonance and various pump-probe experiments. A robust model should seek to satisfy all observations, as rationally as possible. From a slightly different perspective, one of the major tasks of the electronic materials scientist is to attempt to correlate material (and device) preparation conditions and micro/nanostructure with electronic properties. It follows that structural information, from analytical tools such as electron and scanning probe microscopies, x-ray and electron diffraction and Raman spectroscopy, is also of key importance.

A variety of experimental techniques based on photoconductivity have evolved. The information that can be obtained relates to the mechanisms of generation and recombination of carriers and their transport by drift and diffusion. The physical parameters to be extracted include:

- Optical absorption coefficient
- Quantum efficiency
- Carrier mobility
- Carrier lifetime
- Mobility-lifetime products
- Minority carrier diffusion length
- Density of localized electronic states and their capture and emission properties
- Charge state and energetic location and spatial variations in transport properties.

The information obtained need not be specific to the method, and will depend on the wider context of the measurements. Recombination can be studied via photocurrent decay from the steady state, but the intensity and temperature dependence of steady-state photoconductivity can also be used to identify different recombination mechanisms, while details of the density of states in the bandgap of a semiconductor can be inferred both from the spectral response and temperature dependence of the steady-state photoconductivity and a detailed analysis of transient photoconductivity.

This chapter focuses on the main photoconductivity techniques applied in the investigation of semiconductors whose performance is limited by the presence of localized electronic states. Many are deposited rapidly as thin films, are structurally disordered and tend to possess low carrier mobilities and short free-carrier lifetimes compared to *benchmark* crystalline silicon. Research has been stimulated by their highly successful commercial application for over 70 years, in xerography, solar energy, large-area displays and medical imaging. Photocopiers are a good example of the evolution of thin-film photoconductors, beginning with photoreceptor drums coated with selenium, through amorphous silicon, to organic semiconductors. Each of these materials has been intensively researched with regard to their charge generation and retention properties, using photoconductivity methods.

Finally, mention should be made of the use of computer simulations in photoconductivity studies. This has been an active complementary field of research for over 50 years, and its value in developing a coherent understanding of the physics of photoconductivity, and electronic transport in general, is considerable. The experiments described herein furnish data that can be integrated into, or compared with the predictions of, computer simulations of physical processes, and it is often by *closing the loop* in this way that confirmation of the validity and limitations of proposed physical models or data analysis schemes can best be achieved. In a broader context, simulations are now used routinely in the development of solar cells and other devices.

This book contains several chapters that strongly support this work, and are referred to below. Comprehensive accounts of the principles of photoconductivity may be found in the monographs by *Bube* [7.1, 2], *Ryvkin* [7.3] and *Rose* [7.4]. Detailed information on computer simulation of steady-state and transient photoconductivity, and comparison with experiment, is provided in references [7.5–7]. Background and procedures for computer modeling of solar cells, which has strong relevance to the techniques presented here, is given in [7.8]. Readers wishing to try out their own simulations are encouraged to download *SC-Simul* [7.9], a suite of linked software applicable to both steady-state and transient photoconductivity studies.

7.1 Steady-State Photoconductivity (SSPC)

7.1.1 Definitions and Overview

The simplest photoconductivity experiment uses a constant monochromatic light source to generate equal excess densities of free electrons and holes, $\Delta n = \Delta p$, that lead to a change in the conductivity (Chap. 2) by

$$\Delta\sigma = \sigma_{\mathrm{ph}} = q(\mu_{\mathrm{n}}\Delta n + \mu_{\mathrm{p}}\Delta p)\,, \qquad (7.1)$$

where q is the electronic charge and μ_{n} and μ_{p} are the electron and hole mobilities respectively. σ_{ph} is termed the steady-state photoconductivity, or SSPC. The basic experimental arrangement is illustrated in Fig. 7.1a, where L and A are the length and the cross-sectional area of the sample. When a potential difference V is applied, a photocurrent I_{ph} is recorded. The corresponding photoconductivity $\sigma_{\mathrm{ph}} = I_{\mathrm{ph}}/(AF)$, where $F = V/L$ is the applied electric field. The end surfaces of the sample are normally contacted by metallic electrodes. However, for thin films rather than bulk samples, coplanar electrodes in the form of contact pads are typically prepared by evaporation through a shadow mask onto the film surface, or deposited firstly onto the substrate, which is then coated with the photoconductor. Interdigitated electrodes of the type shown in Fig. 7.1b may be used to increase the photocurrent if required. The contacts to this type of sample are chosen to be ohmic. The current measured in this way is termed a *secondary* photocurrent, since it arises from replenishment of photogenerated carriers exiting the sample in order to maintain charge neutrality. Note that the rate of flow of carriers around the circuit may be many times larger than the rate of flow of photons, by a factor ideally equal to the ratio of the carrier lifetime to its transit time across the sample, $\tau/(L^2/(\mu V))$. This is to be distinguished from the *primary* photocurrent exiting a solar cell for example, where both flows are ideally equal.

A significant fraction of photogenerated carriers may become immobilized by trapping at localized states in the semiconductor (band tails and/or defects) such that not every part of Δn and Δp contributes equally to the photoconductivity in (7.1) (Chap. 9). The effect of trapping is reflected in values for μ_{n} and μ_{p} that are lower than the theoretical free-carrier or band

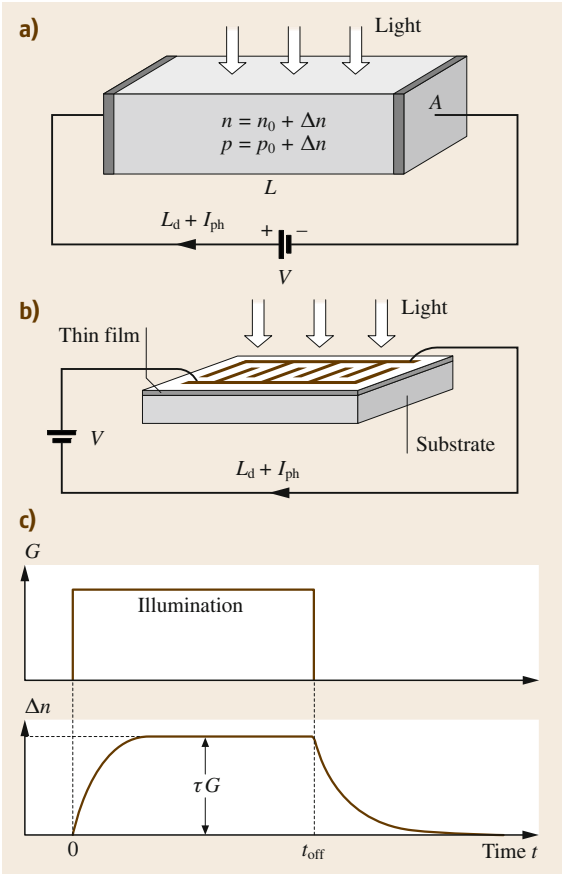


Fig. 7.1 **(a)** Arrangement for photoconductivity measurement. V is the applied voltage, L the sample length and A the cross-sectional area. I_{d}, n_0 and p_0 are the current and the carrier densities in the dark, and I_{ph}, Δn, Δp are the incremental values generated by the illumination. **(b)** Example of interdigitated electrode configuration for a thin film sample. **(c)** Schematic time development of the excess carrier concentration Δn in response to a period of illumination

mobility μ_0. For many materials of interest, either the product $\mu_n \Delta n$ or the product $\mu_p \Delta p$ turns out to be much larger (the majority carrier) than the other (the minority carrier), because of strongly unequal carrier mobilities. For instance, the electron term dominates in intrinsic silicon, while the photocurrent in chalcogenide glasses is carried by holes. In those instances, (7.1) may often be reduced to a one-carrier equation. In $\mu_n \Delta n$ and $\mu_p \Delta p$, the mobility is a material parameter that, in general, will depend on temperature and sample characteristics, while the excess carrier densities Δn, Δp are determined by a combination of material and external parameters. Phenomenologically, the excess density Δn can be written as the product $G\tau_{n,p}$, where G is the rate of generation of free electrons and holes per unit volume, and $\tau_{n,p}$ is the average lifetime of the excess electrons or holes. Substituting into (7.1)

$$\sigma_{ph} = qG(\mu_n \tau_n + \mu_p \tau_p) \,, \tag{7.2}$$

which explicitly displays the mobility-lifetime products frequently used to characterize photoconductors. The relationship between Δn and G is illustrated schematically in Fig. 7.1c. When the light is turned on, Δn increases gradually with time, and eventually reaches the steady-state value we address in this section. When the light is turned off, Δn decays gradually to zero. In the steady state, the generation and recombination rates are equal and constant; in the transient regions, they are not. Time-dependent aspects of photoconductivity are addressed in later sections of this chapter.

The average carrier generation rate G is defined by

$$G = \eta\phi(1-R)\frac{1-\exp(-\alpha d)}{d} \,, \tag{7.3}$$

where η is the quantum efficiency of the generation process, ϕ is the photon flux, the number of photons of energy $E_{ph} = h\nu$ incident on the sample per unit area per unit time, R is the reflection coefficient, α the optical absorption coefficient and d the sample thickness. For monochromatic light, the relation between the intensity I_0 and photon flux is $\phi = I_0/E_{ph}$. A quantum efficiency $\eta < 1$ signifies that not every photon absorbed generates a free electron and hole that contributes to the photocurrent, due to geminate recombination of the carriers, or exciton formation. The values of the parameters η, R and α depend, in general, on the wavelength of the illuminating light. Consequently, monochromatic illumination from a tunable light source can be used to obtain energy-resolved information about the sample, while illumination with white light will provide a global average.

In many measurements on thin films, the condition $\alpha d \ll 1$ will hold over a significant energy range, i. e., when the sample thickness d is small with respect to the optical absorption depth $1/\alpha$. Equation (7.3) can then be simplified to

$$G \cong \eta\phi(1-R)\alpha \,. \tag{7.4}$$

Under these conditions the incident light is uniformly absorbed, and so (neglecting surface effects) G is everywhere constant rather than an average value. Since SSPC is frequently nonlinearly dependent on G, the uniformly absorbed condition is often strongly preferred, especially in quantitative work. However there are instances where nonuniform absorption may be useful, for example to probe surface effects.

The free-carrier lifetimes of excess electrons and holes, τ_n and τ_p, in (7.2) are governed by recombination with carriers of opposite sign. Assuming, for simplicity, the frequently encountered case of photoconductivity dominated by the majority carrier (assumed here to be electrons), the recombination rate can be written as $\tau_n^{-1} = b(p_0 + \Delta p)$, where b is a recombination constant, and p_0 and Δp are the equilibrium and excess minority carrier densities respectively. It then follows that the SSPC may be expressed as

$$\sigma_{ph} \propto \Delta n = G\tau_n = \frac{G}{b(p_0 + \Delta p)} = \frac{G}{b(p_0 + \Delta n)} \,. \tag{7.5}$$

Equation (7.5) indicates that a linear relationship $\sigma_{ph} \propto G$ holds for $\Delta n \ll p_0$ (at low excess carrier density), while higher excitation levels with $\Delta n \gg p_0$ lead to $\sigma_{ph} \propto G^{1/2}$. The resulting linear and quadratic regimes are referred to as *monomolecular* and *bimolecular* recombination. For a given light source and temperature, variations in G correspond to variations in the flux ϕ, and therefore $\sigma_{ph} \propto \phi^{\gamma}$ with the photoconductivity index $0.5 \le \gamma \le 1$. It can also be shown from the above that the response time $\tau \propto \phi^{(\gamma-1)}$. The value of γ also depends on ϕ. However, ϕ is not the only factor that determines the value of γ; intermediate γ values may indicate a $\Delta n \approx p_0$ condition, but they may equally well be a result of an energetic distribution of recombination centers [7.4], as discussed below. In addition, the equilibrium free-carrier densities n_0 and p_0, which play a role in recombination, depend on temperature through the Fermi–Dirac distribution function, thus making recombination a temperature-dependent process.

7.1.2 Example Applications in Materials Research

In photoconductors, recombination is mediated by states in the bandgap. The presence of discrete, or a nar-

row energy range, of trapping levels leads to thermally activated photocurrents, with the activation energy indicating the energetic positions of the traps. *Main* and *Owen* [7.10] and *Simmons* and *Taylor* [7.11] showed that the positive photocurrent activation energy in the monomolecular recombination regime corresponds to the energy above the Fermi level of a donor-like center, while a negative activation energy value in the bimolecular region corresponds to the energy above the valence band edge of an acceptor-like center. Figure 7.2 illustrates this photocurrent behavior for amorphous As_2Se_3 [7.12]. This is characteristic of chalcogenide glasses, where the intrinsic charged defects act as recombination centers [7.13]. SSPC measurements of this type can thus determine the recombination mechanism and the energetic position of discrete defects.

The temperature dependence of the SSPC in amorphous silicon (a-Si:H) does not exhibit a definite activation energy, due to the presence of a more distributed and complex set of traps including band tails and charged and neutral defects (Chaps. 9 and 24). A simplified electronic density of states (DOS) is shown schematically in Fig. 7.3a.

Whenever the DOS in the bandgap of a photoconductor consists of a distribution of traps, quasi-Fermi levels (QFLs) can usefully be defined. In thermal equilibrium, the QFLs and the Fermi level coincide. Under illumination the QFLs *split* from the equilibrium Fermi level such that they define the occupancy of states separately for electrons and for holes. The QFL for electrons may be written

$$E_{\mathrm{Fn}} = E_{\mathrm{F}} + k_{\mathrm{B}}T \ln\left(1 + \frac{\Delta n}{n_0}\right) \qquad (7.6)$$

with an equivalent expression for holes. In the case of the majority carrier, E_{F} may be determined from the activation energy of the dark conductivity σ_{d} and $\Delta n/n_0 = \sigma_{\mathrm{ph}}/\sigma_{\mathrm{d}}$. To a first approximation (see [7.11] for detailed considerations), the QFLs correspond to demarcation energies dividing the DOS into a shallower part, where carriers may be trapped and subsequently emitted, and a deeper part where traps have become recombination centers. This is illustrated in Fig. 7.3b. It is clear that as the QFLs depend on the carrier density, varying the light intensity and temperature will influence the role of a given set of localized states in generation and recombination processes. In addition to the variation with energy of the DOS in the bandgap, the capture and recombination constants for the various centers may also be quite different.

Much progress has been made in understanding photoconductivity in materials with broad distributions of localized states, and in applying SSPC as a diagnostic

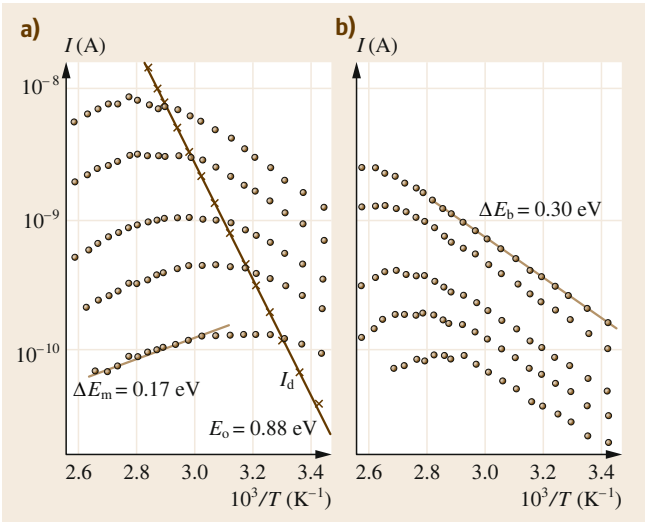


Fig. 7.2a,b Temperature dependence of the steady-state dark and photocurrents in an a-As_2Se_3 sample, **(a)** illuminated at 1.55 eV with intensities of 0.84, 3.5, 9.8, 38 and 120×10^{12} $\mathrm{cm^{-2}s^{-1}}$, and **(b)** illuminated at 1.85 eV with intensities of 0.56, 1.7, 4.6, 27 and 77×10^{12} $\mathrm{cm^{-2}s^{-1}}$. ΔE_{m} and ΔE_{b} represent the photocurrent activation energies in the monomolecular and bimolecular recombination regimes respectively, and E_σ is the activation energy of the dark current I_{d}. (After [7.12])

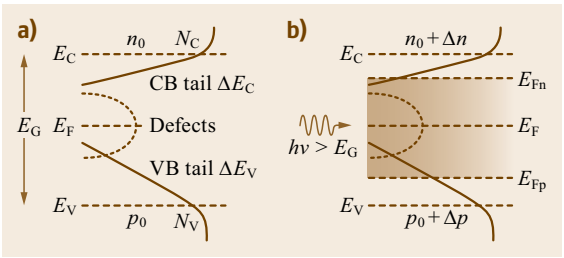


Fig. 7.3 **(a)** Model density of states (DOS) in a disordered semiconductor at thermal equilibrium, electron and hole densities n_0 and p_0 respectively, depicting localized conduction and valence band tail states and defects distributed in midgap. **(b)** Illumination results in splitting of quasi-Fermi levels E_{Fn} and E_{Fp}. Localized states between E_{Fn} and E_{Fp} act as recombination centers, those between E_{Fn} and E_{Fp} and the respective band edges act as carrier traps from which charge is subsequently released

tool. One example is provided by *Brüggemann* [7.14] who studied SSPC in microcrystalline silicon thin films. He applied a result first proposed by *Rose* [7.4] and elaborated by *Bube* [7.2], that the intensity dependence of the SSPC in an intrinsic semiconductor in which electrons are the majority carriers is controlled by local changes in the DOS at E_{Fn}. By adjusting E_{Fn} over a wide range by varying I_0 and T, and measuring γ from

the local slope in a plot of log (σ_{ph}) versus log (I_0), it is possible to reconstruct the DOS in the upper half of the bandgap as a series of segments.

Changing the position of the equilibrium Fermi level by doping also modifies photoconductivity behavior. The effect of n-type doping on the SSPC of a-Si:H has been studied by a number of workers, including *Main* et al. [7.17]. They carried out photoconductivity measurements over a wide range of intensity and temperature, and noted that (i) over certain temperature ranges, around 200 K, values of γ as low as 0.4 could be observed, and (ii) these low values of γ persisted even when the photocurrent was over two orders of magnitude below the dark current. Neither of these results is predicted by the Rose model outlined above. However, they were able to account consistently for the observations in terms of a typical model density of states such as that shown in Fig. 7.3a, but in which recombination takes place solely via defect states. In this case band tails do not act as recombination centers, but are *indirectly* involved as charge reservoirs required to maintain charge neutrality. If the conduction band tail slope ΔE_C is steeper than the valence band tail slope ΔE_V, limiting values of γ in this model are predicted to be $k_B T/\Delta E_V$ at low generation rates, and $\Delta E_C/\Delta E_V$ at high generation rates. Both predictions are borne out experimentally assuming realistic tail slope values (25 meV and 45 meV), and furthermore detailed accordance over the full experimental range was achieved by means of computer modeling. Finally it should be noted that values of $\gamma > 1$ have been observed in a-Si:H, again in conflict with the Rose model, with explanations being offered by a number of workers including *Fritzsche* et al. [7.18] and *Bube* [7.19].

From a materials characterization point of view, SSPC offers the possibility of using (7.4) to determine the absorption coefficient and/or quantum efficiency as a function of the energy of the incoming photons, and thus to explore the bandgap of a semiconductor. When materials with sufficiently well-defined energy levels are studied, local maxima corresponding to specific optical transitions may be seen in the photoconductivity spectra. An example, from the work of *Nesladek* et al. [7.15], is illustrated in Fig. 7.4. This shows the spectrally resolved photocurrent for diamond films prepared by chemical vapor deposition. The rise in photocurrent around 5.5 eV corresponds to the optical bandgap of diamond, while the shoulders at $\approx$ 1.5 eV and $\approx$ 3.5 eV signal the presence of defect distributions

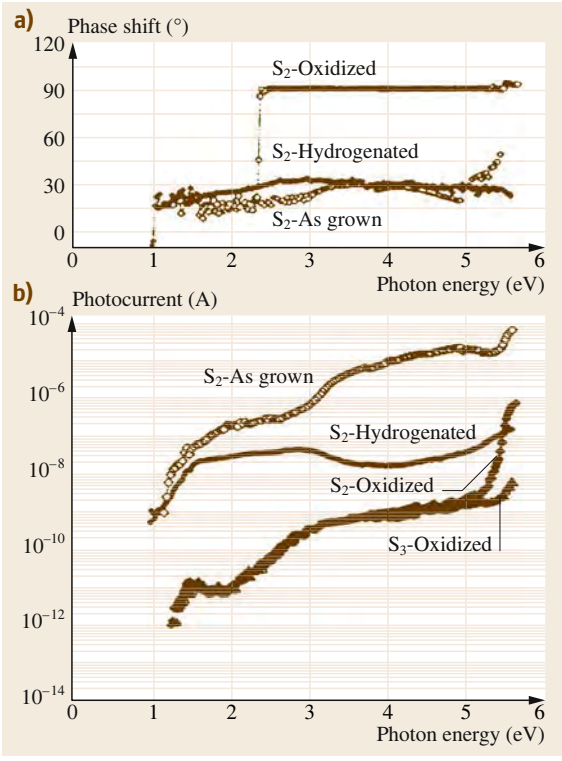


Fig. 7.4a,b Room temperature AC photocurrent spectra, measured at 7 Hz, after various treatments of chemical vapor deposition (CVD) diamond layers deposited at 920 °C (S_2) and 820 °C (S_3). **(a)** Phase shift. **(b)** Magnitude of photocurrent. (After [7.15])

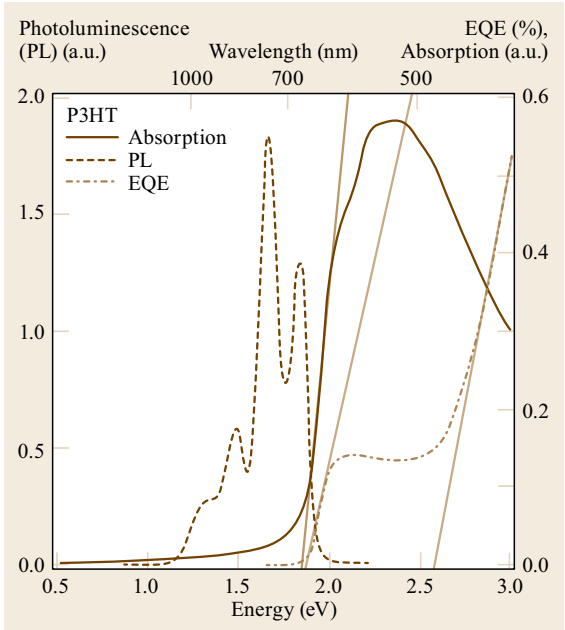


Fig. 7.5 Combination of spectrally-resolved photoconductivity, optical absorption and photoluminescence measurements used to quantify exciton formation in the organic semiconductor P3HT. (After [7.16])

in the bandgap. The data were measured using low-frequency chopped light and a lock-in amplifier. It can be seen that changes in the phase shift between the photocurrent and the chopped light may also reveal specific features in the DOS. This topic is addressed in detail later, in Sect. 7.4 on modulated photoconductivity (MPC).

A further example of the use of spectrally resolved photocurrent, this time to probe organic semiconductors of interest in solar cell development, is shown in Fig. 7.5. *Deibel* et al. [7.16] have measured the quantum efficiency of charge generation in the conjugated polymer P3HT. The onset of strong photogeneration commences at 2.6 eV, compared with a value of 1.9 eV for strong optical absorption in the same sample. The former energy is that required to generate free carriers, whereas the latter is that required to generate an exciton, a bound electron-hole pair (Chap. 3). The difference of 0.7 eV corresponds to the energy required to dissociate the exciton. Photoluminescence (Chap. 38) and photoelectron spectroscopy measurements enable further details of the interactions to be deduced. This example illustrates the benefits of combining a range of measurements to provide linked information on semiconductor properties, which leads in turn to a comprehensive and consistent model.

7.2 Constant Photocurrent Method (CPM) and Related Techniques

7.2.1 CPM

The constant photocurrent method (CPM) is a simple but important development of the basic single-beam measurement described in Sect. 7.1. It was introduced by *Vaněček* et al. [7.20, 21] to determine the optical absorption coefficient $\alpha(E_{\mathrm{ph}})$ of thin-film semiconductors over a very wide range (typically $0.1-10^5\,\mathrm{cm}^{-1}$) as a function of photon energy E_{ph}. In CPM, the photocurrent is kept constant by continually adjusting the light intensity I_0 while E_{ph} is scanned across the spectrum. A constant photocurrent implies that the quasi-Fermi levels have fixed positions, and thus that the free-carrier lifetime τ remains constant.

From (7.2) and (7.4), the photoconductivity of a thin film under uniformly absorbed light ($\alpha d \ll 1$) is

$$\sigma_{\mathrm{ph}} = q\mu\tau(1-R)\eta\alpha\phi\ . \tag{7.7}$$

The product $\alpha\phi$ will thus remain constant, and relative values of α can be determined from it, provided that energy dependencies of the parameters μ, R and η of (7.7) are negligible. The value at which the photocurrent is fixed may be chosen freely, but will in practice be dictated by the low-absorption region of the sample. However, since even low-level photocurrents can still be measured with high precision, the CPM method is especially useful at low values of optical absorption where standard optical transmission measurements lose their accuracy.

It follows from (7.7) that under the above conditions a plot of $1/\phi$ versus E_{ph} will reveal relative values of α. In an improved variant of the CPM technique, termed *absolute* CPM (ACPM) [7.21], optical transmission through the film is measured at the same time as the photocurrent, and the data from the two measurements are combined in order to remove optical interference fringes [7.22] and to calibrate the system using the *Ritter–Weiser* formula [7.23]. The experimental arrangement used in absolute CPM measurements is shown schematically in Fig. 7.6. Depending on requirements, the sample may be mounted in a cryostat to enable operation over a wide temperature range. The choice of monochromator system is important, since at long wavelengths second-order output can easily dominate the sample photocurrent to give erroneous readings. A double monochromator plus order filters are recommended to maintain second- and higher-order output and scattered light below one part in 10^5.

The CPM experiment may be operated with either DC or AC illumination, but as discussed later, the absorption spectra obtained need not be identical. Most researchers use low-frequency AC illumination (chopped light, < 10 Hz) to measure the sample photocurrent and detector outputs, since values may be small (nA to pA) and thus prone to electrical noise and interference

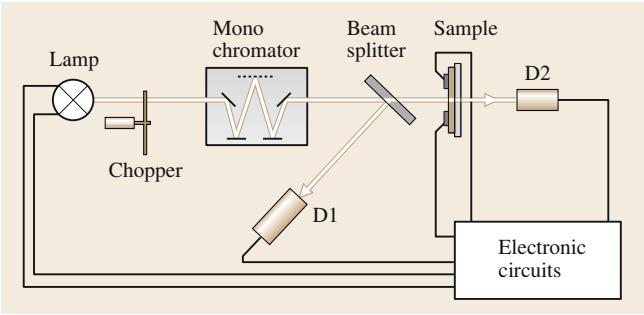


Fig. 7.6 Schematic diagram of an *absolute* CPM setup. Photodetector D1 is used to measure the intensity of light incident on the sample, while detector D2 measures the transmitted light. These signals plus the sample photocurrent are recorded by electronics, the latter being maintained constant by control of the lamp current. Details of operation are given in the text. (After [7.21])

whose effects are minimized by lock-in detection. The CPM measurement process proceeds as follows. The lamp current is adjusted until the sample photocurrent equals the set-point value and remains steady. Readings are taken from the calibrated detectors D1 (proportional to the intensity of the incident beam) and D2 (the transmitted beam). The monochromator wavelength is then decremented, and the lamp current adjusted until the photocurrent once more equals the set-point, D1 and D2 are read, and so on until the desired end point of the wavelength scan is reached. Normally the process is carried out in a closed loop, by means of a control algorithm running on a PC. Depending upon the desired precision, the 30–100 measurement points may take several hours to collect since settling times may be long. For a detailed exposition of calibration and calculations required to generate the final absolute CPM spectrum, the reader is referred to [7.20, 21].

While it is of course useful to know $\alpha(h\nu)$ for a material per se, for example when designing a thin-film solar cell, a wealth of additional information may be extracted. The initial work by *Vaněček* et al. [7.20] was directed at an investigation of thin-film silicon, and CPM emerged as an important technique of choice for assessing material quality as a function of deposition rate, substrate temperature, process gas composition, light exposure and subsequent degradation, and many other factors. Essentially, $\alpha(E_{\mathrm{ph}})$ is a result of all permitted electronic transitions in a given sample. The integral representing this is referred to as the joint density of states (Chap. 3). If a particular set of transitions is dominant, this will be evidenced in the local behavior of $\alpha(E_{\mathrm{ph}})$. The two CPM *figures of merit* most commonly quoted for intrinsic a-Si:H in the literature relate to (i) the Urbach tail slope, the characteristic energy of the exponential rise in optical absorption observed in many disordered materials, and a measure of the degree of structural or compositional disorder, and (ii) the defect shoulder (or bump) in α below the Urbach tail typically at a photon energy of 1.2 eV, that is considered a measure of the dangling bond density. In both cases, transitions from filled states (below the Fermi level) to empty states close to or above the conduction band edge predominate (but see below), since the electron is the majority carrier. It follows that the Urbach energy in a-Si:H revealed by CPM is largely a measure of the valence band tail slope.

As the electronic DOS $g(E)$ is implicitly involved in determining α, it seems reasonable to suppose that there should be an inverse process that enables $g(E)$ to be obtained from α. This turns out to be quite straightforward, provided transitions are limited to those indicated above, and current is carried by a single mobile carrier. The simplest expression, obtained by assuming a constant conduction band DOS and $T = 0$, is that $g(E_{\mathrm{c}} - E_{\mathrm{ph}}) = K\mathrm{d}\alpha/\mathrm{d}E_{\mathrm{ph}}$ between E_{v} and E_{F} [7.24]. The constant K depends on the values of the interaction matrix element and the DOS at the conduction band edge. More elaborate and potentially more accurate procedures assume a parabolic conduction band DOS [7.25], though generally only transitions from filled localized states to electron transport states are considered.

Main et al. [7.26] extended the CPM experiment analysis (on a-Si:H) to include the possibility of optical excitation from the valence band into unoccupied deep defect states (above E_{F}), followed by *thermal emission* into the conduction band. These transitions are weak at high frequencies but significant at low frequencies and DC. Thus DC CPM will tend to return a larger α than AC CPM. These workers also demonstrated that the density of empty states (above E_{F}) could be probed by comparing the AC and DC CPM limiting values.

An increasingly wide range of thin-film semiconductors have been studied by CPM. *Bacioglu* et al. [7.27] determined the Urbach tail slope and estimated the deep defect density in amorphous silicon oxide a-SiO_x : H with varying oxygen concentration [O], calibrated from the Si-O-Si IR stretching peak. Both parameters show a strong dependence, with a linear increase in tail slope from 55 to 120 meV as [O] increases from 0 to 50% and an increase in α at 1.2 eV by a factor of 50 over the same range. This suggests that the compositional disorder of additional oxygen in the structure broadens the band tail, and also results in a greater number of Si dangling bonds. *Willekens* et al. [7.28] studied the organic semiconductors MEH-PPV and PCBM using CPM, both separately and in blended form for application in bulk heterojunction solar cells. They observed differences between the spectra, consistent with charge transfer between components in the blend. *Sharma* et al. [7.29] investigated nanocrystalline CdSe films using CPM following light-soaking and thermal annealing, and report a series of systematic changes in Urbach tail energy and optical bandgap. *Ishihara* et al. [7.30] have demonstrated the usefulness of CPM as a probe of deep defects when investigating transport properties and stability of In-Ga-Zn-O (IGZO) thin films for use in oxide TFTs. This information enabled them to modify preparation to reduce the defect density, which in turn improved threshold voltage stability.

7.2.2 Dual Beam Photoconductivity (DBP)

Like the constant photocurrent method discussed above, the dual-beam photoconductivity (DBP) technique is used to determine the subbandgap optical absorption in a photoconductor [7.31]. DBP achieves constant life-

time conditions by directing an additional constant, uniformly absorbed, *bias* beam of approximately the bandgap energy at the sample. This beam must be set to maintain a generation rate larger than that from the monochromator *probe* beam, throughout the scan. DBP eliminates the need for slow and tedious adjustments of the probe beam required to maintain constant photocurrent conditions in CPM. The probe beam photon energy range may thus be scanned more rapidly than for CPM, using a simpler control and measurement system. DBP requires a very similar experimental setup to CPM, with the addition of a bias light. This may be in the form of a laser beam, directed at the sample obliquely so that the detectors do not become overloaded by it. The probe beam is optically chopped, enabling the resulting sample photocurrent to be separated from the DC bias beam current by lock-in detection and amplification. Similarly, any stray light from the bias beam reaching the optical detectors will be rejected.

A feature of the DBP measurement is that the sample QFLs are more widely split, since the constant DC bias current must be larger than a typical constant photocurrent value in CPM. The resulting differences in occupancy of deep states mean that $\alpha(E_p)$ values recorded by the two techniques may not agree. However the difference(s) can also be used to advantage; the occupancy can be varied predictably by measuring the DC photocurrent, and DBP spectra recorded at different bias beam intensities may be interpreted to yield additional information on defect densities [7.32]. Defect capture properties and charge state, and the presence of surface states, may also be inferred in some cases. An example may be found in *Güneş* et al. [7.33], who studied differences in spectra from annealed and light-soaked hydrogenated amorphous silicon samples in the deep defect region. *Remes* et al. [7.34] have reported that residual nitrogen contamination in nanocrystalline diamond can be identified using this technique.

7.2.3 Fourier-Transform Photocurrent Spectroscopy (FTPS)

The use of a FT spectrometer to record spectrally resolved photocurrent in thin films was first reported by *Inushima* et al. [7.36], and developed as a quantitative technique (FTPS) by *Vaněček* et al. [7.37]. The concept is a development based around a standard technique used in analytical chemistry for many years, the Fourier-transform infrared (FTIR) spectrometer. This technique consists of recording and Fourier-transforming *interferograms* produced by a Michelson setup illuminated by an appropriate broadband light source. In absorption mode, the sample of interest is placed between the light output from the interferometer and a broadband detector, and the interferogram is recorded as the detector output versus optical path difference. The interferogram is then Fourier-transformed and normalized to a background spectrum, yielding a *conventional* IR spectrum of absorbtance versus wavelength. The FT technique effectively scans all wavelengths in parallel and thus has an inherent speed advantage over conventional sequential dispersive spectroscopies. The optical throughput is much higher, so there are considerable benefits in terms of noise cancellation. Overall this makes for a faster, lower noise and potentially higher-resolution approach.

These advantages were recognized as being applicable to the measurement of photoconductive electronic materials at low absorbtance levels [7.36, 37], since the *detector* in the FTIR system may also play a dual role as the *sample*. Many commercial FTIR setups possess an external detector input. In FTPS therefore, the interferogram is recorded by a *detector* consisting of the photoconductor of interest. The photocurrent is amplified and presented to the external input. The scan is carried out, Fourier-transformed, and normalized to a background spectrum previously recorded using a calibrated broadband detector, to yield the absorbtance spectrum of the photoconductor.

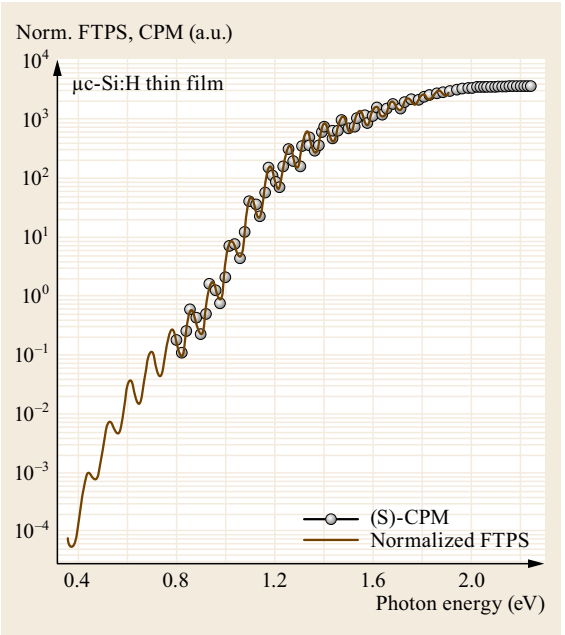


Fig. 7.7 Normalized FTPS spectrum (*full line*) compared with the standard CPM spectrum (*circles*), measured on the same microcrystalline silicon sample. Note good correspondence between the spectra, and greatly extended range of sensitivity available from FTPS. (After [7.35])

The applicability of the FTPS method to a wide range of materials, including amorphous silicon, microcrystalline silicon, diamond layers, nanocrystalline diamond and very thin organic films, was demonstrated in a later publication by *Vaněček* et al. [7.35]. The technique is highly sensitive, and in some cases may be used to extend the measurement of absorptance from the parts per million limit of CPM, to parts per *billion*, as shown in Fig. 7.7 [7.35]. As with standard CPM, interference fringes may appear on an as-recorded spectrum of a thin film. There is no direct equivalent of the transmission measurement used in the absolute CPM approach to remove the fringes, but it is possible to do so if an auxiliary *filter*, optically identical to the sample, is available. The two measurements, with and without filter, can be processed to obtain the thin-film interference-free absorbtance spectrum, by means of the *Ritter–Weiser* formula [7.23]. The FTPS method has been used diagnostically on both films and solar cells. Unlike CPM, the sample photocurrent cannot be kept constant in FTPS, as the interferogram necessarily covers a wide intensity range. However, as with DBP, a constant carrier lifetime may be imposed by applying a uniformly absorbed DC optical bias beam.

It may be concluded that FTPS has several advantages over CPM and DBP, specifically reduced scan time and increased sensitivity, and few disadvantages. A review of the technique and its applications is provided by *Holovsky* [7.38]. Recently, FTPS has been used to extend the range of quantum efficiency measurements on a-Si:H solar cells [7.39]. The increased signal-to-noise ratio over CPM was found to improve resolution of the deep DOS in this case. The use of FTPS to study deep defects in a PbS nanocrystal solar cell has helped quantify aging effects [7.40]. It seems likely that FTPS will enjoy greater usage in the study of photoconductors and photodevices in the coming years, as its potential advantages over other spectrally resolved photocurrent techniques become more widely known.

7.3 Steady-State Photocarrier Grating Method (SSPG)

The SSPG method, first described by *Ritter* et al. [7.41], enables the ambipolar diffusion length L of a semiconductor thin film to be determined. This is a particularly important parameter in materials targeted for use as absorber layers in solar cells, as well as in fundamental studies. The method consists in measuring the photoconductance of a coplanar sample illuminated by a two-beam system when (i) the beams are coherent and form interference fringes parallel to the electrodes, and (ii) the beams are incoherent and superpose as a spatially uniform illumination. A detailed review of the SSPG technique, its theoretical basis, interpretation and applications has been provided by *Brüggemann* [7.42].

A basic experimental arrangement is shown in Fig. 7.8a. Two polarizers are used, to define the plane of polarization (P_1), and to adjust the intensity of the beam emerging from the laser (by rotation of P_2). The beam splitter (BS) provides the requisite two interfering beams. The transmitted beam (main beam) is reflected by the plane mirror M_1 on to the sample S. The reflected beam (probe beam) is likewise reflected by M_2. The intensity of the probe beam is reduced to around 10% of the main beam, ensuring that the modulation amplitude remains small. The sample is mounted on a movable stage, enabling the angle θ between the two beams to be varied by adjustment of M_1 and M_2. The spatial period of the sinusoidal intensity variation, shown in Fig. 7.8b is given by

$$\Lambda = \frac{\lambda}{2\sin\left(\frac{\theta}{2}\right)} . \tag{7.8}$$

A half-wave retarding plate H may be selected and placed in the path of the main beam, at an angle of 45° to the plane of polarization. This rotates the main beam at the sample by 90° relative to the probe beam, and the two beams superpose incoherently. An optical chopper is placed in the path of the probe beam, enabling lock-in detection of the sample secondary photocur-

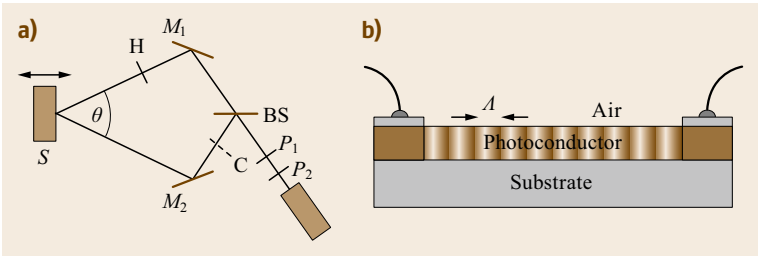


Fig. 7.8 **(a)** SSPG experimental arrangement; optical elements are as described in the text. **(b)** Formation of interference fringes at sample. (After [7.42])

rent under a small external voltage bias. The lock-in amplifier readings U_C and U_I, under coherent and incoherent illumination conditions respectively are recorded for a series of values of Λ, yielding the parameter

$$\beta(\Lambda) = \frac{U_C}{U_I} . \tag{7.9}$$

In turn, β is related to L by

$$\beta(\Lambda) = 1 - \frac{2Z}{\left[1 + \left(\frac{2\pi L}{\Lambda}\right)^2\right]^2} , \tag{7.10}$$

where Z is a constant between 0 and 1, related to various material and system parameters. Low values of Z can indicate poor definition of the grating due to surface light scattering or other inhomogeneities.

Equation (7.10) may be linearized by plotting $1/\Lambda^2$ versus $[2/(1-\beta)]^{1/2}$, yielding a straight line $y = ax - b$, from which $L = 1/2\pi\sqrt{b}$ and $Z = a^2/b^2$ are obtained.

Often the majority carrier $\mu\tau$ product is significantly larger than the minority carrier value, in which case the ambipolar quantity L is determined primarily by the *minority* carrier. The corresponding *majority* carrier product $\mu\tau$ may be determined from the steady-state photoconductivity at the same generation rate, yielding two important figures of merit

$$(\mu\tau)_{\min} = \left(\frac{kT}{2q}\right) L^2, \quad (\mu\tau)_{\text{maj}} = \frac{\sigma_{\text{ph}}}{qG} . \tag{7.11}$$

The SSPG method has been used to investigate diffusion lengths and minority-carrier $\mu\tau$ products in a range of thin-film silicon based semiconductors, including hydrogenated amorphous silicon and silicon-germanium [7.43], polymorphous silicon [7.44] and microcrystalline silicon [7.45]. It has also been applied to thin-film solar cell materials such as GaN [7.46], $CuGaSe_2$ [7.47] and polymer blends [7.48]. Semicommercial automated systems have been developed [7.49]. DOS spectroscopies based on SSPG and moving grating techniques have also been demonstrated [7.50, 51].

The usefulness of the SSPG method in optimization of thin film semiconductor properties is dependent on a number of factors:

1. The position of the Fermi level in a semiconductor changes in response to variations in preparation conditions, doping and degradation effects. This may alter the recombination kinetics and, as a consequence, the diffusion length. Thus it is possible, for example, for L to increase without there being a corresponding improvement in material *quality*.
2. As the method relies upon a uniform photocarrier grating being established, surface roughness or interior structure may scatter light and reduce its uniformity, resulting in inconsistencies and/or inaccuracies in the results.
3. Photocurrents in the SSPG experiment may be small compared with the dark current and this may restrict measurements, especially on doped materials and intrinsic materials at room temperature and above.

These factors are discussed in detail by *Brüggemann* [7.42] and references therein.

7.4 Modulated Photocurrent Spectroscopy (MPC)

7.4.1 MPC Background and Experiment

MPC is a density-of-states spectroscopy, introduced by *Oheda* [7.52] in 1981, used to probe the energetic distribution of unoccupied trap states in the energy gap of a photoconductor. MPC involves the illumination of a coplanar photoconductor sample as shown in Fig. 7.9 with a quasimonochromatic above-gap light source, which is amplitude modulated, typically within the frequency range 0.1 Hz–100 kHz. An LED is a highly suitable source, though chopped light from a monochromator may also be used. A voltage bias is applied to the sample, and the amplitude and phase of the photocurrent relative to the excitation is measured versus modulation frequency using a phase sensitive detector (lock-in amplifier). Generally, the photocurrent lags the light, though there are instances where a genuine phase *lead* may occur.

Both amplitude and phase are sensitive to the transport processes taking place, and both are of use when calculating the DOS versus electron energy. It is important to note that the states probed directly by this experiment are empty trap states that capture majority carriers. This is unlike CPM, where majority carriers are ejected from filled trap states by absorption of a subgap photon. Thus the two techniques are complementary; in the case of amorphous silicon for example, MPC supplies information about the DOS *above* E_F, and CPM about the DOS *below* E_F (but see [7.26]).

Some means of accurately calibrating the amplitude and phase of the light at the sample is essential, as phase shifts introduced by the measurement electron-

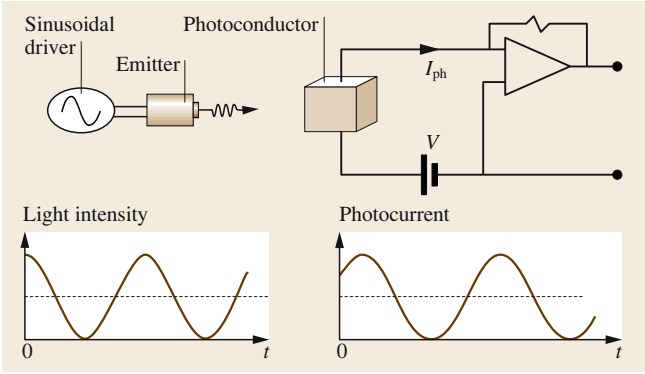


Fig. 7.9 Schematic diagram of an MPC measurement system (*upper frames*), and phase relationship between the exciting light intensity and the photocurrent (*lower frames*)

ics must be eliminated to obtain accurate and consistent results. This may be achieved by connecting a fast pin photodiode to the input of the current pre-amplifier, in addition to the sample. A calibration scan is carried out, with the light source incident on the pin diode and the sample blocked to light, but still electrically connected to the pre-amplifier. Then, the MPC experiment is performed by blocking the pin diode and directing the light at the sample. Complications can arise at high frequency, since *feed-forward* of the electric field from the modulation drive electronics can leak into the preamplifier by capacitive coupling. This manifests itself as a clear phase-leading signal trend. Another source of bogus signals entering the system is the presence of ground loops. Generally it is beneficial to optically isolate the phase reference, or to disable the ground shield on the coaxial cable. A further obstacle is presented by the dark current. AC coupling of instruments will introduce a parasitic phase lead, especially at low frequency. Thus when measuring small photocurrents below 10 Hz in DC mode, it may be necessary to introduce a stable DC offset current at the input to the preamplifier to cancel the dark current, and thus avoid signal overload.

Finally, the sample amplitude is normalized to the calibration amplitude, and the sample phase corrected by subtracting the calibration phase, at each frequency. To probe a suitably wide energy range, several scans at sample temperatures typically between 120 K and 350 K are carried out.

7.4.2 MPC Density of States Analysis

It is possible to extract the DOS from MPC amplitude and phase data by application of straightforward formulae. The degree of success depends upon the sharpness of features to be resolved in the DOS, their energy separation, and the method of analysis chosen. While there are a number of approaches that differ in detail, there are two limiting cases termed the high frequency (HF) and the low frequency (LF) regimes. The former concerns the emission of carriers from traps, the latter the recombination of carriers. A review of the development of these techniques is given by *Kleider* et al. [7.53].

HF MPC

We consider the single carrier case (electrons), with a single species of trap whose capture coefficient is C_n. The electron capture frequency into a set i of traps of density N_{ti} closely spaced in energy and centered at E_i is $\omega_{ti} = C_n N_{ti}$. The emission frequency into transport states just above E_c is $\omega_{ei} = \nu \exp(-E_i/(k_B T))$, where ν is the attempt-to-escape frequency. The capture properties are governed by the trap density, and the emission properties by the trap energy depth. From detailed balance considerations, $\omega_{ti} = \omega_{ei}$ for each set of traps.

The frequency separating the HF from the LF regime is related to the energy at which traps become recombination centers, the quasi-Fermi level E_{Fn}, by $\omega_{Fn} \approx \nu \exp(-E_{Fn}/(k_B T))$ (Sect. 7.1.2). For intrinsic amorphous silicon, ν is often taken as $10^{12}\,\mathrm{s}^{-1}$, and E_F is around 0.8 eV below E_c. Assuming low-level illumination such that the photocurrent is equal to the dark current, $E_{Fn} \approx E_F$, yields $\omega_{Fn} = 0.04\,\mathrm{rad\,s}^{-1}$ at 300 K. The HF regime lies above this, i. e., where the angular frequency of the applied modulation $\omega > \omega_{Fn}$. This threshold depends on the DC generation rate and the temperature. For example, increasing the illumination intensity at constant temperature (300 K) so that the photocurrent equals 100 times the dark current will move E_{Fn} closer to E_c by $k_B T \ln(100) = 0.12$ eV, and increase ω_{Fn} by 100.

In the HF regime, the measured AC photocurrent is furnished predominantly by states whose emission frequency matches the modulation frequency. By solving the rate equations, two useful, though mathematically approximate, relationships linking the DOS to the amplitude and phase of the photocurrent at ω are obtained [7.54, 55]

$$N(E_{\omega n}) = \left(\frac{2}{\pi k_B T}\right)\left(\frac{\mu_n}{C_n}\right)(qAFG_\omega)\left[\frac{\sin(\phi_{n\omega})}{I_{n\omega}}\right] \tag{7.12}$$

and

$$N(E_{\omega n}) = \left(\frac{1}{k_B T}\right)\left(\frac{\mu_n}{C_n}\right)(qAFG_\omega) \times \omega \frac{\mathrm{d}}{\mathrm{d}\omega}\left[\frac{\cos(\phi_{n\omega})}{I_{n\omega}}\right], \tag{7.13}$$

where

$$E_{\omega n} = E_c - k_B T \ln\left(\frac{\nu_n}{\omega}\right). \tag{7.14}$$

G_ω is the AC generation rate. From detailed balance considerations, $\nu_n = C_n N_c$ where N_c is the effective density of states in the conduction band. Typical values are $N_c \approx 10^{20}\,\mathrm{cm}^{-3}$ and $C_n \approx 10^{-8}\,\mathrm{cm}^3\,\mathrm{s}^{-1}$.

Performing the HF MPC experiment over a range of temperatures provides several DOS *sections*, together covering a wide energy range. These sections should overlap provided the data lie within the HF regime and the above parameters are correctly matched. This is more convincing if there is a visible bump or shoulder in the DOS. As the DOS is temperature-independent, any feature should remain fixed at a specific energy for all sections, pinned by a correct choice of ν. A range of capture properties is possible, depending on the type of trap. It might therefore be argued that without additional information HF MPC cannot deliver the absolute DOS, but only the product $N_t C_n$.

An equivalent set of equations for holes may be derived, and in general both electrons and holes will contribute to the photocurrent and thus to the apparent DOS. There is no means of distinguishing between these contributions. However, given the ratio μ/C appears as a scaling factor in (7.12) and (7.13) the HF DOS technique will tend to probe either the upper or the lower half of the bandgap, controlled by trapped electrons or trapped holes.

Of the two expressions, (7.12) is capable of resolving features in the DOS of the order of $k_B T$ apart in energy, whereas (7.13) has an improved resolution, of order $k_B T/2$. The price to pay is a poorer signal-to-noise ratio in the latter case, since a derivative of the experimental data must be taken. Their relative merits have been compared in detail using computer simulations and by experiments on amorphous silicon, which we return to later.

LF MPC

The LF MPC technique is based on the relationship between the recombination lifetime and the DOS at the quasi-Fermi level for trapped carriers. *Koropecki* et al. [7.56] showed that for the LF MPC experiment the DOS at a probe energy E_{pr} is given by

$$N(E_{pr}) = \frac{\tau_R G_{DC}}{k_B T \ln(2)}\,, \tag{7.15}$$

where

$$\tau_R = \frac{\tan(\phi)}{\omega} \quad \text{and} \quad E_{pr} = E_{Fn} + \frac{4}{3} k_B T \ln(2)\,. \tag{7.16}$$

τ_R is obtained from the slope of a plot of tan (ϕ) versus ω, and E_{Fn} from the ratio of the photocurrent to the dark current (Sect. 7.1.2).

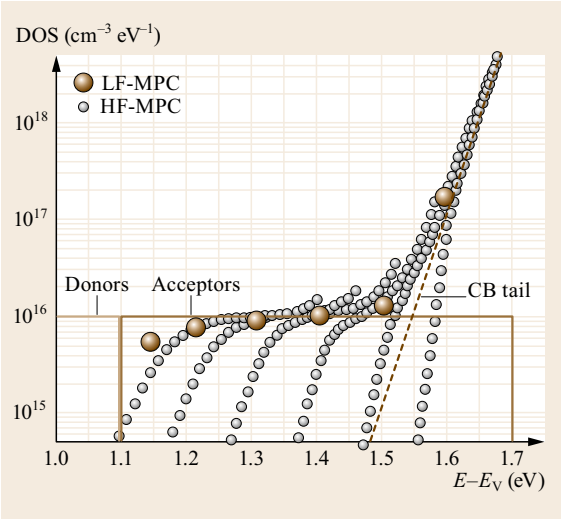


Fig. 7.10 Comparison of HF and LF MPC techniques. The model DOS shown was used to computer generate MPC $I(\omega)$ data over the range 100–350 K in 50 K steps, interpreted using (7.12) and (7.15) to yield the DOS points shown (*gray circle* and *brown circle*). The range of validity of HF MPC is illustrated by a rapid fall in the apparent HF DOS close to the QFL at each temperature. (After [7.57])

The effectiveness of the LF technique in comparison to the HF technique has been investigated in detail [7.56], and a comparison based on data generated by a computer model is shown in Fig. 7.10. The primary advantages of the LF technique are (i) since LF MPC is performed at higher generation rates, the signal-to-noise ratio of the recovered signal is improved; and (ii) the relationship does not involve the attempt to escape frequency. From (ii) it follows that the intersection of the two techniques may therefore be used as a calibration to deduce the trap parameters. This and other developments in interpretation are explored by *Schmidt* et al. [7.57]. One apparent disadvantage of LF MPC in comparison to HF MPC is that the improved resolution offered by (7.13) is not available.

7.4.3 MPC Applications

Following the initial demonstration of MPC by *Oheda* [7.52] in 1981, development was focused on prototype disordered semiconductors such as amorphous As_2Se_3 [7.58, 59], a-Si:H [7.55, 60] and alloys such as a-Si:Ge:H [7.61]. An example of the use of MPC in the study of light-induced defect creation and subsequent thermal annealing in polymorphous silicon [7.62] is shown in Fig. 7.11. Application has extended to a wider range of materials, including the following examples: *Gorgolis* et al. [7.63] have studied the DOS in thin films

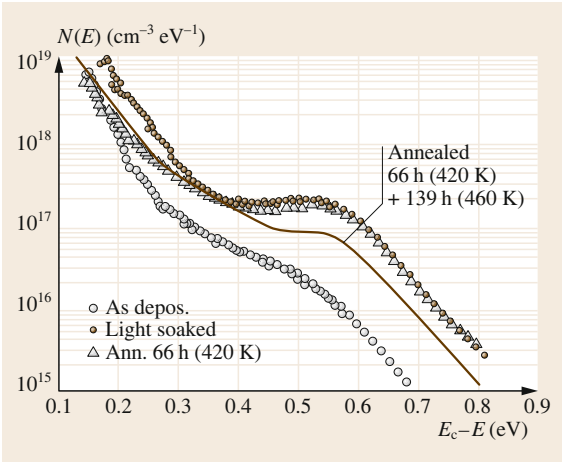


Fig. 7.11 DOS below the CB edge of a polymorphous silicon sample deposited at 423 K and measured by MPC as follows: (i) as-deposited (*gray dots*), (ii) after light soaking (*brown dots*), (iii) annealed for 66 h at 420 K (*triangles*), (iv) annealed for further 139 h at 460 K (*solid line*). (After [7.62])

of the oligomeric semiconductor pentacene, used in thin-film transistor applications. They conclude that localized states in this material comprise exponentially distributed band tails due to disorder, plus a Gaussian distribution of metastable deep defects that arise from impurities. $CuGaSe_2$ films similar to those used in chalcopyrite solar cells have been studied by *Krysztopa* et al. [7.64]. They identified five distinct defect species, correlated with Cu-rich and Ga-rich phases, and measured their capture properties. Results were consistent with junction capacitance measurements on defect levels. The hole conductor a-GeTe and phase-change alloy a-$Ge_2Sb_5Te_5$ were investigated by *Luckas* et al. [7.65], who identified features in the spectrum associated with defects, and demonstrated that correlation of MPC spectra at different temperatures was substantially improved when variations in the bandgap with temperature are taken into account. *Erslev* et al. [7.66] mapped a wide energy range of localized gap states in the transparent semiconductor amorphous zinc tin oxide (ZTO). They confirmed that the conduction band tail is extremely steep (10 meV), consistent with the known high electron mobility in ZTO despite its amorphous structure.

7.5 Switch-on and Switch-off Transients

7.5.1 Switch-on Transient

Immediately after switch-on, G is constant and the photocarrier density evolves with time, through trapping, emission, and recombination. This may be highly complex, depending on material properties and external conditions of intensity and temperature. However there may be circumstances where the situation can be simplified, for example through an assumption of a specific DOS profile, or if certain processes are considered to dominate over a particular time interval.

The switch-on regime has not been greatly studied in comparison to SSPC, or the switch-off regime covered in the following subsection, due to the complexity introduced if both G and R are nonzero. However *Kastner* and *Monroe* [7.67], and *Street* [7.68] analyzed switch-on assuming an exponentially distributed tail of states. They followed what is referred to as the TROK [7.69, 70] model, coined from the initials of the authors, which was successful in explaining steady-state photoconductivity in a number of disordered semiconductors, and other aspects of transient photoconductivity addressed later (Chap. 9). In the steady state, it is predicted that $\sigma_{ph} \propto G^{\gamma}$, where $\gamma = (1-\alpha)^{-1}$, the dispersion parameter $\alpha = k_B T/\Delta E_0$, and $\Delta E_0 = k_B T_0$ is the characteristic energy of the exponential band tail DOS, $g(E) = N_0 \exp(-E/\Delta E_0)$. The early part of the switch-on characteristic was predicted to follow $\sigma_{ph}(t) \propto t^{\alpha}$, with σ_{ph} directly proportional to G, which was confirmed experimentally for a-Si:H. Detailed behavior depends on the ratio of trapping and recombination coefficients. In some instances overshoot of the steady-state value is predicted, and has been observed. Although the possibility of utilizing switch-on as a more detailed probe of trapping and recombination has been identified [7.67, 68], relatively little further work has since taken place.

Some ten years earlier, *Main* [7.71] observed a linear increase in switch-on photocurrent at times of the order of 100 ns, in the chalcogenide glass As_2Te_3, as reproduced in Fig. 7.12a. Assuming deep trapping and recombination to be negligible, he attributed this to the progressive filling of a discrete set, or a rapidly falling (linear or Gaussian) tail, of shallow traps. These traps maintain a *reservoir* of charge, in quasi-thermal equilibrium with the band, with the ratio of free to trapped charge remaining constant over the duration of the linear rise in current. In this case the trap-limited mobility is given by

$$\mu_{\text{eff}} = \frac{\left(\frac{\mathrm{d}I_{\text{ph}}}{\mathrm{d}t}\right)_{t\approx 0}}{q\phi F \eta L(1-R)} \,. \tag{7.17}$$

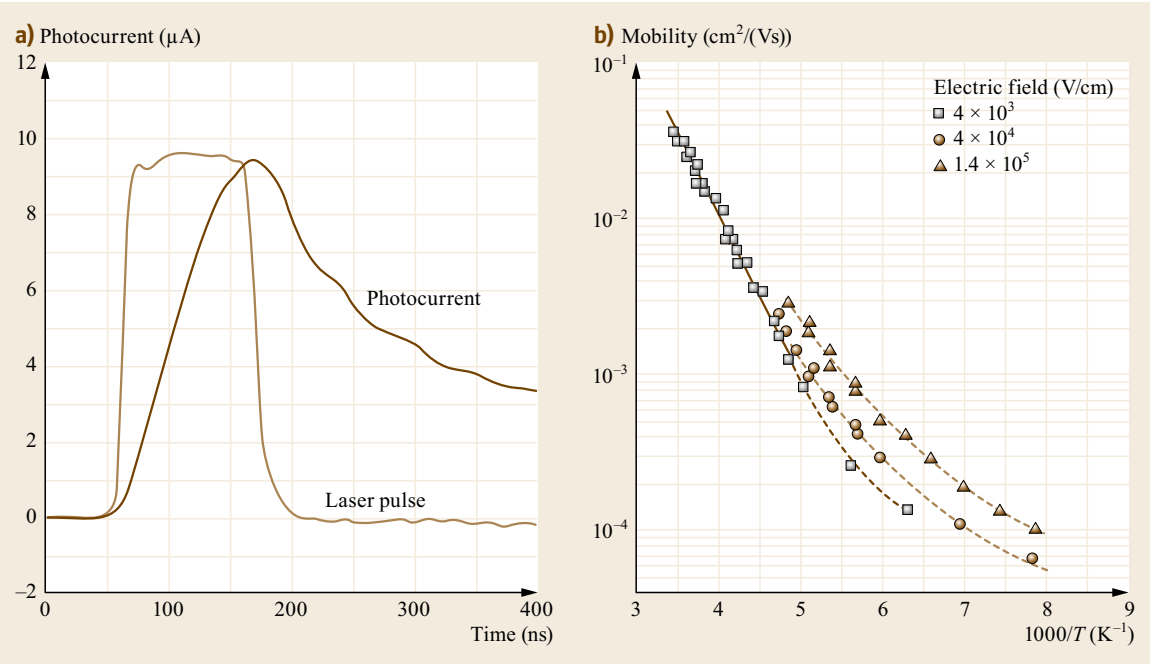


Fig. 7.12 **(a)** Switch-on measurement on an As_2Te_3 sample. The photocurrent increases approximately linearly over the duration of the pulse. **(b)** The rate of current increase is interpreted to yield a shallow-trap limited mobility that is temperature- and electric field-dependent. (After [7.71])

As shown in Fig. 7.12b μ_{eff} is thermally activated, with an activation energy related to the reservoir depth. This may be compared with the well-established procedure for measuring trap-limited mobility, the time-of-flight (TOF) or *Spear* method described later. The switch-on method could be used as an alternative if the dielectric relaxation time is short relative to the TOF drift time, where the uniform field assumption is invalidated. However, TOF has the advantage that it can also be used to measure both electron and hole mobilities separately.

7.5.2 Switch-off Transient

At switch-off the generation term is eliminated from the rate equation that describes the nonequilibrium carrier distribution, but the recombination kinetics are not immediately altered. Consequently, the initial photocurrent decay is governed by whatever recombination mode existed under SSPC conditions. Because of the presence of traps, the initial decay is not directly related to the free carrier lifetime, since carriers trapped at the instant the light is switched off must be emitted to the band at least once to become free, and in general may be trapped and emitted many times before they recombine. Given that in addition, recombination may be a monomolecular or a bimolecular process, the switch-off current is expected to exhibit a rich variety of forms. *Main* et al. [7.72] reviewed this topic by means of analysis supported by computer simulation. For parity with existing work they assumed an exponential band tail and examined the current decay under strong or weak retrapping conditions, governed by the relative probabilities of free carrier capture by traps and recombination centers. These conditions may be selected by the value of G prior to switch-off. It was shown that even for this relatively simple DOS, the initial rate of decay of current need not involve emission from or multitrapping in states near the quasi-Fermi energy, and thus any analysis making such an assumption will be limited in its range of validity. This demonstrates the value of computer modeling in supporting interpretation of photocurrent behavior.

In a different context, the switch-off photocurrent decay in (usually crystalline or polycrystalline) semiconductors with relatively well-defined deep trap(s) has yielded a mainstream tool, in use for some 40 years; it is often termed photoinduced current transient spectroscopy (PICTS), [7.73]. PICTS involves measurement of the quasi-exponential decay of photocurrent associated with emission from discrete or narrow bands of traps at a given energy depth, using the rate-window concept. Trap density and trapping coefficients are obtained from a series of measurements versus temperature, often in conjunction with addi-

tional complementary measurements, enabling species of traps to be identified. This is a different scenario to the broadly distributed DOS discussed above, though the influence of strong and weak retrapping and recombination mechanisms is of importance in both cases.

7.6 Transient Photocurrent Spectroscopy (TPC)

7.6.1 TPC Principles

In the transient photoconductivity (TPC) experiment, free carriers are excited into the bands of a semiconductor sample at time $t = 0$ by a short pulse of uniformly absorbed above-gap light. Since the carrier distributions are in thermal equilibrium at the beginning of the experiment, the trapping sites for electrons above E_F and for holes below E_F are empty, and photogenerated carriers are not excluded from any of those sites. Trapping is dependent on the density of states, a high density resulting in a rapid trapping rate. When a carrier becomes trapped it is immobilized until emitted some time later, in a thermally activated process with the trap depth being the activation energy. As shallower states release trapped carriers sooner, retrapping will lead to increased occupation of the deeper states as time proceeds, and a consequent reduction in the free carrier density. This *thermalization* of charge to progressively deeper energies continues until recombination sets in, at which point charge is stripped away. This is manifested as a steeper fall in the measured current, often at times > 1 s at 300 K. However this does not correspond to the free carrier recombination time, as on average carriers spend far more time in traps than in the band.

The TPC experiment is carried out in secondary photocurrent mode, making use of the photoconductive gain. The current, measured over as wide a time range as possible, will contain information on the density of localized states and their depth. How that information might reliably be extracted to yield the DOS versus energy is considered shortly.

7.6.2 TPC Experiment

The TPC experimental setup is shown in Fig. 7.13. Carriers may be generated by means of a pulse of order 1 ns in duration, typically from a nitrogen-pumped dye laser. Longer pulses from an overdriven LED may also be used. Peak laser powers of the order of 100 kW need to be attenuated, typically by a factor of 1000, to avoid saturating the traps one wishes to measure. The transient current decay is recorded using a fast oscilloscope, with an analogue bandwidth of around 1 GHz, signals being transferred to a PC for further processing. The current transient fidelity at short times is limited by RC time constants and impedance matching, to around 1–5 ns. Performance may also be limited by electromagnetic interference from the laser, which can be reduced by screening and/or by use of a fiber-optic delay line. Repetitive averaging is highly effective in reducing uncorrelated noise.

At short times the transient current may be recorded as the volt-drop across a low-value sampling resistor, 50 Ω being a common choice, but *RC* limitations mean a larger value of resistor is often unsatisfactory at longer times and a fast transimpedance pre-amplifier is a better option. A suitable unit provides a gain of 10^5–10^9 V/A with a rise time of 100 ns–1 ms. The pre-amp may usefully possess a DC offset option to enable the dark current to be nulled. At long times a recording limit is often imposed by the drift of dark current, and so a stable cryostat is desirable. It is important to note that sufficient time should be allowed between pulses to enable the system to return close to thermal equilibrium, otherwise deep traps may become significantly filled during repetitive averaging.

The current transient can contain useful information out to some 100 s, thus a dynamic range of around ten orders of magnitude in both time and current may be necessary. To achieve this necessitates the transient be-

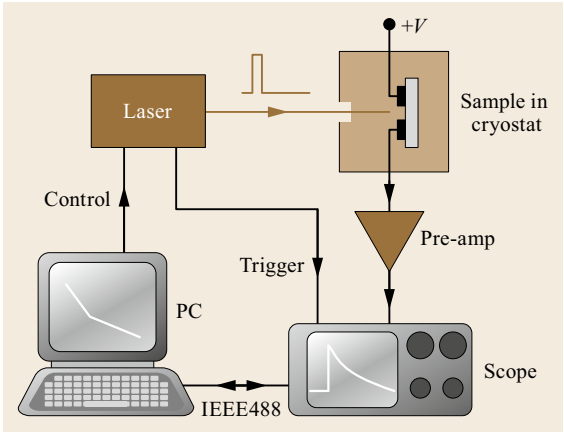


Fig. 7.13 Typical coplanar TPC measurement system. A similar arrangement is used for TOF except the sample is a sandwich structure as shown in Fig. 7.17, $+V$ is a bias step rather than DC, and the laser pulse is delayed for 0.1–1 μs e.g., by means of a reel of optical fiber

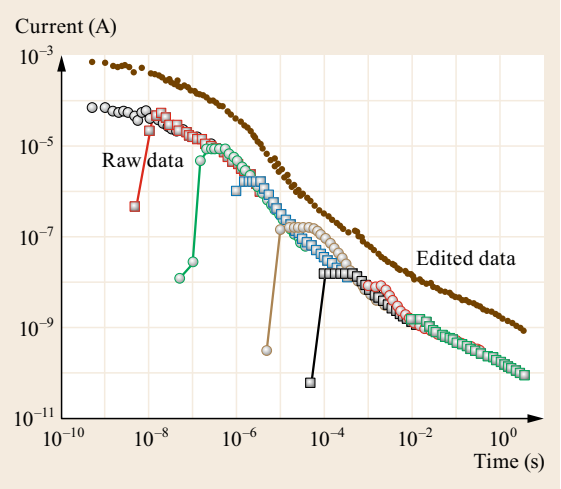


Fig. 7.14 Sections of TPC decays obtained from an a-Si:H sample at room temperature, recorded at progressively longer times (and with greater current amplification). By overlapping the sections and then removing obvious instrumental artifacts, a generally smooth and consistent curve extending over some ten orders of magnitude of time may be obtained

ing recorded in 4–6 sections, overlapped and edited to give a single log-log trace for subsequent processing, as illustrated in Fig. 7.14. Several such traces are measured over the temperature range 100–400 K to give broad energy coverage.

7.6.3 TPC Density-of-States Analysis

Central to the TROK model introduced earlier is the concept of the thermalization energy

$$E_{\mathrm{th}} = k_{\mathrm{B}} T \ln(\nu t) \,. \tag{7.18}$$

This is the energy at which traps will, on average, have released once after a time t. In the case of a purely exponential tail of states, the traps that control the instantaneous current are located at this energy depth. Provided the experimental temperature is less than the band tail temperature ($T < T_0$) the TROK analysis yields $I(t) \propto t^{-(1-\alpha)}$, i. e., the current decays according to a power law that is related directly to the DOS slope. Essentially exponential distributions have been shown to dominate the valence band tail of equilibrated amorphous As_2Se_3 samples over a wide energy range [7.74], but there are few other examples of extensive featureless decays.

In general, the instantaneous current is the outcome of a range of kinetic interactions, and $I(t)$ versus t cannot be mapped to the DOS in such a simple way. For example, if in addition to the band tail there is a *bump* in the DOS at deeper energies due to a species of defects, *capture* into this bump will commence at a time when the thermalizing charge has reached a depth in the tail at which the bump DOS equals the band tail DOS (assuming both have the same capture coefficient). This may abruptly reduce the measured current, but it does not signify anything *unusual* existing in the DOS at E_{th}.

To enable TPC to be applied reliably as a *DOS spectroscopy* requires circumstances such as these to be interpreted correctly, and the fact that both trapping and emission may in general contribute to the measured current throughout the time range requires a more holistic treatment. This has been a topic of interest for several decades, and a variety of techniques for inverting the $I(t)$ curve have emerged [7.75–77]. The more enduring of these involve either Fourier [7.78] or Laplace [7.79, 80] transformations. As the impulse response and the HF-MPC response constitute a Fourier transform pair, a discrete Fourier transformation of the entire TPC time series may be carried out to obtain the frequency domain response

$$\begin{aligned} I(\omega_{\mathrm{n}}) &= \sum_k i(t_k)[\cos(\omega_{\mathrm{n}} t) - \mathrm{j}\sin(\omega_{\mathrm{n}} t)]\Delta t_k \\ &= i_{\mathrm{p}} - \mathrm{j} i_{\mathrm{q}} \,. \end{aligned} \tag{7.19}$$

The in-phase i_{p} and quadrature i_{q} components define the amplitude and phase terms input to the HF-MPC expressions (7.12)–(7.14) presented earlier. This pathway to extracting the DOS makes no a priori assumptions as to its form, though it includes some mathematical approximations, and may suffer from artifacts due to short-time data restrictions. The DFT approach [7.78] has proved to be robust, and is applicable to both pre- and post-recombination regimes. Its performance has been characterized rigorously by inverting computer-simulated $i(t)$ curves generated from model DOS distributions, and then comparing the result with the model [7.78–80]. Figure 7.15 shows an example of two FT DOS methods, of lower and higher resolution, applied to the study of band tails and defects in a-Si:H.

Experimentally, the main advantage of the TPC approach over MPC as a DOS spectroscopy is that the high-frequency limit is somewhat extended. At a given temperature, this may result in improved definition of shallow states. There may be discrepancies [7.81] between the two techniques regarding the recovery of deep levels, since the DC optical bias inherent in HF-MPC alters the occupancy of these states whereas TPC either does not do so, or does so rather differently. Strictly speaking, HF-MPC should be compared to TPC with the addition of optical bias.

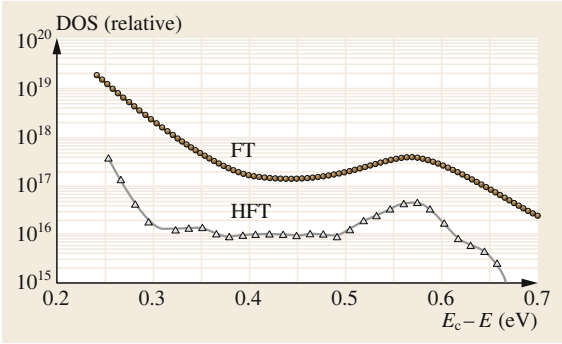


Fig. 7.15 DOS below the conduction band edge in a-Si:H, obtained through Fourier transformation of identical TPC decays. The curve marked *FT* uses the analysis of (7.12). *HFT* is a high-resolution analysis described in [7.77]. The latter offers improved resolution of both band-tail and defect *bump*, at the expense of somewhat increased noise

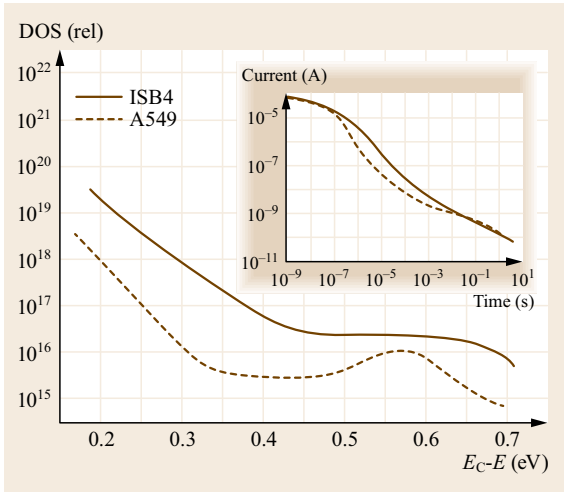


Fig. 7.16 Comparison of DOS in two samples of a-Si:H. Sample ISB4 is from a commercial PECVD reactor. Falling solar-cell efficiency was traced to a number of quality control issues. The conduction band tail slope (35 meV) and deep defect density (7×10^{15} cm^{-3}) were both poor compared with sample A549 University of Dundee *A* gear PECVD material (24 meV and 1.5×10^{15} cm^{-3} respectively). *Inset* shows current-time curves

7.6.4 TPC Applications

Unlike the time-of-flight (TOF) technique discussed below, TPC has the disadvantage that it cannot be used to measure carrier mobility, or at least not directly. In addition, the current-time curve is determined by transport of majority carriers, and thus TPC DOS spectroscopy is restricted to probing the localized states with which majority carriers interact. It is only occasionally possible to probe the DOS on the other side of the bandgap, and this somewhat by accident – in n-type amorphous silicon, the current decay at intermediate times is controlled by emission of holes from the valence band tail [7.82]. However TPC has the advantage of simplicity; there are normally no issues with dielectric relaxation time, space-charge or difficult contacts, and the current measured in the prerecombination region is generally comfortably large. As current is traveling in the plane of the film it may be useful to compare results with the TOF technique applied to the same material, where current travels through the film. In this way, anisotropies in defect density and band tail profile in microcrystalline silicon have been identified [7.83]. A further benefit of the coplanar technique is that it is possible to estimate surface versus bulk defect densities by comparing results using uniformly absorbed and strongly absorbed light respectively [7.84].

Figure 7.16 illustrates how TPC has assisted in identifying a quality control problem in an industrial amorphous silicon PECVD reactor. The reader should not however surmise that practical application is restricted to thin silicon films. Similarly to MPC, TPC has been used to study transport in a diverse range of materials, including short-chain organic molecules [7.85], spray-coated WO_3 [7.86], crystalline CdTe:Sn [7.87] and TiO_2 single crystals [7.88].

7.7 Time-of-Flight (TOF) and Related Techniques

7.7.1 TOF Mobility and DOS Measurements

The time-of-flight (TOF) measurement technique, initially referred to as the *Spear* method, was developed in the 1960s from the Haynes–Shockley experiment as a means of measuring transport properties in low-mobility, high-resistivity semiconductors such as chalcogenide glasses [7.89]. A recent review of the technique has been given by *Bronger* [7.90].

While the experimental setup is similar to that of TPC shown in Fig. 7.13, there are key differences:

1. The sample is in sandwich configuration, typically 1–100 μm thick.
2. The contacts must be blocking to carriers.
3. At least one contact is semitransparent.

A typical TOF sample configuration is shown in Fig. 7.17. Thin metal films forming Schottky contacts

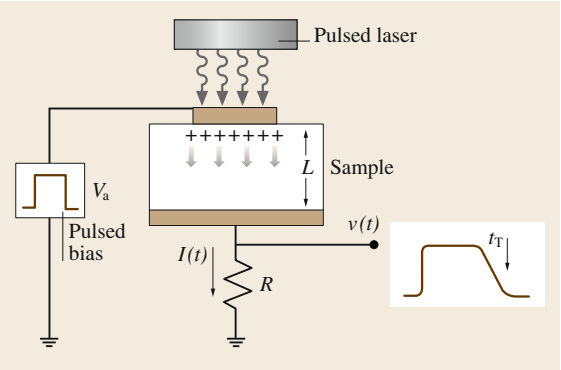


Fig. 7.17 TOF measurement sample arrangement, shown for the case of holes drifting through length L under a positive applied voltage V_a. Transit time is t_T

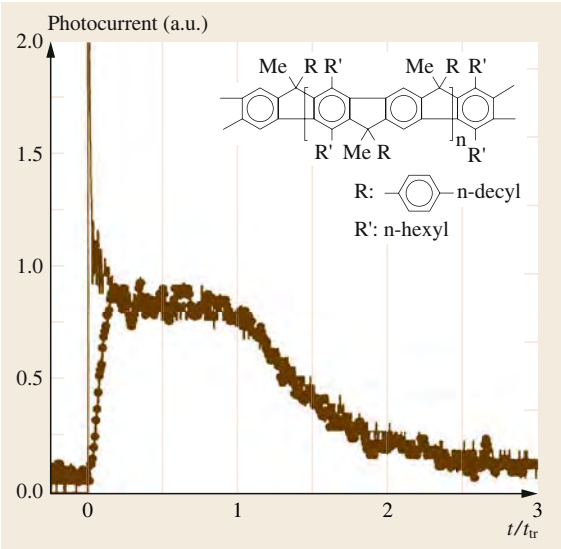


Fig. 7.18 Time-of-flight transients measured at 243 K in methyl-substituted ladder-type poly(para)phenylene (MeLPPP) with 60 kV/cm (*line*) and 300 kV/cm (*dot*) applied, and normalized to a transit time set to 90% of the pretransit current. The *inset* shows the chemical structure of MeLPPP. (After [7.91])

(Chap. 8), or a p-i-n or other solar cell type structure, are commonly employed. The measurement procedure is as follows. A reverse-bias voltage step is applied, establishing a uniform electric field $F = (V_a + V_{bi})/L$, where L is the sample length and V_{bi} the built-in voltage. Following a brief settling time, a strongly absorbed (laser) light pulse generates a sheet of carriers close to one or other contact. In hole TOF for example, the laser is directed at the positive contact; excess electrons are removed rapidly, and excess holes drift across the sample to be extracted at the negative contact. Neither carrier is replenished, as the contacts are blocking. This primary photocurrent is recorded as an equivalent displacement current in the external circuit. What may be seen is a period of relatively constant drift mobility, as carriers traverse the sample, followed by a period where the current falls more rapidly, as carriers are extracted. This is illustrated in Fig. 7.18 for the organic semiconductor MeLPPP [7.91]. The changeover in behavior denotes the transit time t_T, the average time taken for carriers to cross the sample, and the resulting *drift mobility* is $\mu_D = L/(F\,t_T)$. The experiment may be carried out by directing the laser at either side, so that in principle both electron and hole drift mobilities may be determined from separate experiments on the same sample. It should be noted that the electric field remains constant for a period somewhat less than the dielectric relaxation time, which thus defines the maximum timescale of the experiment. The photocharge must be limited to a fraction of the product CV_a, where C is the sample capacitance, to avoid distortion of the field.

Many materials exhibit dispersive transport [7.75], due to capture and subsequent emission of carriers from traps distributed at various energy depths, rather than conventional diffusion of the charge sheet. In such cases it is often more revealing to plot the current-time curve on a log-log scale. Figure 7.19 shows a temperature series of hole transits in amorphous silicon (Chap. 25) deposited by the expanding thermal plasma technique [7.92], as an illustration. A geometric construction may be made to define the transit time, or alternatively, the time at which a proportion (say 50%) of the total charge has been collected may be adopted. In dispersive transport the mobility is not a constant for a given material, since charge remaining in the sample for longer will have an opportunity to become more deeply trapped, which further delays its progress. Thus a TOF experiment performed at low electric field will yield a lower mobility than that at high field. It is thus important to report sample details and experimental conditions when comparing data from sample to sample, and from lab to lab [7.93]. Figure 7.19 illustrates that as the experimental temperature is decreased at constant electric field, the current is reduced and the transit is lengthened. The total collected charge however remains constant.

A complication in the TOF experiment is that if deep trapping occurs it may produce an artifact similar in appearance to a transit [7.94], but is distinguishable by its field independence. A high density of deep traps will thus curtail the useful TOF time range.

While majority carriers dominate photocurrents in TPC, TOF allows independent measurements of ma-

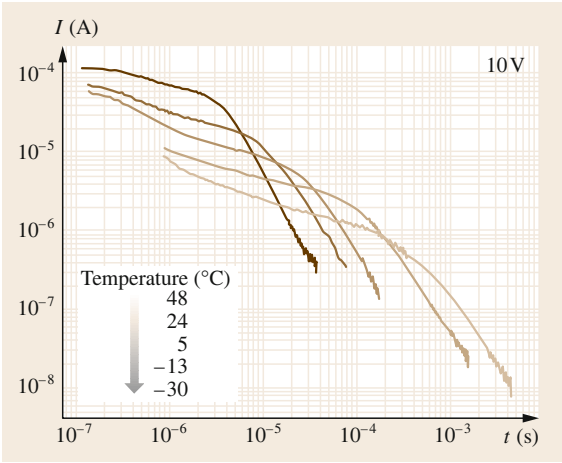


Fig. 7.19 Example of TOF hole transients measured at several temperatures, with 10 V applied across a 5.6 µm-thick a-Si:H sample grown in an expanding thermal plasma at 0.85 nm/s and 250 °C substrate temperature, sandwiched between Mo contacts. (After [7.92])

jority and minority carrier interactions with localized states, and separate examinations of the conduction and valence band sides of the bandgap is possible. Following TROK [7.69, 70] in the case of purely exponential band tails, pre- and post-transit currents take the form $I(t) \propto t^{-(1-\alpha)}$ and $I(t) \propto t^{-(1+\alpha)}$, respectively. Since $\alpha = k_B T/\Delta E_0$, the conduction band tail slope is obtained from electron TOF, and the valence band tail slope from hole TOF. The curves in Fig. 7.19 may thus be interpreted in terms of an exponential valence band tail, at least to first order. Just as with TPC, the current-time curves may be analyzed spectroscopically to yield a less prescriptive DOS, using procedures such as the Fourier transformation method. The approach is valid for both recombination (TPC) and extraction (TOF) of carriers, since analytically both are represented by a generic *time constant*.

In the post-transit regime, current arises from emission of carriers that are more likely to be extracted than to be retrapped as deeply again. Here DOS analysis need consider only a distribution of emission times, simplifying interpretation [7.94]. Specific procedures offering higher energy resolution under these conditions have also been developed [7.95].

TOF is now a well-established characterization technique, used in the study of a wide range of contemporary thin-film electronic materials. The main criteria concerning its applicability are that:

1. The dielectric relaxation time must be sufficiently long to allow a constant electric field to be maintained over the transit.
2. Suitable blocking contacts must be established.
3. A sample of appropriate thickness must be prepared for transit times to lie within the measurement window.

A wide range of organic semiconductors (Chap. 37) satisfy these criteria, and TOF in combination with complementary techniques has supported the advancement of transport models in this field [7.96]. TOF has also been applied successfully in the study of a wide range of inorganic semiconductors, including:

- Microcrystalline silicon [7.83, 97]
- Crystalline CdTe and CdZnTe [7.98]
- Single-crystal diamond [7.99]
- Polycrystalline CIGS [7.100]
- Sillenite ($Bi_{12}TiO_{20}$) [7.101].

7.7.2 Interrupted Field TOF (IFTOF)

The IFTOF experiment pioneered by *Kasap* et al. [7.102] differs from the time-of-flight experiment described in the previous section, in that the applied field that drives the photogenerated carrier packet through the sample is turned off for some period of time before the carriers have completed their transit. As illustrated in Fig. 7.20, a lower current intensity is measured when the field is turned on again, signaling that some carriers have become immobilized in deep traps. The fall in current versus interruption time t_i, enables the deep-trapping lifetime of the carriers to be evaluated. By *precharging* a sample with carriers of one polarity

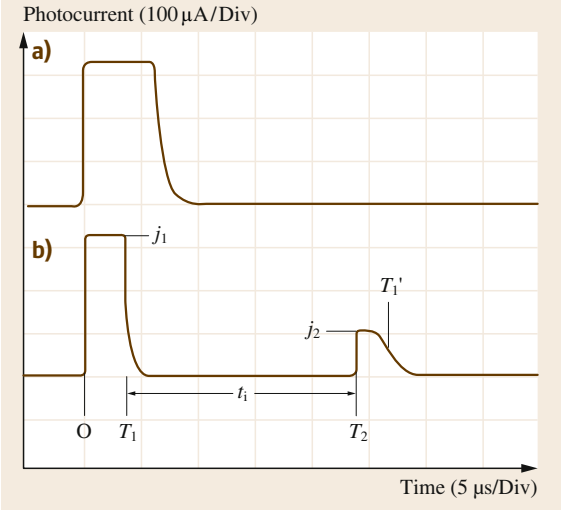


Fig. 7.20a,b Comparison of current traces obtained in **(a)** TOF, and **(b)** IFTOF experiments. The electric field is turned off in **(b)** for a length of time t_i. (After [7.102])

before performing an IFTOF experiment that drifts carriers of the opposite polarity through the sample, recombination parameters can also be studied [7.103].

Another modification involves generating free carriers through one contact and drifting the slower carrier into the sample, followed by a second pulse through the other contact, which sends faster carrier in the opposite direction. The two packets will cross, and some will recombine, affecting the observed currents and providing a way to study recombination. An example of this technique applied to amorphous selenium is given in *Haugen* and *Kasap* [7.104]. a-Se was one of the earliest amorphous semiconductors to be studied in detail, driven by its application at that time in xerography. While selenium is no longer used in this role, it remains an important x-ray photoconductor in flat-panel medical x-ray imaging (Chap. 45). A combination of TOF and IFTOF has been used to study the relaxation of arsenic-stabilized a-Se, following deposition, annealing and cooling to room temperature [7.105], and the influence of the stabilizer on the properties of oxygen and chlorine doped a-Se [7.106]. Both electron and hole mobilities may be studied independently in this way.

7.8 Other Photoconductivity-Related Techniques

This chapter has for the most part addressed how photoconductivity may be used as a tool in the detection, identification and measurement of electronic and/or optical properties of photoconductors. We conclude with brief descriptions of three techniques, related directly to photoconductivity, that do not feature elsewhere in this *handbook* but which readers may find relevant to their fields of interest.

7.8.1 Surface Photovoltage (SPV)

This technique involves the detection of changes in contact potential between a semiconductor and an electrode (Kelvin probe) placed close to its surface, under varying illumination conditions. A space-charge region forms in the bulk close to the surface, balanced by charge residing in surface states. Light incident on the semiconductor generates free charge, which redistributes under the built-in field. The steady-state charge distribution is sensitive to the semiconductor ambipolar diffusion length L, which in turn is reflected in the SPV. Details may be found in [7.107].

Measurement of L involves stepping the wavelength of the incident light, and adjusting its intensity after each step so as to maintain a constant SPV. Each wavelength corresponds to a specific absorption depth $1/\alpha$ in the semiconductor, and the value of L is obtained as the intercept of a plot of light intensity versus absorption depth. SPV is a popular technique for measuring L, particularly in bulk materials, although inaccuracies are claimed to arise when used on thin films.

SPV is versatile and may be configured to measure surface state density, and charge separation lengths in materials such as molecular clusters and quantum dots. It also forms the basis of a surface potential microscopy (Kelvin probe force microscopy) [7.108] when combined with an AFM system. The electrostatic force between tip and semiconductor is adjusted to zero by application of a bias potential, whose value is plotted versus position to yield a surface map.

7.8.2 Spin-Dependent Recombination

As was discussed in Sect. 7.1, recombination is influenced both by material properties and by external parameters. A further external variable, alongside temperature and light intensity, is afforded by the possibility of altering recombination kinetics by means of spin flipping. For a recombination event to occur, both electron and hole must form a singlet spin state. Recombination via the more probable triplet state is excluded, but may be enabled by flipping one of the spins by spin resonance. Thus by shining light on a photoconductive sample and scanning in an electron spin resonance (ESR) system, the resonance condition may be detected as a decrease in photoconductivity [7.109]. This detection method, termed electrically detected magnetic resonance or EDMR, is is many orders of magnitude more sensitive than a standard ESR experiment, which measures absorbed power using a microwave bridge. When combined with measured g-values and conductivity changes, spin-dependent photoconductivity provides detailed information on recombination pathways, and may be applied to devices such as solar cells and LEDs as well as to material samples.

7.8.3 Time-Resolved Microwave Conductivity (TRMC)

TRMC consists of measuring changes in microwave reflectivity of a photoconductor when excess photogenerated carriers are created by a short flash of light. The reflectivity is proportional to the free carrier density,

yielding a transient decay over many orders of magnitude of time similar to the TPC method. A principal advantage is that it is a contactless technique, not requiring deposition of electrodes on to the sample, and may thus be used with powders, aggregates or nanoporous materials difficult or impossible to measure using TPC. This has led to the adoption of TRMC when working with large surface area materials such as TiO_2 [7.110], used in photocatalysis and dye-sensitized solar cells where surface interactions are of primary importance.

Acknowledgments. The authors are grateful to the Engineering and Physical Sciences Research Council, and the Fonds voor Wetenschappelijk Onderzoek – Vlaanderen for financial support of parts of this work. Particular thanks are due to Dr. C. Main and Dr. A.C. Reynolds, for their critical reading of the manuscript and constructive comments. Finally, we are indebted to colleagues and students past and present, and to the numerous academic visitors to the Dundee and Leuven labs, from whom we have learned much and share many fond memories.

References

7.1 R.H. Bube: *Photoconductivity of Solids* (Wiley, New York 1960)
7.2 R.H. Bube: *Photoelectronic Properties of Semiconductors* (Cambridge Univ. Press, Cambridge 1992)
7.3 S.M. Ryvkin: *Photoelectric Effects in Semiconductors* (Consultants Bureau, New York 1964)
7.4 A. Rose: *Concepts in Photoconductivity and Allied Problems* (Krieger, Huntington 1978)
7.5 C. Main, J. Berkin, A. Merazga: Photoconductivity in amorphous semiconductors – Experiment and computer modelling. In: *Proceedings of the Sixth International School of Condensed Matter Physics, Varna*, ed. by M. Borissov, N. Kirov, J.M. Marshall, A. Vavrék (World Scientific, Singapore 1991) p. 55
7.6 C. Main, R. Brüggemann: Electronic and optoelectronic materials for the 21st century. In: *Proceedings of the Seventh International School of Condensed Matter Physics, Varna*, ed. by N. Kirov, A. Vavrék (World Scientific, Singapore 1993) p. 270
7.7 R. Brüggemann, C. Main: In: *Proceedings of the Seventh International School of Condensed Matter Physics, Varna*, ed. by J.M. Marshall, N. Kirov, A. Viavrék (World Scientific, Singapore 1993) p. 279
7.8 D. Abou-Ras, T. Kirchartz, U. Rau (Eds.): *Advanced Characterization Techniques for Thin Film Solar Cells*, 2nd edn. (Wiley-VCH, Weinheim 2016)
7.9 The SC-Simul software is available from R. Brüggemann, University of Oldenburg, and may be downloaded at: http://www.physik.uni-oldenburg.de/greco/download.html (last accessed Jan 2016)
7.10 C. Main, A.E. Owen: In: *Electronic and Structural Properties of Amorphous Semiconductors*, ed. by P.G. Le Comber, J. Mort (Academic, London 1973) p. 527
7.11 J.G. Simmons, G.W. Taylor: J. Phys. C **7**, 3051 (1974)
7.12 G.J. Adriaenssens: Philos. Mag. B **62**, 79 (1990)
7.13 G.J. Adriaenssens, N. Qamhieh: J. Mater. Sci. Mater. Electron. **14**, 605 (2003)
7.14 R. Brüggemann: J. Appl. Phys. **92**, 2540 (2002)
7.15 M. Nesládek, L.M. Stals, A. Stesmans, K. Iakoubovskij, G.J. Adriaenssens, J. Rosa, M. Vaněček: Appl. Phys. Lett. **72**, 3306 (1998)
7.16 C. Deibel, D. Mack, J. Gorenflot, A. Schöll, S. Krause, F. Reinert, D. Rauh, V. Dyakonov: Phys. Rev. B **81**, 085202 (2010)
7.17 C. Main, F. Dick, S. Reynolds, W. Gao, R.A.G. Gibson: J. Non-Cryst. Solids **198**, 263 (1996)
7.18 H. Fritzsche, B.-G. Yoon, D.-Z. Chi, M.Q. Tran: J. Non-Cryst. Solids **141**, 123 (1992)
7.19 R.H. Bube: J. Appl. Phys. **74**, 5138 (1993)
7.20 M. Vaněček, J. Kočka, J. Stuchlík, A. Třiska: Solid State Commun. **39**, 1199 (1981)
7.21 M. Vaněček, J. Kočka, A. Poruba, A. Fejfar: J. Appl. Phys. **78**, 6203 (1995)
7.22 M. Sasaki, S. Okamoto, Y. Hishikawa, S. Tsuda, S. Nakano: Sol. Energy Mater. Sol. Cells **34**, 541 (1994)
7.23 D. Ritter, K. Weiser: Optics Commun. **57**, 336 (1986)
7.24 K. Pierz, H. Mell, J. Terjukov: J. Non-Cryst. Solids **77**, 547 (1985)
7.25 P. Jensen: Solid State Commun. **76**, 1301 (1990)
7.26 C. Main, S. Reynolds, I. Zrinscak, A. Merazga: J. Non-Cryst. Solids **338**, 228 (2004)
7.27 A. Bacioglu, A.O. Kodolbas, O. Oktu: Sol. Energy Mater. Sol. Cells **89**, 49 (2005)
7.28 J. Willekens, M. Brinza, T. Aernouts, J. Poortmans, G.J. Adriaenssens: J. Non-Cryst. Solids **352**, 1675 (2006)
7.29 K. Sharma, A.S. Al-Kabbi, G.S.S. Saini, S.K. Tripathi: Curr. Appl. Phys. **13**, 964 (2013)
7.30 N. Ishihara, M. Tsubuku, Y. Nonaka, R. Watanabe, K. Inoue, H. Shishido, K. Kato, S. Yamazaki: Optical properties and evaluation of localized level in gap of In-Ga-Zn-O thin film, Proc. 19th Int. Workshop Active-Matrix Flatpanel Disp. Devices (2012) p. 143
7.31 C. Wronski, B. Abeles, T. Tiedje, G.D. Cody: Solid State Commun. **44**, 1423 (1982)
7.32 S. Lee, S. Kumar, C.R. Wronski: J. Non-Cryst. Solids **114**, 316 (1989)
7.33 M. Günes, C. Wronski, T.J. McMahon: J. Appl. Phys. **76**, 2260 (1994)
7.34 Z. Remes, T. Izak, A. Kromka, M. Vanecek: Diam. Relat. Mater. **19**, 453 (2010)
7.35 M. Vaněček, A. Poruba: Thin Solid Films **515**, 7499 (2007)
7.36 T. Inushima, M.H. Brodsky, J. Kanicki, R.J. Serino: AIP Conf. Proc. **120**, 24 (1984)
7.37 M. Vaněček, A. Poruba: Appl. Phys. Lett. **80**, 719 (2002)

7.38 J. Holovsky: Fourier transform photocurrent spectroscopy on non-crystalline semiconductors. In: *Fourier Transforms – New Analytical Approaches and FTIR Strategies*, ed. by G. Nikolic (Intech, Rijeka 2011)
7.39 J. Melskens, M. Schouten, R. Santbergen, M. Fischer, R. Vasudevan, D.J. van der Vlies, R.J.V. Quax, S.G.M. Heirman, K. Jäger, V. Demontis, M. Zeman, A.H.M. Smets: Sol. Energy Mater. Sol. Cells **129**, 70 (2014)
7.40 D. Bozyigit, S. Volk, O. Yarema, V. Wood: Nano Lett. **13**, 5284 (2013)
7.41 D. Ritter, E. Zeldov, K. Weiser: Appl. Phys. Lett. **49**, 791 (1986)
7.42 R. Brüggemann: Steady-state photocarrier grating method. In: *Advanced Characterization Techniques for Thin Film Solar Cells*, 2nd edn., ed. by D. Abou-Ras, T. Kirchartz, U. Rau (Wiley-VCH, Weinheim 2016)
7.43 C.D. Abel, G.H. Bauer: Prog. Photovolt. **1**, 269 (1993)
7.44 Y.M. Soro, A. Abramov, M.E. Gueunier-Farret, E.V. Johnson, C. Longeaud, P. Roca i Cabarrocas, J.P. Kleider: J. Non-Cryst. Solids **354**, 2092 (2008)
7.45 M. Goerlitzer, N. Beck, P. Torres, J. Meier, N. Wyrsch, A. Shah: J. Appl. Phys. **80**, 5111 (1996)
7.46 M. Niehus, R. Schwarz: Superlattices Microstruct. **40**, 350 (2006)
7.47 R. Menner, S. Zweigert, R. Klenk, H.W. Schock: Jpn. J. Appl. Phys. Part 1 **32**, 45 (1993)
7.48 A. FathAllah, F. Ventosinos, C. Longeaud: J. Phys. Conf. Ser. **558**, 012011 (2014)
7.49 C. Longeaud: Rev. Sci. Instrum. **84**, 055101 (2013)
7.50 J.A. Schmidt, C. Longeaud: Phys. Rev. B **71**, 125208 (2005)
7.51 C. Longeaud, F. Ventosinos, J.A. Schmidt: J. Appl. Phys. **112**, 023709 (2012)
7.52 H. Oheda: J. Appl. Phys. **52**, 6693 (1981)
7.53 J.P. Kleider, C. Longeaud, M.E. Gueunier: J. Non-Cryst. Solids **338**, 390 (2004)
7.54 R. Brüggemann, C. Main, J. Berkin, S. Reynolds: Philos. Mag. B **62**, 29 (1990)
7.55 K. Hattori, Y. Adachi, M. Anzai, H. Okamoto, Y. Hamakawa: J. Appl. Phys. **76**, 2841 (1994)
7.56 R.R. Koropecki, J.A. Schmidt, R. Arce: J. Appl. Phys. **91**, 8965 (2002)
7.57 J.A. Schmidt, C. Longeaud, R.R. Koropecki, R. Arce, J.P. Kleider: J. Non-Cryst. Solids **354**, 2914 (2008)
7.58 H. Naito, M. Nakaishi, T. Matsushita, M. Okuda: J. Non-Cryst. Solids **77**, 1183 (1985)
7.59 C. Main, D.P. Webb, R. Brüggemann, S. Reynolds: J. Non-Cryst. Solids **137**, 951 (1991)
7.60 J.P. Kleider, C. Longeaud, O. Glodt: J. Non-Cryst. Solids **137**, 447 (1991)
7.61 J.D. Cohen: Sol. Energy Mater. Sol. Cells **78**, 399 (2003)
7.62 C. Longeaud, D. Roy, O. Saadane: Phys. Rev. B **65**, 85206 (2002)
7.63 S. Gorgolis, A. Giannopoulou, P. Kounavis: J. Appl. Phys. **113**, 123102 (2013)
7.64 A. Krysztopa, M. Igalson, P. Zabierowski, J. Larsen, Y. Aida, S. Siebentritt, L. Gütay: Thin Solid Films **519**, 7308 (2011)
7.65 J. Luckas, S. Kremers, D. Krebs, M. Salinga, M. Wuttig, C. Longeaud: J. Appl. Phys. **110**, 013719 (2011)
7.66 P.T. Erslev, E.S. Sundholm, R.E. Presley, D. Hong, J.F. Wager, J.D. Cohen: Appl. Phys. Lett. **95**, 192115 (2009)
7.67 M.A. Kastner, D. Monroe: Sol. Energy Mater. **8**, 41 (1982)
7.68 R. Street: Solid St. Commun. **39**, 263 (1981)
7.69 T. Tiedje, A. Rose: Solid State Commun. **37**, 49 (1981)
7.70 J. Orenstein, M. Kastner: Phys. Rev. Lett. **46**, 1421 (1981)
7.71 C. Main: Photoconductivity and Noise in Amorphous Chalcogenide Semiconductors, Ph.D. Thesis (Univ. Edinburgh, Edinburgh 1973)
7.72 C. Main, S. Reynolds, R. Brüggemann: Phys. Status Solidi (c) **1**, 1194 (2004)
7.73 J.C. Balland, J.P. Zielinger, C. Noguet, M. Tapiero: J. Phys. D Appl. Phys. **19**, 57 (1986)
7.74 G.J. Adriaenssens, H. Michiel: AIP Conf. Proc. **120**, 110 (1984)
7.75 J.M. Marshall: Rep. Prog. Phys. **46**, 1235 (1983)
7.76 C. Main: MRS Symp. Proc. Ser. **467**, 167 (1997)
7.77 C. Main: J. Non-Cryst. Solids **299**, 525 (2002)
7.78 C. Main, R. Brüggemann, D.P. Webb, S. Reynolds: Solid State Commun. **83**, 401 (1992)
7.79 H. Naito, J. Ding, M. Okuda: Appl. Phys. Lett. **64**, 1830 (1994)
7.80 T. Nagase, K. Kishimoto, H. Naito: J. Appl. Phys. **86**, 5026 (1999)
7.81 S. Reynolds, C. Main, D.P. Webb, M.J. Rose: Philos. Mag. B **80**, 547 (2000)
7.82 C. Main, R. Russell, J. Berkin, J.M. Marshall: Philos. Mag. Lett. **55**, 189 (1987)
7.83 S. Reynolds: MRS Symp. Proc. Ser. **910**, 3 (2007)
7.84 R. Brüggemann, C. Main, S. Reynolds: J. Phys. Condens. Matter **14**, 6909 (2002)
7.85 S. Reynolds, J.T. Shepherd, C. Main, J.M. Marshall, J.M. Maud: J. Non-Cryst. Solids **266**, 994 (2000)
7.86 J. Hao, S.A. Studenikin, M. Cocivera: J. Appl. Phys. **90**, 5064 (2001)
7.87 M.J. Gueorguieva, C. Main, S. Reynolds, R. Brüggemann, C. Longeaud: J. Non-Cryst. Solids **299**, 541 (2002)
7.88 Y. Tamada, Y. Kanemitsu: Appl. Phys. Lett. **101**, 133907 (2012)
7.89 W.E. Spear: J. Non-Cryst. Solids **1**, 197 (1969)
7.90 T. Bronger: Time-of-flight analysis. In: *Advanced Characterization Techniques for Thin Film Solar Cells*, 2nd edn., ed. by D. Abou-Ras, T. Kirchartz, U. Rau (Wiley-VCH, Weinheim 2016)
7.91 D. Hertel, A. Ochse, V.I. Arkhipov, H. Bässler: J. Imaging Sci. Technol. **43**, 220 (1999)
7.92 M. Brinza, E.V. Emelianova, G.J. Adriaenssens: Phys. Rev. B **71**, 115209 (2005)
7.93 S. Dinca, G. Ganguly, E.A. Schiff, V. Vlahos, C.R. Wronski, Q. Yuan: Mater. Res. Soc. Symp. Proc. **762**, A7.1 (2003)
7.94 G.F. Seynhaeve, R.P. Barclay, G.J. Adriaenssens, J.M. Marshall: Phys. Rev. B **39**, 10196 (1989)
7.95 C. Main, S. Reynolds, R.I. Badran, J.M. Marshall: J. Appl. Phys. **88**, 1190 (2000)

7.96 A. Köhler, H. Bässler: *Electronic Processes in Organic Semiconductors, an Introduction* (Wiley, Weinheim 2015)
7.97 T. Dylla, F. Finger, E.A. Schiff: Appl. Phys. Lett. **87**, 032103 (2005)
7.98 K. Suzuki, S. Seto, T. Sawada, K. Imai: IEEE Trans. Nucl. Sci. **49**, 1287 (2002)
7.99 M. Nesladek, A. Bogdan, W. Deferme, N. Tranchant, P. Bergonzo: Diam. Relat. Mater. **17**, 1235 (2008)
7.100 S.A. Dinca, E.A. Schiff, W.N. Shafarman, B. Egaas, R. Noufi, D.L. Young: Appl. Phys. Lett. **100**, 103901 (2012)
7.101 C. Longeaud, H. Belgacem, C. Douay: J. Phys. Condens. Matter **19**, 476202 (2007)
7.102 S. Kasap, B. Polishuk, D. Dodds: Rev. Sci. Instrum. **61**, 2080 (1990)
7.103 S. Kasap, B. Fogal, M.Z. Kabir, R.E. Johanson, S.K. O'Leary: Appl. Phys. Lett. **84**, 1991 (2004)
7.104 C. Haugen, S.O. Kasap: Philos. Mag. B **71**, 91 (1995)
7.105 C. Allen, G. Belev, R. Johanson, S.O. Kasap: J. Non-Cryst. Solids **354**, 2711 (2008)
7.106 G. Belev, D. Tonchev, B. Fogal, C. Allen, S.O. Kasap: J. Phys. Chem. Solids **68**, 972 (2007)
7.107 L. Kronik, Y. Shapira: Surf. Sci. Rep. **37**, 1 (1999)
7.108 W. Melitz, J. Shen, A.C. Kummel, S. Lee: Surf. Sci. Rep. **66**, 1 (2011)
7.109 M. Stutzmann, M.S. Brandt, M.W. Bayerl: J. Non-Cryst. Solids **266**, 1 (2000)
7.110 C. Colbeau-Justin, M. Kunst, D. Huguenin: J. Mater. Sci. **38**, 2429 (2003)

8. Electronic Properties of Semiconductor Interfaces

Winfried Mönch

In this chapter, we discuss electronic properties of semiconductor interfaces. Semiconductor devices contain metal–semiconductor, insulator–semiconductor, insulator–metal, and/or semiconductor–semiconductor interfaces. The electronic properties of these interfaces determine characteristics of the device. The band structure lineup at all these interfaces is determined by one unifying concept, the continuum of interface-induced gap states (IFIGS). These intrinsic interface states are the wave function tails of electron states that overlap the fundamental band gap of a semiconductor at the interface; in other words, they are caused by the quantum mechanical tunneling effect. IFIGS theory quantitatively explains the experimental barrier heights of well-characterized metal–semiconductor or Schottky contacts as well as the valence-band offsets of semiconductor–semiconductor interfaces or semiconductor heterostructures. Insulators are viewed as semiconductors with wide band gaps.

8.1 **Experimental Database** 177
8.1.1 Barrier Heights of Laterally Homogeneous Schottky Contacts 177
8.1.2 Band Offsets of Semiconductor Heterostructures 180
8.2 **IFIGS-and-Electronegativity Theory** 181
8.3 **Comparison of Experiment and Theory** 183
8.3.1 Barrier Heights of Schottky Contacts 183
8.3.2 Band Offsets of Semiconductor Heterostructures 184
8.4 **Modifications of Schottky Contacts** 187
8.4.1 Nonalloyed Ohmic Contacts or MUTIS Schottky Contacts 187
8.4.2 Atomic Interlayers 189
8.5 **Graphene Schottky Contacts** 190
8.6 **Final Remarks** 190
References 191

In his pioneering article entitled *Semiconductor Theory of the Blocking Layer*, *Schottky* [8.1] finally explained the rectifying properties of metal–semiconductor contacts, which had first been described by *Braun* [8.2], as being due to a depletion of the majority carriers on the semiconductor side of the interface. This new depletion-layer concept immediately triggered a search for a physical explanation of the barrier heights observed in metal–semiconductor interfaces, or Schottky contacts as they are also called in order to honor Schottky's many basic contributions to this field.

The early *Schottky–Mott* rule [8.3, 4] proposed that n-type (p-type) barrier heights were equal to the difference between the work function of the metal and the electron affinity (ionization energy) of the semiconductor. A plot of the experimental barrier heights of various metal–selenium rectifiers versus the work functions of the corresponding metals did indeed reveal a linear correlation, but the slope parameter was much smaller than unity [8.4]. To resolve the failure of the very simple and therefore attractive Schottky–Mott rule, *Bardeen* [8.5] proposed that electronic interface states in the semiconductor band gap play an essential role in the charge balance at metal–semiconductor interfaces.

Heine [8.6] considered the quantum-mechanical tunneling effect at metal–semiconductor interfaces and noted that *for energies in the semiconductor band gap, the volume states of the metal have tails in the semiconductor*. *Tejedor* and *Flores* [8.7] applied this same idea to semiconductor heterostructures where, for energies in the band-edge discontinuities, the volume states of one semiconductor tunnel into the other. The continua of interface-induced gap states (IFIGS), as these evanescent states were later called, are an intrinsic property of semiconductors and they are the *fundamental* physical mechanism that determines the band-structure lineup at both metal–semiconductor contacts and semiconductor heterostructures: in other words, at all semiconductor interfaces. Insulator interfaces are also included in this, since insulators may be de-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_8

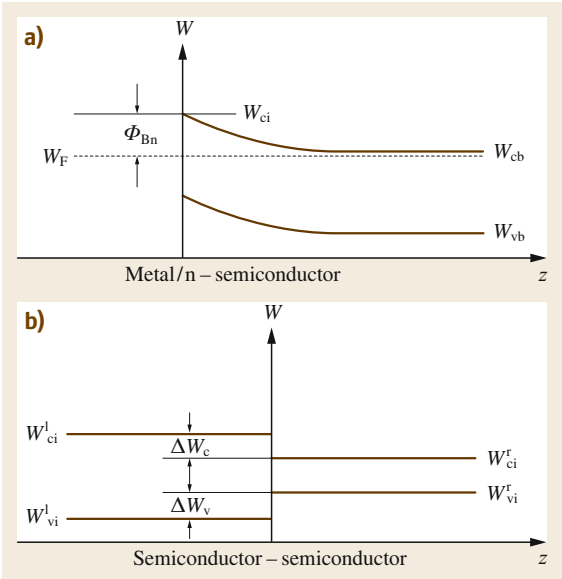


Fig. 8.1a,b Schematic energy-band diagrams of **(a)** metal–semiconductor contacts and **(b)** semiconductor heterostructures. W_F, Fermi level; Φ_{Bn}, barrier height; W_v and W_c, valence-band maximum and conduction-band minimum, respectively; ΔW_v and ΔW_c, valence- and conduction-band offset, respectively; i and b, values at the interface and in the bulk, respectively; r and l, right and left side, respectively

scribed as wide-gap semiconductors. Figure 8.1 shows schematic band diagrams of an n-type Schottky contact and a semiconductor heterostructure.

The IFIGS continua derive from both the valence- and the conduction-band states of the semiconductor. The energy at which their predominant character changes from valence-band-like to conduction-band-like is called their branch point. The position of the Fermi level relative to this branch point then determines the sign and the amount of the net charge in the IFIGS. Hence, the IFIGS give rise to intrinsic interface dipoles. Both the barrier heights of Schottky contacts and the band offsets of heterostructures thus divide up into a zero-charge-transfer term and an electric-dipole contribution.

From a more chemical point of view, these interface dipoles may be attributed to the partial ionic character of the covalent bonds between atoms right at the interface. Generalizing *Pauling*'s [8.8] electronegativity concept, the difference in the electronegativities of the atoms involved in the interfacial bonds also describes the charge transfer at semiconductor interfaces. Combining the physical IFIGS and the chemical electronegativity concept, the electric-dipole contributions of Schottky barrier heights as well as those of heterostructure band offsets vary proportional to the difference in the electronegativities of the metal and the semiconductor and of the two semiconductors, respectively.

The theoreticians appreciated Heine's IFIGS concept at once. The initial reluctance of most experimentalists was motivated by the observation that the predictions of the IFIGS theory only marked upper limits for the barrier heights observed with *real* Schottky contacts [8.9]. *Schmitsdorf* et al. [8.10] finally resolved this dilemma. They found a linear decrease in the effective barrier heights with increasing ideality factors for their Ag/n-Si(111) diodes. Such behavior has been observed for all of the Schottky contacts investigated so far. *Schmitsdorf* et al. attributed this correlation to patches of decreased barrier heights and lateral dimensions smaller than the depletion layer width [8.10, 11]. Consequently, they extrapolated their plots of effective barrier height versus ideality factor to the ideality factor determined by the image-force or *Schottky* effect [8.12] alone; in this way, they obtained the barrier heights of the laterally homogeneous contacts. The barrier heights of laterally uniform contacts can also be determined from capacitance–voltage measurements (C/V) and by applying ballistic-electron-emission microscopy (BEEM) and internal photoemission yield spectroscopy (IPEYS). The I/V, C/V, BEEM, and IPEYS data agree within the margins of experimental error.

Mönch [8.13] found that the barrier heights of laterally homogeneous Schottky contacts as well as the experimentally observed valence-band offsets of semiconductor heterostructures agree excellently with the predictions of the IFIGS-and-electronegativity theory.

8.1 Experimental Database

8.1.1 Barrier Heights of Laterally Homogeneous Schottky Contacts

I/V Characteristics

The current transport in real Schottky contacts occurs via thermionic emission over the barrier provided the doping level of the semiconductor is not too high [8.14]. For doping levels larger than approximately $10^{18}\,\mathrm{cm}^{-3}$, the depletion layer becomes so narrow that tunnel or field emission through the depletion layer prevails. The current–voltage characteristics then become ohmic rather than rectifying.

For thermionic emission over the barrier, the current–voltage characteristics of n-type Schottky contacts may be written as [8.13]

$$I_{\mathrm{te}} = I_{\mathrm{s}} \exp\left(\frac{e_0 V_{\mathrm{c}}}{n k_{\mathrm{B}} T}\right)\left(1 - \exp\left(\frac{-e_0 V_{\mathrm{c}}}{k_{\mathrm{B}} T}\right)\right) , \quad (8.1)$$

where

$$I_{\mathrm{s}} = A A_{\mathrm{R}}^{*} T^2 \exp\left(\frac{-\Phi_{\mathrm{Bn}}^{\mathrm{eff}}}{k_{\mathrm{B}} T}\right) \quad (8.2)$$

is the saturation current and A is the diode area, A_{R}^{*} is the effective Richardson constant of the semiconductor, and k_{B}, T, and e_0 are Boltzmann's constant, the temperature, and the electronic charge, respectively. The effective Richardson constant is defined as

$$A_{\mathrm{R}}^{*} = \frac{4\pi e_0 k_{\mathrm{B}} m_{\mathrm{n}}^{*}}{h^3} = A_{\mathrm{R}} \frac{m_{\mathrm{n}}^{*}}{m_0} , \quad (8.3)$$

where $A_{\mathrm{R}} = 120\,\mathrm{A\,cm^{-2}\,K^{-2}}$ is the Richardson constant for thermionic emission of nearly free electrons into vacuum, h is Planck's constant, and m_0 and m_{n}^{*} are the vacuum and the effective conduction-band mass of electrons, respectively. The externally applied bias V_{a} divides up into a voltage drop V_{c} across the depletion layer of the Schottky contact and an IR drop at the series resistance R_{s} of the diode, so that $V_{\mathrm{c}} = V_{\mathrm{a}} - IR_{\mathrm{s}}$. For *ideal* (intimate, abrupt, defect-free, and, above all, laterally homogeneous) Schottky contacts, the effective zero-bias barrier height $\Phi_{\mathrm{Bn}}^{\mathrm{eff}}$ equals the difference $\Phi_{\mathrm{Bn}}^{\mathrm{hom}} - \delta\Phi_{\mathrm{if}}^{0}$ between the homogeneous barrier height and the zero-bias image-force lowering [8.13]

$$\delta\Phi_{\mathrm{if}}^{0} = e_0 \left[\frac{2 e_0^2 N_{\mathrm{d}}}{(4\pi)^2 \varepsilon_{\infty}^2 \varepsilon_{\mathrm{b}} \varepsilon_0^3}\left(e_0 \left|V_{\mathrm{i}}^{0}\right| - k_{\mathrm{B}} T\right)\right]^{1/4} , \quad (8.4)$$

where N_{d} is the donor density, $e_0|V_{\mathrm{i}}^{0}|$ is the zero-bias band bending, ε_{∞} and ε_{b} are the optical and the bulk dielectric constant, respectively, and ε_0 is the permittivity of vacuum. The ideality factor n describes the voltage dependence of the barrier height and is defined by

$$1 - \frac{1}{n} = \frac{\partial \Phi_{\mathrm{Bn}}^{\mathrm{eff}}}{\partial e_0 V_{\mathrm{c}}} . \quad (8.5)$$

For *real* diodes, the ideality factors n are generally found to be larger than the ideality factor

$$n_{\mathrm{if}} = \left(1 - \frac{\delta\Phi_{\mathrm{if}}^{0}}{4 e_0 |V_{\mathrm{i}}^{0}|}\right)^{-1} , \quad (8.6)$$

which is determined by the image-force effect only.

The effective barrier heights and the ideality factors of real Schottky diodes fabricated under experimentally identical conditions vary from one specimen to the next. However, the variations of both quantities are correlated, and the $\Phi_{\mathrm{Bn}}^{\mathrm{eff}}$ values become smaller as the ideality factors increase. As an example, Fig. 8.2 displays $\Phi_{\mathrm{Bn}}^{\mathrm{eff}}$ versus n data for Ag/n-Si(111) contacts with $(1\times1)^{\mathrm{i}}$-unreconstructed and $(7\times7)^{\mathrm{i}}$-reconstructed interfaces [8.10]. The dashed and dash-dotted lines are the linear least-squares fits to the data points. The linear dependence of the effective barrier height on the ideality factor may be written as

$$\Phi_{\mathrm{Bn}}^{\mathrm{eff}} = \Phi_{\mathrm{Bn}}^{\mathrm{nif}} - \varphi_{\mathrm{p}}(n - n_{\mathrm{if}}) , \quad (8.7)$$

where $\Phi_{\mathrm{Bn}}^{\mathrm{nif}}$ is the barrier height at the ideality factor n_{if}. Several conclusions may be drawn from this relation. First, the $\Phi_{\mathrm{Bn}}^{\mathrm{eff}} - n$ correlation shows that more than one physical mechanism determines the barrier heights of *real* Schottky contacts. Second, the extrapolation of $\Phi_{\mathrm{Bn}}^{\mathrm{eff}}$ versus n curves to n_{if} removes all mechanisms that

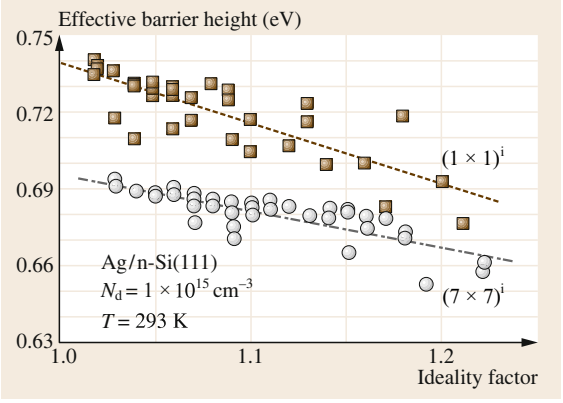


Fig. 8.2 Effective barrier heights versus ideality factors determined from I/V characteristics of Ag/n-Si(111)-$(7\times7)^{\mathrm{i}}$ and -$(1\times1)^{\mathrm{i}}$ contacts at room temperature. The *dashed* and *dash-dotted* lines are the linear least-squares fits to the data. (After [8.10])

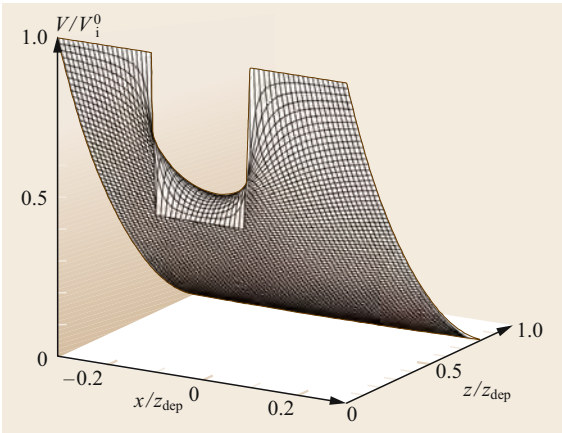


Fig. 8.3 Calculated potential distribution underneath and around a patch of reduced interface potential embedded in a region of larger interface band-bending. The lateral dimension and the interface potential reduction of the patch are set to two-tenths of the depletion layer width z_{dep} and one-half of the interface potential of the surrounding region. (After [8.13])

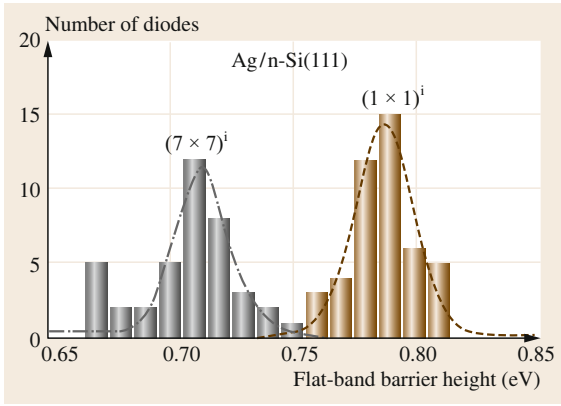


Fig. 8.4 Histograms of flat-band barrier heights determined from C/V characteristics of Ag/n-Si(111)-$(7 \times 7)^i$ and -$(1 \times 1)^i$ contacts at room temperature. The data were obtained with the same diodes discussed in Fig. 8.2. The *dashed* and *dash-dotted lines* are the Gaussian least-squares fits to the data. (After [8.10])

cause a larger bias dependence of the barrier height than the image-force effect itself from consideration. Third, the extrapolated barrier heights Φ_{Bn}^{nif} are equal to the zero-bias barrier height $\Phi_{Bn}^{hom} - \delta\Phi_{if}^{0}$ of the laterally homogeneous contact.

The laterally homogeneous barrier heights obtained from Φ_{Bn}^{eff} versus n curves to n_{if} are not necessarily characteristic of the corresponding *ideal* contacts. This is illustrated by the two data sets displayed in Fig. 8.2, which differ in the interface structures of the respective diodes. Quite generally, structural rearrangements such as the $(7 \times 7)^i$ reconstruction are connected with a redistribution of the valence charge. The bonds in perfectly ordered bulk silicon, the example considered here, are purely covalent, and so reconstructions are accompanied by electric $Si^{+\Delta q} - Si^{-\Delta q}$ dipoles. The Si(111)-(7×7) reconstruction is characterized by a stacking fault in one half of its unit mesh [8.15]. *Schmitsdorf* et al. [8.10] quantitatively explained the experimentally observed reduction in the laterally homogeneous barrier height of the $(7 \times 7)^i$ with regard to the $(1 \times 1)^i$ diodes by the electric dipole associated with the stacking fault of the Si(111)-(7×7) reconstruction.

Patches of reduced barrier height with lateral dimensions smaller than the depletion layer width that are embedded in large areas of laterally homogeneous barrier height is the only known model that explains a lowering of effective barrier heights with increasing ideality factors. In their phenomenological studies of such patchy Schottky contacts, *Freeouf* et al. [8.11] found that the potential distribution exhibits a saddle point in front of such nanometer-size patches of reduced barrier height. Figure 8.3 explains this behavior. The saddle-point barrier height strongly depends on the voltage drop V_c across the depletion layer. Freeouf et al. simulated the current transport in such patchy Schottky contacts and found a reduction in the effective barrier height and a correlated increase in the ideality factor as they reduced the lateral dimensions of the patches. However, they overlooked the fact that the barrier heights of the laterally homogeneous contacts may be obtained from Φ_{Bn}^{eff} versus n plots, by extrapolating to n_{if}.

It has to be mentioned that the effective barrier heights and the ideality factors of *real* Schottky contacts generally vary as a function of temperature. The so-called Richardson plots $\ln(I_s/T^2)$-versus-$1/T$ will neither result in correct values of effective barrier heights nor will the effective Richardson constants obtained equal the theoretical values because such graphs inherently assume the barrier heights to be independent of temperature [8.16–18] in contrast to what is experimentally observed with real Schottky contacts.

C/V Characteristics

Both the space charge and the width of the depletion layers at metal–semiconductor contacts vary as a function of the externally applied voltage. The space-charge theory gives the variation in the depletion layer capacitance per unit area as [8.13]

$$C_{dep} = \left(\frac{e_0^2 \varepsilon_b \varepsilon_0 N_d}{2} \left[e_0 \left(|V_i^0| - V_c \right) - k_B T \right] \right)^{1/2} . \quad (8.8)$$

The current through a Schottky diode biased in the reverse direction is small, so the *IR* drop due to the series resistance of the diode may be neglected. Consequently, the extrapolated intercepts on the abscissa of $1/C_{\mathrm{dep}}^2$ versus V_{a} plots give the band-bending $e_0|V_{\mathrm{i}}^0|$ at the interface, and together with the energy distance $W_{\mathrm{n}} = W_{\mathrm{F}} - W_{\mathrm{cb}}$ from the Fermi level to the conduction band minimum in the bulk, one obtains the flat-band barrier height $\Phi_{\mathrm{Bn}}^{\mathrm{fb}} \equiv \Phi_{\mathrm{Bn}}^{\mathrm{hom}} = e_0|V_{\mathrm{i}}^0| + W_{\mathrm{n}}$, which equals the laterally homogeneous barrier height of the contact.

As an example, Fig. 8.4 displays the flat-band barrier heights of the same Ag/n-Si(111) diodes that are discussed in Fig. 8.2. The dashed- and dash-dotted lines are the Gaussian least-squares fits to the data from the diodes with $(1\times1)^{\mathrm{i}}$ and $(7\times7)^{\mathrm{i}}$ interface structures, respectively. Within the margins of experimental error, the peak C/V values agree with the laterally homogeneous barrier heights obtained from the extrapolations of the I/V data shown in Fig. 8.2. These data clearly demonstrate that barrier heights characteristic of laterally homogeneous Schottky contacts can be only obtained from I/V or C/V data from many diodes fabricated under identical conditions rather than from a single diode. However, the effective barrier heights and the ideality factors vary as a function of the diode temperature. Hence, the effective barrier heights and ideality factors evaluated from the I/V characteristics for one and the same diode recorded at different temperatures are also suitable for determining the corresponding laterally homogeneous barrier height [8.13].

Ballistic-Electron-Emission Microscopy

In BEEM [8.19], a tip injects almost monoenergetic electrons into the metal film of a Schottky diode. These tunnel-injected electrons reach the semiconductor as ballistic electrons provided that they lose no energy on their way through the metal. Hence, the collector current I_{coll} is expected to set in when the ballistic electrons surpass the metal–semiconductor barrier; in other words, if the voltage V_{tip} applied between tip and metal film exceeds the *local* potential barrier $\Phi_{\mathrm{Bn}}^{\mathrm{loc}}(z)/e_0$. *Bell* and *Kaiser* [8.20] derived the square law

$$I_{\mathrm{coll}}(z) = R^* I_{\mathrm{tip}} \left(e_0 V_{\mathrm{tip}} - \Phi_{\mathrm{Bn}}^{\mathrm{loc}}(z)\right)^2 \qquad (8.9)$$

for the BEEM $I_{\mathrm{coll}}/V_{\mathrm{tip}}$ characteristics, where I_{tip} is the injected tunnel current. BEEM measures *local* barrier heights; most specifically, the saddle-point barrier heights in front of nanometer-sized patches rather than their lower barrier heights right at the interface.

BEEM is *the* experimental tool for measuring spatial variations in the barrier height on the nanometer-scale. The local barrier heights are determined by fitting relation (8.9) to measured $I_{\mathrm{coll}}/V_{\mathrm{tip}}$ characteristics recorded at successive tip positions along lateral line scans. Figure 8.5 displays histograms of the local BEEM barrier heights of two Pd/n-6*H*-SiC(0001) diodes [8.21]. The diodes differ in their ideality factors, 1.06 and 1.49, which are close to and much larger, respectively, than the value $n_{\mathrm{if}} = 1.01$ determined solely by the image-force effect. Obviously, the *nanometer-scale* BEEM histograms of the two diodes are identical, although their *macroscopic* ideality factors and therefore their patchinesses differ. Two important conclusions were drawn from these findings. First, these data suggest the existence of two different types of patches: intrinsic and extrinsic ones. The intrinsic patches might be correlated with the random distributions of the ionized donors and acceptors which cause nanometer-scale lateral fluctuations in the interface potential. A few gross interface defects of extrinsic origin, which escape BEEM observations, are then responsible for the variations in the ideality factors. Second, Gaussian least-squares fits to the histograms of the local BEEM barrier heights yield peak barrier heights of 1.27 ± 0.03 eV. Within the margins of experimental error, this value agrees with the laterally homogeneous value of 1.24 ± 0.09 eV, which was obtained by extrapolation of the linear least-squares fit to a $\Phi_{\mathrm{Bn}}^{\mathrm{eff}}$ versus n plot to n_{if}. The nanometer-scale BEEM histograms and the macroscopic I/V characteristics thus provide identical barrier heights of laterally homogeneous Schottky contacts.

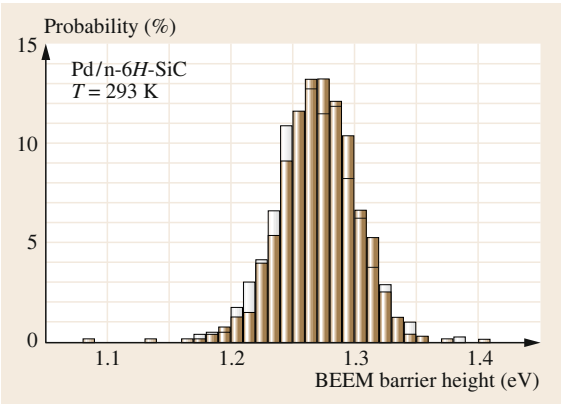


Fig. 8.5 Histograms of local BEEM barrier heights of two Pd/n-6*H*-SiC(0001) diodes with ideality factors of 1.06 (*gray solid bars*) and 1.49 (*empty bars*). The data were obtained by fitting the square law (8.9) to 800 BEEM $I_{\mathrm{coll}}/V_{\mathrm{tip}}$ spectra each. (After *Im* et al. [8.21])

Internal Photoemission Yield Spectroscopy

Metal–semiconductor contacts show a photoelectric response to optical radiation with photon energies smaller than the width of the bulk band gap. This effect is caused by photoexcitation of electrons from the metal over the

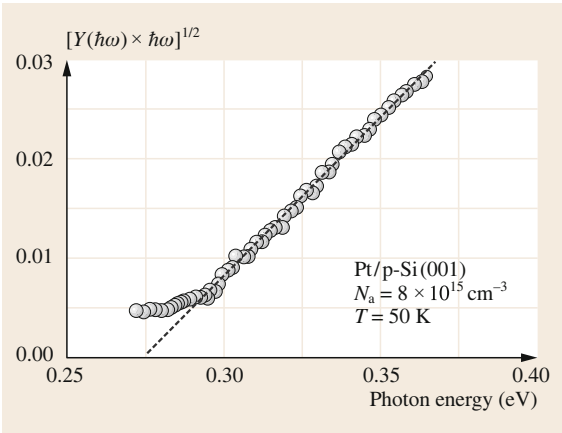


Fig. 8.6 Spectral dependence of the internal photoemission yield of a Pt/p-Si(001) diode versus the photon energy of the exciting light. The *dashed line* is the linear least-squares fit to the data for photon energies larger than 0.3 eV. (After *Turan* et al. [8.23])

interfacial barrier into the conduction band of the semiconductor. Experimentally, the internal photoemission yield, which is defined as the ratio of the photoinjected electron flux across the barrier into the semiconductor to the flux of the electrons excited in the metal, is measured as a function of the energy of the incident photons. Consequently, this technique is called internal photoemission yield spectroscopy (IPEYS). *Cohen* et al. [8.22] derived that the internal photoemission yield varies as a function of the photon energy $\hbar\omega$ as

$$Y(\hbar\omega) \propto \frac{1}{\hbar\omega}\left(\hbar\omega - \Phi_{\mathrm{Bn}}^{\mathrm{IPEYS}}\right)^2 . \qquad (8.10)$$

Patches only cover a small portion of the metal–semiconductor interface, so the threshold energy $\Phi_{\mathrm{Bn}}^{\mathrm{IPEYS}}$ will equal the barrier height $\Phi_{\mathrm{Bn}}^{\mathrm{hom}}$ of the laterally homogeneous part of the contact minus the zero-bias image-force lowering $\delta\Phi_{\mathrm{if}}^{0}$.

In Fig. 8.6, experimental $(Y(\hbar\omega)\times\hbar\omega)^{1/2}$ data for a Pt/p-Si(001) diode [8.23] are plotted versus the energy of the exciting photons. The dashed line is the linear least-squares fit to the data. The deviation of the experimental $(Y(\hbar\omega)\times\hbar\omega)^{1/2}$ data towards larger values slightly below and above the threshold is caused by the shape of the Fermi–Dirac distribution function at finite temperatures and by the existence of patches with barrier heights smaller and larger than $\Phi_{\mathrm{Bn}}^{\mathrm{hom}}$.

8.1.2 Band Offsets of Semiconductor Heterostructures

Semiconductors generally grow layer-by-layer, at least initially. Hence, core-level photoemission spectroscopy (PES) is a very reliable tool and the one most widely used to determine the band-structure lineup at semiconductor heterostructures. The valence-band offset may be obtained from the energy positions of core-level lines in x-ray photoelectron spectra recorded with bulk samples of the semiconductors in contact and with the interface itself [8.24]. Since the escape depths of the photoelectrons are on the order of just 2 nm, one of the two semiconductors must be sufficiently thin. This condition is easily met when heterostructures are grown by molecular beam epitaxy (MBE) and photoemission spectra (PES) are recorded during growth interrupts. The valence-band discontinuity is then given by (Fig. 8.7)

$$\begin{aligned}\Delta W_{\mathrm{v}} &= W_{\mathrm{vir}} - W_{\mathrm{vil}} = W_{\mathrm{i}}(n_{\mathrm{r}}l_{\mathrm{r}}) - W_{\mathrm{i}}(n_{\mathrm{l}}l_{\mathrm{l}}) \\ &\quad + [W_{\mathrm{vbr}} - W_{\mathrm{b}}(n_{\mathrm{r}}l_{\mathrm{r}})] - [W_{\mathrm{vbl}} - W_{\mathrm{b}}(n_{\mathrm{l}}l_{\mathrm{l}})] ,\end{aligned} \qquad (8.11)$$

where $n_{\mathrm{r}}l_{\mathrm{r}}$ and $n_{\mathrm{l}}l_{\mathrm{l}}$ denote the core levels of the semiconductors on the right (r) and the left (l) side of the interface, respectively. The subscripts i and b characterize interface and bulk properties, respectively. The energy difference $W_{\mathrm{i}}(n_{\mathrm{r}}l_{\mathrm{r}}) - W_{\mathrm{i}}(n_{\mathrm{l}}l_{\mathrm{l}})$ between the core levels of the two semiconductors at the interface is determined from energy distribution curves of photoelectrons recorded during MBE growth of the heterostructure. The energy positions $W_{\mathrm{vbr}} - W_{\mathrm{b}}(n_{\mathrm{r}}l_{\mathrm{r}})$ and $W_{\mathrm{vbl}} - W_{\mathrm{b}}(n_{\mathrm{l}}l_{\mathrm{l}})$ of the core levels relative to the valence-band maxima in the bulk of the two semiconductors are evaluated separately.

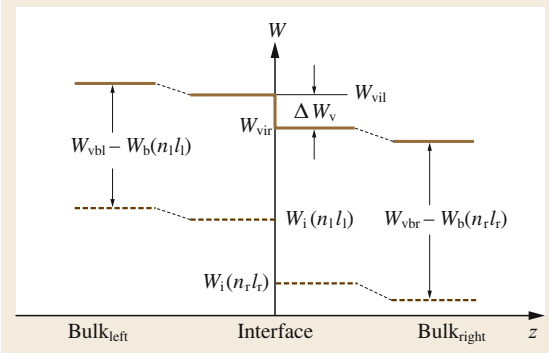


Fig. 8.7 Schematic energy band diagram at semiconductor heterostructures. W_{vb} and W_{vi} are the valence-band maxima and $W_{\mathrm{b}}(nl)$ and $W_{\mathrm{i}}(nl)$ are the core levels in the bulk and at the interface, respectively. The subscripts l and r denote the semiconductors on the right and the on the left side of the interface. ΔW_{v} is the valence-band offset. The *thin dashed lines* account for possible band-bending from space-charge layers

Another widely used technique for determining band offsets in heterostructures is internal photoemission yield spectroscopy. The procedure for evaluating the IPEYS signals is the same as described in Sect. 8.1.1.

8.2 IFIGS-and-Electronegativity Theory

Because of the quantum-mechanical tunneling effect, the wave functions of bulk electrons decay exponentially into vacuum at surfaces or, more generally speaking, at solid–vacuum interfaces. A similar behavior occurs at interfaces between two solids [8.6, 7]. In energy regions of Schottky contacts and semiconductor heterostructures where occupied band states overlap a band gap, the wave functions of these electrons will tail across the interface. The only difference to solid–vacuum interfaces is that the wave function tails oscillate at solid–solid interfaces. Figure 8.8 schematically explains the tailing effects at surfaces and semiconductor interfaces. For the band-structure lineup at semiconductor interfaces, only the tailing states within the gap between the top valence and the lowest conduction band are of any real importance since the energy position of the Fermi level determines their charging state. These wave function tails or interface-induced gap states (IFIGS) derive from the continuum of the virtual gap states (ViGS) of the complex semiconductor band structure. Hence, the IFIGS are an intrinsic property of the semiconductor.

The IFIGS are made up of valence-band and conduction-band states of the semiconductor. Their net charge depends on the energy position of the Fermi level relative to their branch point, where their character changes from predominantly donor- or valence-band-like to mostly acceptor- or conduction-band-like. The band-structure lineup at semiconductor interfaces is thus described by a zero-charge-transfer term and an electric dipole contribution.

In a more chemical approach, the charge transfer at semiconductor interfaces may be related to the partly ionic character of the covalent bonds at interfaces. *Pauling* [8.8] described the ionicity of single bonds in diatomic molecules by the difference between the electronegativities of the atoms involved. The binding energies of core-level electrons are known to depend on the chemical environment of the atoms or, in other words, on the ionicity of their chemical bonds. Figure 8.9 displays experimentally observed chemical shifts for Si(2*p*) and Ge(3*d*) core levels induced by metal adatoms on silicon and germanium surfaces as a function of the difference $X_m - X_s$ between the Pauling atomic electronegativity of the metal and that of the semiconductor atoms. The covalent bonds between metal and substrate atoms still persist at metal–semiconductor interfaces, as ab-initio calculations [8.25] have demonstrated for the example of Al/GaAs(110) contacts. The pronounced linear correlation of the data displayed in Fig. 8.9 thus justifies the application of Pauling's electronegativity concept to semiconductor interfaces.

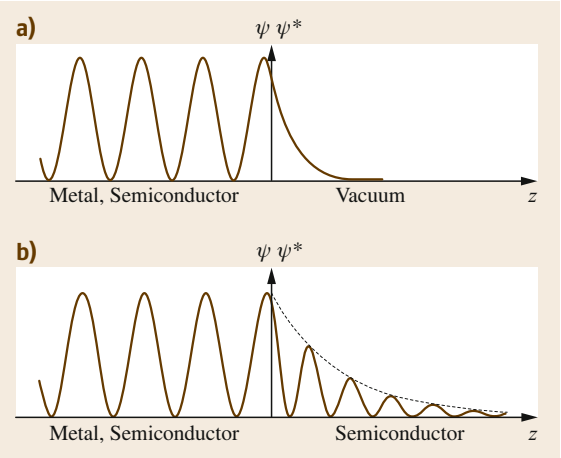


Fig. 8.8a,b Wave functions at clean surfaces **(a)** and at metal–semiconductor and semiconductor–semiconductor interfaces **(b)** (schematically)

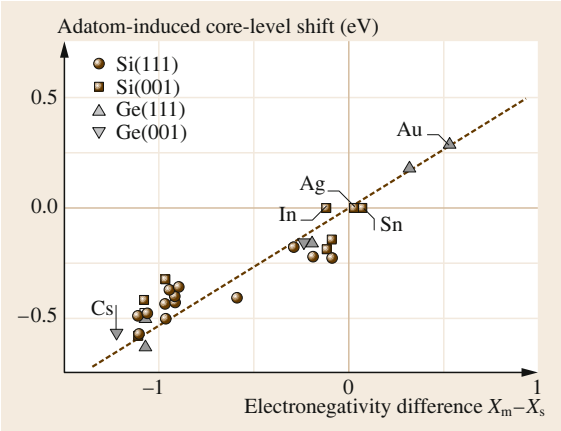


Fig. 8.9 Chemical shifts of Si(2*p*) and Ge(3*d*) core levels induced by metal adatoms on silicon and germanium surfaces, respectively, as a function of the difference $X_m - X_s$ in the metal and the semiconductor electronegativities in Pauling units. (After [8.13])

The combination of the physical IFIGS and the chemical electronegativity concept yields the barrier heights of ideal p-type Schottky contacts and the valence-band offsets of ideal semiconductor heterostructures as

$$\Phi_{\mathrm{Bp}} = \Phi^{\mathrm{p}}_{\mathrm{bp}} - S_{\mathrm{X}}(X_{\mathrm{m}} - X_{\mathrm{s}}) \tag{8.12}$$

and

$$\Delta W_{\mathrm{v}} = \Phi^{\mathrm{p}}_{\mathrm{bpr}} - \Phi^{\mathrm{p}}_{\mathrm{bpl}} + D_{\mathrm{X}}(X_{\mathrm{sr}} - X_{\mathrm{sl}}) \,, \tag{8.13}$$

respectively, where $\Phi^{\mathrm{p}}_{\mathrm{bp}} = W_{\mathrm{bp}} - W_{\mathrm{v}}(\Gamma)$ is the energy distance from the valence-band maximum to the branch point of the IFIGS or the p-type branch-point energy. It has the physical meaning of a zero-charge-transfer barrier height. The slope parameters S_{X} and D_{X} are explained at the end of this section.

The IFIGS derive from the ViGS of the complex band structure of the semiconductor. Their branch point is an average property of the semiconductor. *Tersoff* [8.26] calculated the branch-point energies $\Phi^{\mathrm{p}}_{\mathrm{bp}}$ of Si, Ge, and 13 of the III–V and II–VI compound semiconductors. He used a linearized augmented plane-wave method and the local density approximation. Such extensive computations may be avoided. *Mönch* [8.27] applied *Baldereschi*'s concept [8.28] of mean-value k-points to calculate the branch-point energies of zincblende-structure compound semiconductors. He first demonstrated that the quasi-particle band gaps of diamond, silicon, germanium, 3C-SiC, GaAs, and CdS at the mean-value k-point equal their average or dielectric band gaps [8.29]

$$W_{\mathrm{dg}} = \frac{\hbar\omega_{\mathrm{p}}}{\sqrt{\varepsilon_{\infty} - 1}} \,, \tag{8.14}$$

where $\hbar\omega_{\mathrm{p}}$ is the plasmon energy of the bulk valence electrons. Mönch then used Tersoff's $\Phi^{\mathrm{p}}_{\mathrm{bp}}$ values, calculated the energy dispersion of the topmost valence band in the empirical tight-binding approximation (ETB), and plotted the resulting branch-point energies at the mean-value k-point versus the widths of the dielectric band gaps W_{dg}. The linear least-squares fit to the data of the zincblende-structure compound semiconductors [8.30]

$$\Phi^{\mathrm{p}}_{\mathrm{bp}} = 0.449 W_{\mathrm{dg}} - [W_{\mathrm{v}}(\Gamma) - W_{\mathrm{v}}(\underline{k}_{\mathrm{mv}})]_{\mathrm{ETB}} \tag{8.15}$$

indicates that the branch points of these semiconductors lie 5% below the middle of the energy gap at the mean-value k-point k_{mv}. Table 8.1 displays the calculated p-type branch-point energies of the Group IV elemental semiconductors, of SiC, of III–V and II–VI compound semiconductors and of some I–III–VI_2 chalcopyrites.

Table 8.1 Optical dielectric constants, widths of the dielectric band gap, and branch-point energies of diamond-, zincblende, and chalcopyrite-structure semiconductors. (After [8.26, 29, 30])

Semiconductor	ε_{∞}	W_{dg} (eV)	$\Phi^{\mathrm{p}}_{\mathrm{bp}}$ (eV)
C	5.70	14.40	1.77
Si	11.90	5.04	0.36[a]
Ge	16.20	4.02	0.18[a]
3C-SiC	6.38	9.84	1.44
3C-AlN	4.84	11.92	2.97
AlP	7.54	6.45	1.13
AlAs	8.16	5.81	0.92
AlSb	10.24	4.51	0.53
3C-GaN	5.80	10.80	2.37
GaP	9.11	5.81	0.83
GaAs	10.90	4.97	0.52
GaSb	14.44	3.8	0.16
3C-InN	–	6.48	1.51
InP	9.61	5.04	0.86
InAs	12.25	4.20	0.50
InSb	15.68	3.33	0.22
ZnS	5.14	8.12	2.05
ZnSe	5.70	7.06	1.48
ZnTe	7.28	5.55	1.00
CdS	5.27	7.06	1.93
CdSe	6.10	6.16	1.53
CdTe	7.21	5.11	1.12
$CuGaS_2$	6.15	7.46	1.43[b]
$CuInS_2$	6.3*	7.02	1.47[b]
$CuAlSe_2$	6.3*	6.85	1.25[b]
$CuGaSe_2$	7.3*	6.29	0.93[b]
$CuInSe_2$	9.00	5.34	0.75[b]
$CuGaTe_2$	8.0*	5.39	0.61[b]
$CuInTe_2$	9.20	4.78	0.55[b]
$AgGaSe_2$	6.80	5.96	1.09[b]
$AgInSe_2$	7.20	5.60	1.11[b]

* $\varepsilon_{\infty} = n^2$, [a] [8.26], [b] [8.29]

A simple phenomenological model of Schottky contacts with a continuum of interface states and a constant density of states D_{is} across the semiconductor band gap yields the slope parameter [8.31, 32]

$$S_{\mathrm{X}} = \frac{A_{\mathrm{X}}}{\left(1 + \frac{e_0^2}{\varepsilon_{\mathrm{i}}\varepsilon_0} D_{\mathrm{is}}\delta_{\mathrm{is}}\right)} \,, \tag{8.16}$$

where ε_{i} is an interface dielectric constant. The parameter A_{X} depends on the electronegativity scale chosen and amounts to 0.86 eV/Miedema-unit and 1.79 eV/Pauling-unit. For $D_{\mathrm{is}} \to 0$, relation (8.16) yields $S_{\mathrm{X}} \to A_{\mathrm{X}}$ or, in other words, if no IFIGS were present at the metal–semiconductor interfaces one would obtain the Schottky–Mott rule. The extension δ_{is} of the interface states may be approximated by their charge decay

length $1/2q_{is}$. *Mönch* [8.32] used theoretical D_{gs}^{mi} and q_{gs}^{mi} data for metal-induced gap states (MIGS), as the IFIGS in Schottky contacts are traditionally called, and plotted the $(e_0^2/\varepsilon_0)D_{gs}^{mi}/(2q_{gs}^{mi})$ values versus the optical susceptibility $\varepsilon_\infty - 1$. The linear least-squares fit to the data points yielded [8.32]

$$\frac{A_X}{S_X} - 1 = 0.1(\varepsilon_\infty - 1)^2 \,, \tag{8.17}$$

where the reasonable assumption $\varepsilon_i \approx 3$ was made.

To a first approximation, the slope parameter D_X of heterostructure band offsets may be equated with the slope parameter S_X of Schottky contacts, since the IFIGS determine the intrinsic electric-dipole contributions to both the valence-band offsets and the barrier heights.

In early *Gedanken experiments*, Schottky contacts and semiconductor heterostructures were assembled by gradually approaching the respective two solids in vacuum [8.3, 33]. As long as a vacuum gap exists between the two surfaces facing each other, the vacuum level W_{vac} is the appropriate reference and, for example, the work function $W_{vac} - W_F$ of the metal and the electron affinity $W_{vac} - W_c$ or the ionization energy $W_{vac} - W_v$ of the semiconductor may be used to describe the charge distribution and the energy-band alignment in such a metal–vacuum–semiconductor array. As the vacuum gap is closed, that is, an intimate Schottky contact is made, chemical bonds between metal and semiconductor atoms will have formed at the now intimate metal–semiconductor interface. Hence, the work function of the metal as well as the electron affinity or the ionization energy of the semiconductor are inappropriate reference properties, but instead the electronegativities of the metal and the semiconductor may be used to describe the electric dipole layer at Schottky contacts [8.34]. The same arguments also hold for semiconductor heterostructures.

8.3 Comparison of Experiment and Theory

8.3.1 Barrier Heights of Schottky Contacts

Experimental barrier heights of intimate, abrupt, clean, and (above all) laterally homogeneous Schottky contacts on n-Si, n-GaAs, n-GaN, and SiO_2, as an example for insulators, are plotted in Figs. 8.10 and 8.11 versus the difference in the Miedema electronegativities of the metals and the semiconductors. *Miedema*'s electronegativities [8.35] are preferred since they were derived from properties of metal alloys and intermetallic compounds, while *Pauling* [8.8] considered covalent bonds in small molecules. The p- and n-type branch-point energies, $\Phi_{bp}^{p} = W_{bp} - W_v(\Gamma)$ and $\Phi_{bp}^{n} = W_c - W_{bp}$, respectively, add up to the fundamental band-gap energy $W_g = W_c - W_v(\Gamma)$. Hence, the barrier heights of n-type Schottky contacts are

$$\Phi_{Bn}^{hom} = \Phi_{bp}^{n} + S_X(X_m - X_s) \,. \tag{8.18}$$

The electronegativity of a compound is taken as the geometric mean of the electronegativities of its constituent atoms.

First, the experimental data plotted in Figs. 8.10 and 8.11 clearly confirm that the different experimental techniques, I/V, C/V, BEEM, IPEYS, and PES, yield barrier heights of laterally homogeneous Schottky contacts which agree within the margins of experimental error.

Second, the experimental Si, GaAs, and GaN data are quantitatively explained by the branch-point energies (8.15) and the slope parameters (8.17) of the IFIGS-and-electronegativity theory. As was already mentioned in Sect. 8.1.1, the stacking fault, which is part of the interfacial Si(111)-$(7\times7)^i$ reconstruction [8.15], causes an *extrinsic* electric dipole in addition to the *intrinsic* IFIGS electric dipole. The latter one is present irrespective of whether the interface structure is reconstructed or $(1\times1)^i$-unreconstructed. The extrinsic stacking fault-induced electric dipole quantitatively explains the experimentally observed barrier-height lowering of 76 ± 2 meV [8.10].

Third, the IFIGS line in Fig. 8.11a was drawn using the branch-point energy calculated for *cubic* 3*C*-GaN since relation (8.15) was derived for zincblende-structure compounds only. However, the Schottky contacts were prepared on hexagonal *wurtzite-structure* 2*H*-GaN. The good agreement between the experimental data and the IFIGS line indicates that the p-type branch-point energy seems to be rather insensitive to the specific bulk lattice structure of the semiconductor. This conclusion is further justified by the barrier heights of laterally homogeneous Schottky contacts on cubic 3*C*-SiC and its hexagonal polytypes 4*H* and 6*H* [8.13] as well as by the experimentally observed band-edge discontinuities of semiconductor heterostructures, which will be discussed in Sect. 8.3.2.

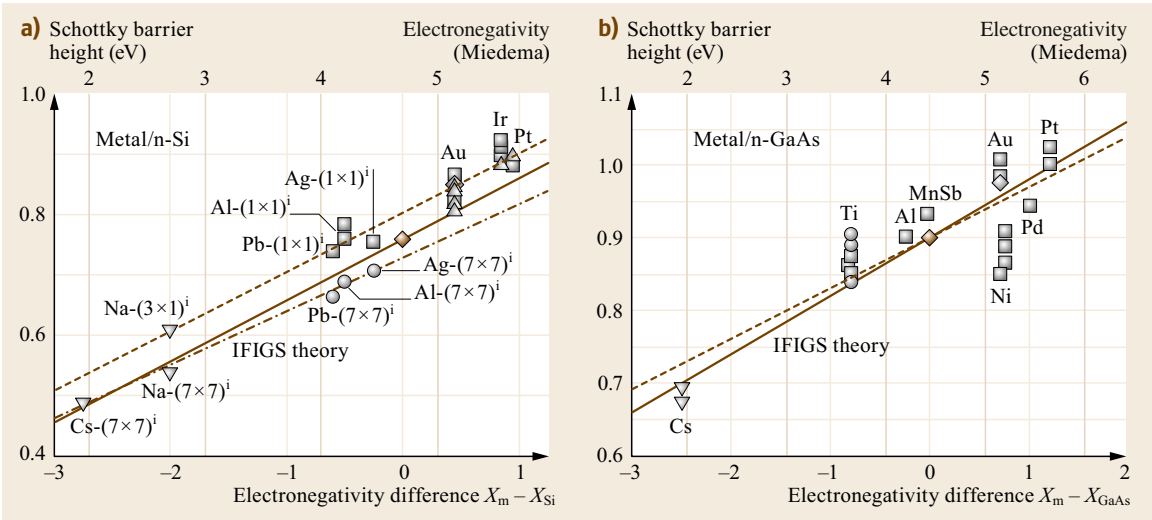


Fig. 8.10a,b Barrier heights of laterally homogeneous n-type silicon **(a)** and GaAs Schottky contacts **(b)** versus the difference in the Miedema electronegativities of the metals and the semiconductors. The □ and ○, △, ◇, and ▽ *symbols* differentiate the data from I/V, BEEM, IPEYS, and PES measurements, respectively. The *dashed* and the *dash-dotted lines* are the linear least-squares fits to the respective experimental data. The *solid* IFIGS *lines* are drawn with $S_X = 0.101\,\text{eV/Miedema-unit}$ and $\Phi_{bp}^{p} = 0.36\,\text{eV}$ for silicon **(a)** and with $S_X = 0.08\,\text{eV/Miedema-unit}$ and $\Phi_{bp}^{p} = 0.5\,\text{eV}$ for GaAs **(b)**. (After [8.13])

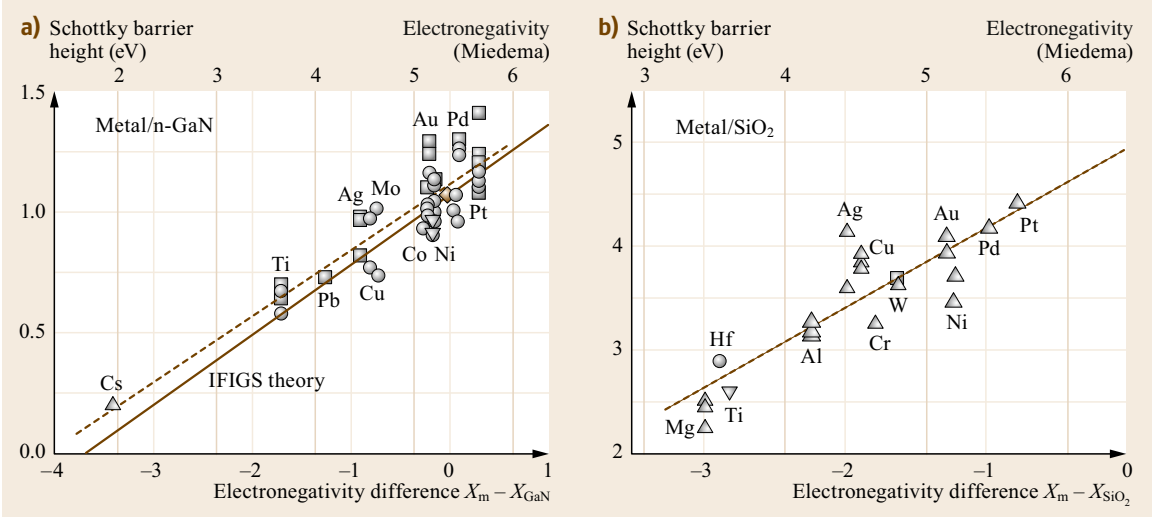


Fig. 8.11a,b Barrier heights of laterally homogeneous n-GaN(0001) **(a)** and SiO_2 Schottky contacts **(b)** versus the difference in the Miedema electronegativities of the metals and the semiconductors. **(a)** The *solid* IFIGS *line* is drawn with $S_X = 0.29\,\text{eV/Miedema-unit}$ and $\Phi_{bp}^{p} = 2.37\,\text{eV}$. **(a,b)** The □, ○, △, and ▽ *symbols* differentiate the data from I/V, C/V, IPEYS, and PES measurements. The *dashed lines* are the linear least-squares fits to the respective data. (After [8.13, 36])

8.3.2 Band Offsets of Semiconductor Heterostructures

In the bulk and at interfaces of sp^3-coordinated semiconductors, the chemical bonds are covalent. The simplest semiconductor–semiconductor interfaces are *lattice-matched* heterostructures. However, if the bond lengths of the two semiconductors differ, then the interface will respond with tetragonal lattice distortions. Such *pseudomorphic* interfaces are under tensile or

compressive stress. If the strain energy becomes too large, then it is energetically more favorable to release the stress by the formation of misfit dislocations. Such *metamorphic* interfaces are almost relaxed.

In contrast to isovalent heterostructures, the chemical bonds at heterovalent interfaces require special attention, since interfacial donor- and acceptor-type bonds may cause interfacial electric dipoles [8.37]. No such extrinsic electric dipoles will exist normal to nonpolar (110) interfaces. However, polar (001) interfaces behave quite differently. Acceptor bonds or donor bonds normal to the interface would exist at abrupt heterostructures. But, for reasons of charge neutrality, they have to be compensated by a corresponding density of donor bonds and acceptor bonds, respectively. This may be achieved by an intermixing at the interface which, on the other hand, causes extrinsic electric dipoles. Their components normal to the interface will add an extrinsic electric-dipole contribution to the valence-band offset.

The valence-band offsets at nonpolar heterostructures of compound semiconductors should equal the difference in the branch-point energies of the two semiconductors in contact provided the intrinsic IFIGS electric-dipole contribution can be neglected, see relation (8.13). This assumption is justified for heterostructures among the elemental Group IV semiconductors Ge and Si, SiC, and the III–V and II–VI compounds – with the nitrides and the oxides being excepted – because the elements that constitute these semiconductors are all placed in the middle of the Periodic Table so that their electronegativities only differ slightly. Figure 8.12a displays respective experimental results for diamond- and zincblende-structure semiconductors as a function of the difference in the calculated branch-point energies given in Table 8.1. The dashed line clearly demonstrates that the experimental data are excellently explained by the theoretical branch-point energies or, in other words, by the IFIGS theory.

As an example of lattice-matched and isovalent heterostructures, Fig. 8.13 shows valence-band offsets of $Al_{1-x}Ga_xAs/GaAs$ heterostructures as a function of the alloy composition x. The IFIGS branch-point energies of the alloys were calculated assuming virtual cations [8.30] and were found to vary linearly as a function of composition between the values of AlAs and GaAs. More refined first-principles calculations yielded identical results [8.38–40]. Figure 8.13 reveals that the theoretical IFIGS valence-band offsets excellently fit the experimental data.

Figure 8.12b displays valence-band offsets of metamorphic heterostructures versus the difference in the branch-point energies of the two semiconductors. The dashed line indicates that the experimental results are again excellently described by the theoretical IFIGS

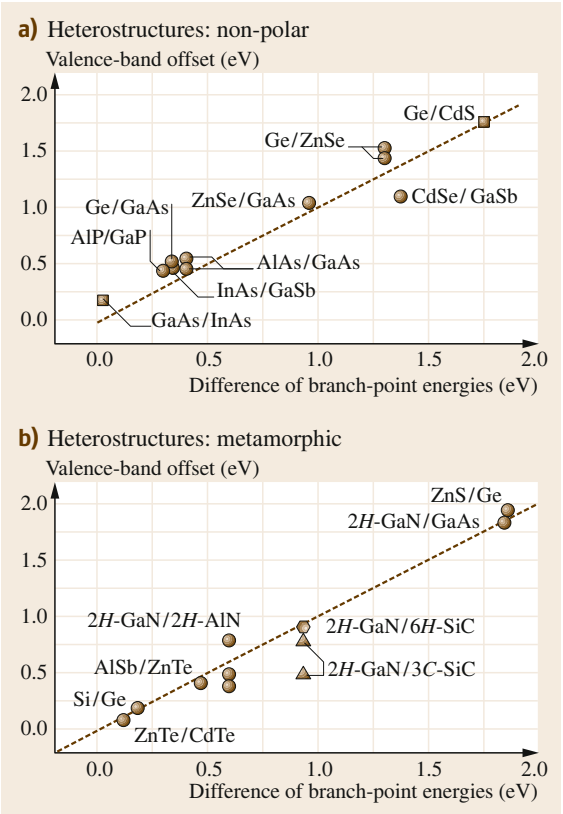


Fig. 8.12a,b Valence-band offsets at nonpolar (110)-oriented **(a)** and metamorphic semiconductor heterostructures **(b)** versus the difference between the p-type branch-point energies of the semiconductors in contact. (After [8.13])

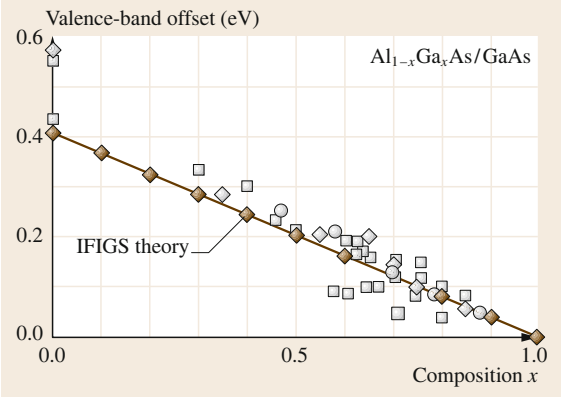


Fig. 8.13 Valence-band offsets of lattice-matched and isovalent heterostructures as a function of alloy composition x. (After [8.13])

data. This is true not only for heterostructures between cubic (C) zincblende- and hexagonal (H) wurtzite-

structure compounds but also for wurtzite-structure Group III nitrides grown on both cubic 3*C*- and hexagonal 6*H*-SiC substrates. These observations suggest the following conclusions. First, all of the heterostructures considered in Fig. 8.12b are only slightly (if at all) strained, although their lattice parameters differ by up to 19.8%. First-principles calculations [8.41] revealed that misfit dislocations change the valence-band offsets of metamorphic heterostructures by less than 0.1 eV only. Second, the calculations of the IFIGS branch-point energies assumed zincblende-structure semiconductors. These values, on the other hand, reproduce the experimental valence-band offsets irrespective of whether the semiconductors have zincblende, wurtzite, or, as in the case of 6*H*-SiC, another hexagonal-polytype structure. These findings again support the conclusion drawn from the GaN Schottky barrier heights in the previous section that the IFIGS branch-point energies are rather insensitive to the specific semiconductor bulk lattice structure.

Quite generally, the contribution of the IFIGS electric-dipole term $S_X(X_{sl} - X_{sr})$ to the valence-band offset ΔW_v of heterostructures may not be neglected. A simple rearrangement of terms in relation (8.13) gives

$$\Delta W_v - S_X(X_{sr} - X_{sl}) = \Phi^p_{bpr} - \Phi^p_{bpl} \,. \qquad (8.19)$$

For heterostructres of one and the same semiconductor, the valence-band offsets less the IFIGS-related electric-dipole contributions thus vary as the p-type branch-point energies of the respective other semiconductors. Figure 8.14 displays experimental valence-band offsets of SiO_2 heterostructures minus the respective electric-dipole contributions as a function of the calculated branch-point energies of the respective other semiconductors [8.26, 30] (Table 8.1); the S_X value of SiO_2 was adopted from Fig. 8.11b. The linear least-squares fit has a slope parameter of 0.94 ± 0.07 which value is close to 1 as is to be expected from (8.19), and its intercept on the ordinate gives the empirical p-type branch-point energy of SiO_2 as 3.73 ± 0.07 eV.

Figures 8.15 and 8.16 display experimental $\Delta W_v - S_X(X_{semi} - X_{Si,Ge,GaN})$ data of Si, Ge, and GaN heterostructures as a function of the branch-point energies of the other semiconductors and insulators involved. For the semiconductors, the calculated values [8.26, 30] were taken again while for the insulators empirical values [8.42–44] were considered. Again, the slope parameters of the linear least-squares fits (0.97 ± 0.03, 0.90 ± 0.04, and 0.87 ± 0.04) are close to 1, the value expected from (8.19). The empirical p-type branch-point energies result as 0.30 ± 0.09 eV for Si, 2.34 ± 0.4 eV for GaN, and 0.08 ± 0.10 eV for Ge. Within the margins of experimental error, these empirical values excellently agree with the calculated ones (Table 8.1).

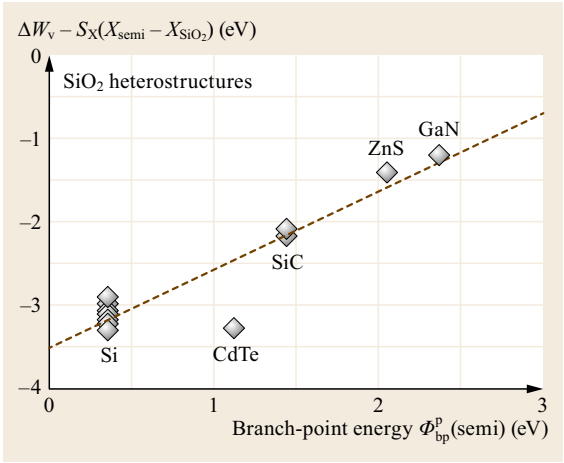


Fig. 8.14 Experimental valence-band offsets of SiO_2 heterostructures minus the respective IFIGS electric-dipole contributions as a function of the calculated p-type branch-point energies of the respective other semiconductors. The *dashed line* is the linear least-squares fit to the data points. (After [8.42])

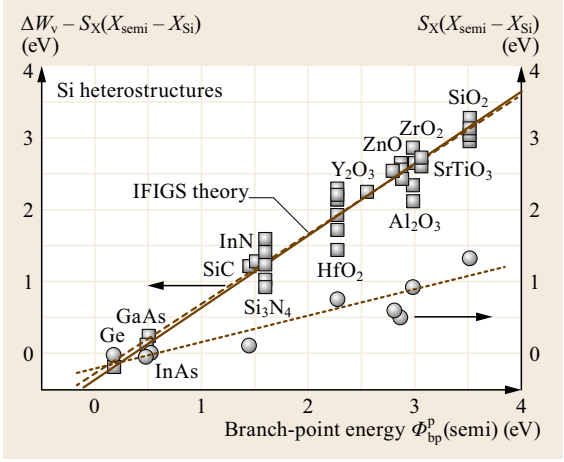


Fig. 8.15 Experimental valence-band offsets of Si heterostructures minus the respective IFIGS electric-dipole contributions (□) and the IFIGS electric-dipole contributions (○) as a function of calculated and empirical p-type branch-point energies of the respective other semiconductors and insulators, respectively. The *dashed* and *dash-dotted lines* are the linear least-squares fits to the data points. (After [8.43])

Figure 8.15 also displays some of the IFIGS electric-dipole contributions $S_X(X_{semi} - X_{Si})$ in Si heterostructures. The example of SiO_2/Si heterostructures demonstrates that this intrinsic dipole term measures 1.3 eV and this large value thus amounts to approximately 30% of the total valence-band offset.

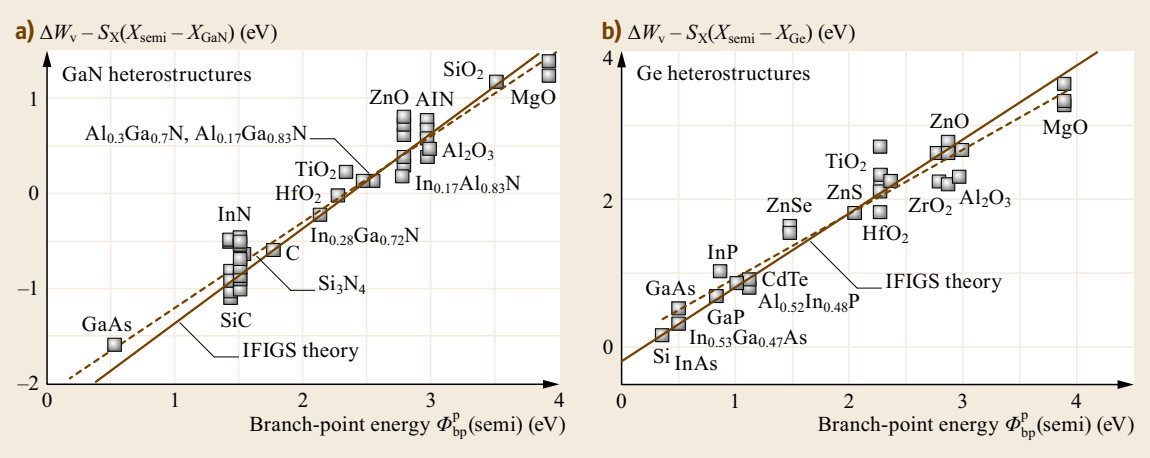


Fig. 8.16a,b Experimental valence-band offsets of GaN **(a)** and Ge heterostructures **(b)** minus the respective IFIGS electric-dipole contributions as a function of the calculated and empirical p-type branch-point energies of the respective other semiconductors and insulators, respectively. The *dashed lines* are the linear least-squares fits to the data points. (After [8.43])

Table 8.2 Fundamental band gaps W_g and empirical branch-point energies Φ_{bp}^{n} and Φ_{bp}^{p} as obtained from barrier heights of Schottky contacts and valence-band offsets of heterostructures, respectively [8.13, 36, 42–46]

Semiconductor	W_g (eV)	Φ_{bp}^{n} (eV)	Φ_{bp}^{p} (eV)	$\Phi_{bp}^{n} + \Phi_{bp}^{p}$ (eV)
Si	1.12	0.80 ± 0.01	0.30 ± 0.09	1.10
Ge	0.66	0.59 ± 0.01	0.08 ± 0.10	0.67
GaN	3.39	1.12 ± 0.03	2.34 ± 0.09	3.46
GaAs	1.42	0.90 ± 0.01	0.68 ± 0.11	1.58
InP	1.34	0.50 ± 0.02	0.86 ± 0.11	1.36
SiO_2	8.7–8.9	4.95 ± 0.19	3.73 ± 0.07	8.68
TiO_2	3.3–3.5	$1.13\ (\pm 0)$	2.34 ± 0.36	3.47
HfO_2	5.7–6.0	4.01 ± 0.19	2.27 ± 0.14	6.28
Al_2O_3	6.7–7.0	4.38 ± 0.34	2.98 ± 0.26	7.36
Ga_2O_3	4.9	1.34 ± 0.08	3.57 ± 0.50	4.91

As was already mentioned, the p- and n-type branch-point energies, $\Phi_{bp}^{p} = W_{bp} - W_v(\Gamma)$ and $\Phi_{bp}^{n} = W_c - W_{bp}$, respectively, add up to the fundamental band-gap energy $W_g = W_c - W_v(\Gamma)$. Table 8.2 shows empirical Φ_{bp}^{n} and Φ_{bp}^{p} values that were obtained from experimental barrier heights of n-type Schottky contacts and valence-band offsets of heterostructures, respectively; see, for example, Figs. 8.10, 8.11, 8.14–8.16, 8.19. Within the margins of experimental error, the empirical p- and n-type branch-point energies of these semiconductors and insulators indeed add up to the widths of their fundamental band gaps. These findings again strongly support the IFIGS-and-electronegativity concept of the band-structure alignment at semiconductor interfaces. The empirical Φ_{bp}^{n} values are of special importance since they were evaluated from experimental barrier heights with no other input than the metal and semiconductor electronegativities.

8.4 Modifications of Schottky Contacts

8.4.1 Nonalloyed Ohmic Contacts or MUTIS Schottky Contacts

Ideal metal–semiconductor or Schottky contacts are generally rectifying while genuine low-resistance or ohmic contacts are obtained with p-type Ge and GaSb and n-type InN, InAs, and InSb only because their IFIGS branch points are close to their valence-band maxima and above their conduction-band minima, respectively (Table 8.1). Recently, *Connelly* et al. [8.47, 48] studied a new path towards Ohmic contacts. They avoided alloying and demonstrated that the resistance of

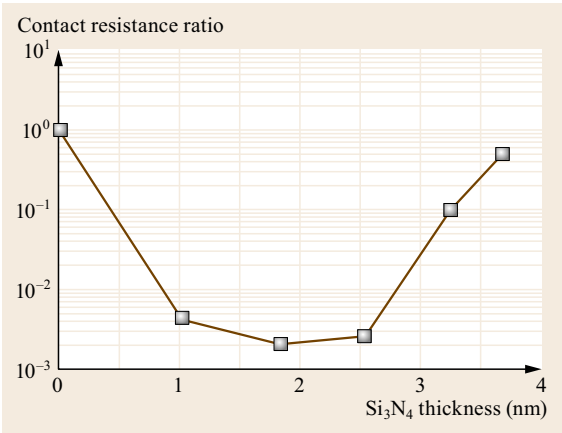


Fig. 8.17 Contact resistance ratio of Al/Si_3N_4/n-GaAs Schottky contacts versus the thickness of the Si_3N_4 interlayer. (After [8.49])

Si Schottky contacts is considerably lowered by ultrathin Si_3N_4 layers placed between the metal and silicon. As a typical example, Fig. 8.17 displays the contact resistance of Al/Si_3N_4/n-GaAs Schottky contacts normalized to the value of the intimate Al/n-GaAs contact in dependence on the insulator thickness [8.49]. The initial dramatic decrease of the contact resistance is accompanied by a gradual reduction of the Schottky barrier heights to a then constant value while the subsequent rise is easily attributed to the increasing tunnel resistance through the then thicker insulator interlayers.

Figure 8.18 schematically shows the energy-band diagram of ideal metal-ultrathin insulator–semiconductor or MUTIS structures resulting from the IFIGS-and-electronegativity concept. It is assumed that the electronegativity of the insulator is intermediate between the values of the metal and the semiconductor. In order to avoid that the IFIGS of the metal–insulator and the insulator–semiconductor interface overlap the thickness of the insulator has to be larger than four to five times the decay length at the IFIGS branch point of the insulator. Theoretical calculations found these dacay lengths to range between 0.09 and 0.28 nm [8.50]. The pronounced minimum value of the contact resistance ratio displayed in Fig. 8.17 is compatible with this estimate. Considering (8.12) and (8.13), one obtains

$$W_F - W_v^i = \Phi_{bp}^{pi} - S_X^i(X_m - X_i) \tag{8.20}$$

at the metal–insulator interface and

$$\begin{aligned} W_F - W_v^i &= \Delta W_v + W_g^s - \Phi_{Bn}^s \\ &= \Phi_{bp}^{pi} - \Phi_{bp}^{ps} + S_X^i(X_s - X_i) + W_g^s - \Phi_{Bn}^s \end{aligned} \tag{8.21}$$

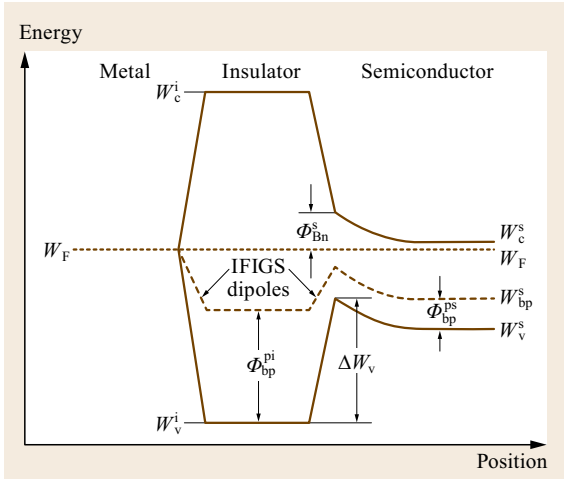


Fig. 8.18 Band-structure lineup at metal-ultrathin insulator–semiconductor (MUTIS) structures (schematically) with the assumption $X_m < X_i > X_s$. On the semiconductor side the extension of the band-bending regime is not drawn to scale. (After [8.50])

at the insulator–semiconductor interface where the sub- and superscripts m, i, and s indicate metals, insulators, and semiconductors, respectively. W_g^s designates the width of the fundamental band gap of the semiconductor. A rearrangement of (8.20) and (8.21) then gives the effective n-type barrier heights of the MUTIS structures as

$$\begin{aligned} \Phi_{Bn}^s &= W_g^s - \Phi_{bp}^{ps} + S_X^i(X_m - X_s) \\ &= \Phi_{bp}^{ns} + S_X^i(X_m - X_s) \end{aligned} \tag{8.22}$$

The effective n-type barrier height of an ideal n-type MUTIS structure is thus determined by the n-type branch-point energy Φ_{bp}^{ns} of the semiconductor and the slope parameter S_X^i of the insulator. The optical dielectric constant ε_∞ of Si_3N_4, and the same holds for other insulators such as Al_2O_3, Ge_3N_4, and MgO, is considerably smaller than the respective values of semiconductors as, for example, Ge, Si, and GaAs and, hence, (8.17) gives $S_X^i > S_X^s$ for the slope parameters. The insertion of the ultrathin insulators in MUTIS contacts thus alleviates the pinning of the Fermi level in comparison with the respective intimate MS contacts. Furthermore, (8.22) reveals that Ohmic rather than rectifying n-type MUTIS contacts will be obtained provided metals are chosen such that their electronegativities are sufficiently smaller than the values of the respective semiconductors.

Experimental barrier heights of Si_3N_4/n-Si and intimate n-Si Schottky contacts are plotted versus the electronegativity difference $X_m - X_{Si}$ in Fig. 8.19. The

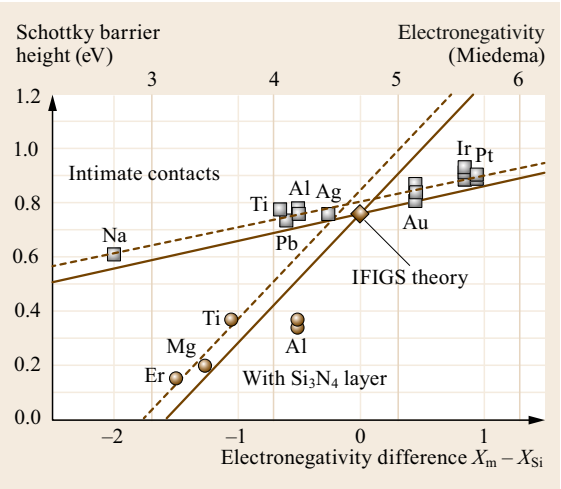


Fig. 8.19 Schottky barrier heights of intimate metal/n-Si (from Fig. 8.10a) and metal/Si_3N_4/n-Si versus the difference $X_m - X_{Si}$ of the metal and the Si electronegativities. The *full* and the *dashed lines* are the predictions of the IFIGS theory and the linear-least squares fits to the experimental data, respectively. (After [8.50])

thickness of the Si_3N_4 layers measured approximately 1 nm and is thus sufficiently larger than the estimated decay length of the IFIGS at their branch point. Within the margins of experimental error, the experimental barrier heights of the metal/Si_3N_4/n-Si junctions clearly confirm prediction (8.22) of the IFIGS-and-electronegativity concept in that the MUTIS barrier heights are determined by the branch-point energy of the semiconductor and the slope parameter of the insulator. This observation most directly demonstrates the alleviation of the Fermi-level pinning in MUTIS structures in comparison with intimate MS interfaces.

8.4.2 Atomic Interlayers

The barrier heights of metal–semiconductor contacts may also be modified by atomic interlayers. Figure 8.20 displays barrier heights of p-type silicon and diamond Schottky contacts without and with hydrogen interlayers [8.48, 50, 51]. Irrespective of the metals used, the H-induced modifications amount to -1.4 eV for p-diamond(001) but to $+0.3$ eV for p-Si(111) and to $+0.16$ eV for p-Si(001) Schottky contacts. In contrast to the insertion of ultrathin insulators, the H-interlayers do not alter the slope parameters S_X.

The metal-independent H-induced modifications of the Schottky barrier heights were explained by the partial ionic character of the covalent bonds between the H atoms and the interface atoms of the substrates [8.52]. The electronegativity of hydrogen ranges between the values of silicon and carbon, that is, $X_{Si} < X_H < X_C$ [8.8]. Hence, the partial ionic character of covalent Si–H and C–H bonds differs in sign, and $H^{-\Delta q}$-$Si^{+\Delta q}$ but $H^{+\Delta q}$-$C^{-\Delta q}$ bonds will exist at silicon and diamond interfaces, respectively. The opposite directions of the H-induced interface dipoles easily explain why the H-induced variations of the barrier

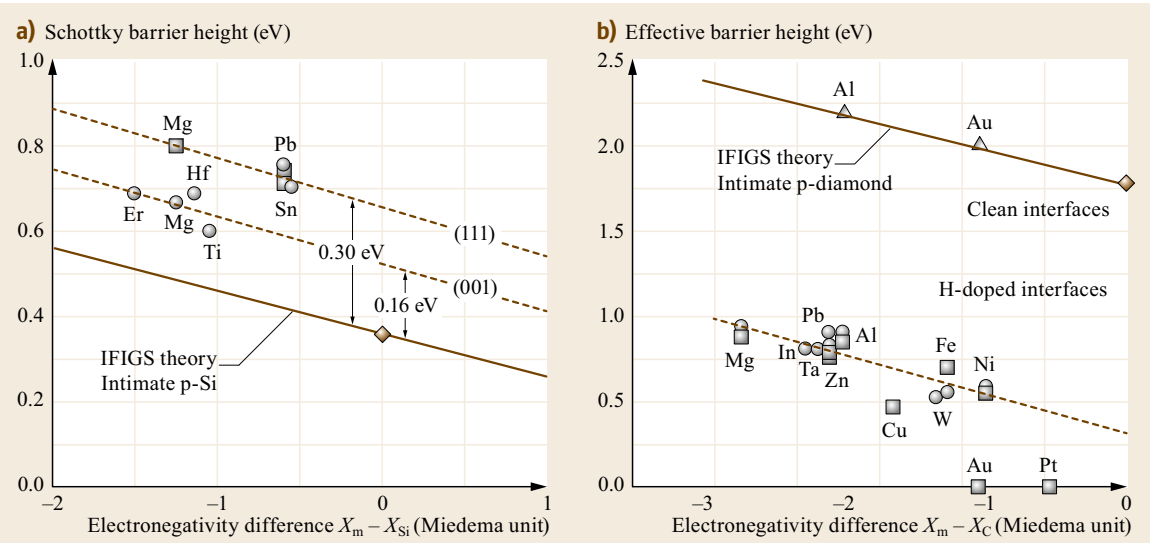


Fig. 8.20a,b Barrier heights of p-Si **(a)** and p-diamond Schottky contacts **(b)** with and without H-interlayers. The *full* and the *dashed lines* are the predictions of the IFIGS theory and the linear least-squares fits to the experimental data, respectively. (After [8.50])

height exhibit different signs with p-Si and p-diamond Schottky contacts.

These early conclusions were excellently confirmed by the recent observation that F-interlayers reduce the barrier heights of n-Ge Schottky contacts and do not change the slope parameter S_X [8.53]. The electronegativity of F is larger than the one of Ge so that $F^{-\Delta q}$-$Ge^{+\Delta q}$ dipoles are to be expected at F-doped Ge interfaces. These dipoles will reduce the barrier heights of metal/F:n-Ge Schottky contacts while, as with H:p-Si contacts, an increase would be observed with metal/F:p-Ge interfaces. The reduction of the barrier height observed with Fe/S:n-Ge contacts [8.54] is easily explained along the same lines.

The experimental data presented in Fig. 8.20 and [8.53] clearly demonstrate that atomic interlayers in Schottky contacts cause no depinning of the Fermi level and no alleviation of the Fermi-level pinning either.

8.5 Graphene Schottky Contacts

Graphene (Gr) is a single layer of sp^2-bonded carbon atoms arranged in a hexagonal crystal lattice. The remaining $2p_z$ orbitals are perpendicular to the graphene plane and form contiguous π and π^* bands so that graphene may be described as a zero-overlap semimetal. As the *conventional* metal-semiconductor contacts discussed in the preceding chapters graphene-semiconductor contacts indeed exhibit rectifying properties. However, *conventional* and Gr Schottky contacts differ in that the bonds between the metal and the semiconductor atoms are covalent while the bonds between graphene and semiconductor substrates are of van der Waals type. *Mönch* [8.55] verified that the experimental barrier heights of intimate Gr contacts on *conventional* n-type semiconductors as well as on layered n-MoS_2 agree with the values predicted by (8.17) and (8.18) but do not follow the Schottky–Mott rule. Hence, electric dipoles exist at Gr-semiconductor interfaces because graphene has a larger electronegativity than the semiconductors, with SiO_2 being the only exemption. The respective electronic charge transfer from the semiconductors towards the more electronegative Gr layer was also confirmed by theoretical calculations (see Refs. cited in [8.55]). The barrier heights of graphene Schottky contacts are thus explained by the π and π^* wave-function tails of the graphene across the interface in the energy range where the π–π^* bands of the graphene overlap the band gap of the semiconductor. The IFIGS-and-electronegativity concept thus also explains the band structure lineup at graphene-semiconductor interfaces rather than the Schottky–Mott rule as assumed in almost all the articles reporting barrier heights of such contacts.

8.6 Final Remarks

The local density approximation to density functional theory (LDA-DFT) is the most powerful and widely used tool in theoretical studies of the ground-state properties of solids. However, excitation energies such as the width of the energy gaps between the valence and conduction bands of semiconductors cannot be correctly obtained from such calculations. The fundamental band gaps of the elemental semiconductors C, Si, and Ge as well as of the III–V and II–VI compounds are notoriously underestimated by 25−50%. However, it became possible to compute quasi-particle energies and band gaps of semiconductors from first principles using the so-called GW approximation for the electron self-energy [8.56, 57]. The resulting band-gap energies agree to within 0.1−0.3 eV with experimental values.

For some specific metal–semiconductor contacts, the band-structure lineup was also studied by state-of-the-art ab-initio LDA-DFT calculations. The resulting LDA-DFT barrier heights were then subjected to a-posteriori corrections, which consider quasi-particle effects and, if necessary, spin–orbit interactions and semicore-orbital effects. However, comparison of the theoretical results with experimental data gives an inconsistent picture. The *mean* values of the barrier heights of Al- and Zn/p-ZnSe contacts, which were calculated for different interface configurations using ab-initio LDA-DF theory and a-posteriori spin–orbit and quasi-particle corrections [8.58, 59], agree with the experimental data to within the margins of experimental error. The same conclusion was reached for Al/$Al_xGa_{1-x}As$ Schottky contacts [8.60]. However, ab-initio LDA-DFT barrier heights of Al-, Ag-, and Au/p-GaN contacts [8.61, 62], as well as of Al– and Ti/3*C*-SiC(001) interfaces [8.63, 64], strongly deviate from the experimental results.

As already mentioned, ab-initio LDF-DFT valence-band offsets of heterostructures [8.39–41] reproduce the experimental results well. The same holds for *mean* values of LDF-DFT valence-band offsets computed for different interface configurations of GaN- and AlN/SiC-heterostructures [8.65–69].

The main difficulty, which the otherwise extremely successful ab-initio LDF-DFT calculations encounter when describing semiconductor interfaces, is not the precise exchange-correlation potential, which may be estimated in the GW approximation, but their remarkable sensitivity to the geometrical and compositional structure right at the interface. This aspect is more serious at metal–semiconductor interfaces than at heterostructures between two sp^3-bonded semiconductors. The more conceptual IFIGS-and-electronegativity theory, on the other hand, quantitatively explains not only the barrier heights of ideal Schottky contacts but also the valence-band offsets of semiconductor heterostructures. Here again, the Schottky contacts are the more important case, since their zero-charge-transfer barrier heights equal the branch-point energies of the semiconductors, while the valence-band offsets are determined by the differences in the branch-point energies of the semiconductors in contact.

References

8.1 W. Schottky: Naturwissenschaften **26**, 843 (1938)
8.2 F. Braun: Pogg. Ann. Physik Chemie **153**, 556 (1874)
8.3 N.F. Mott: Proc. Camb. Philos. Soc. **34**, 568 (1938)
8.4 W. Schottky: Phys. Z. **41**, 570 (1940)
8.5 J. Bardeen: Phys. Rev. **71**, 717 (1947)
8.6 V. Heine: Phys. Rev. **138**, A1689 (1965)
8.7 C. Tejedor, F. Flores: J. Phys. C **11**, L19 (1978)
8.8 L.N. Pauling: *The Nature of the Chemical Bond* (Cornell Univ. Press, Ithaca 1939)
8.9 W. Mönch: Phys. Rev. B **37**, 7129 (1988)
8.10 R. Schmitsdorf, T.U. Kampen, W. Mönch: Surf. Sci. **324**, 249 (1995)
8.11 J.L. Freeouf, T.N. Jackson, S.E. Laux, J.M. Woodall: Appl. Phys. Lett. **40**, 634 (1982)
8.12 W. Schottky: Phys. Z. **15**, 872 (1914)
8.13 W. Mönch: *Electronic Properties of Semiconductor Interfaces* (Springer, Berlin, Heidelberg 2004)
8.14 H.A. Bethe: Theory of the Boundary Layer of Crystal Rectifiers, MIT Radiation Lab. Rep. 43/12 (1942)
8.15 K. Takayanagi, Y. Tanishiro, M. Takahashi, S. Takahashi: Surf. Sci. **164**, 367 (1985)
8.16 T. Arizumi, M. Hirose: Jpn. J. Appl. Phys. **8**, 749 (1969)
8.17 Y.A. Goldberg, E.A. Posse, B.V. Tsarenkov: Sov. Phys. Semicond. **9**, 337 (1975)
8.18 W. Mönch: Appl. Phys. A **87**, 359 (2007)
8.19 W.J. Kaiser, L.D. Bell: Phys. Rev. Lett. **60**, 1406 (1988)
8.20 L.D. Bell, W.J. Kaiser: Phys. Rev. Lett. **61**, 2368 (1988)
8.21 H.-J. Im, Y. Ding, J.P. Pelz, W.J. Choyke: Phys. Rev. B **64**, 075310 (2001)
8.22 J. Cohen, J. Vilms, R.J. Archer: Hewlett-Packard R&D Report AFCRL-69-0287 (1969)
8.23 R. Turan, B. Aslan, O. Nur, M.Y.A. Yousif, M. Willander: Appl. Phys. A **72**, 587 (2001)
8.24 R.W. Grant, J.R. Waldrop, E.A. Kraut: Phys. Rev. Lett. **40**, 656 (1978)
8.25 S.B. Zhang, M.L. Cohen, S.G. Louie: Phys. Rev. B **34**, 768 (1986)
8.26 J. Tersoff: Phys. Rev. Lett. **52**, 465 (1984)
8.27 W. Mönch: J. Appl. Phys. **80**, 5076 (1996)
8.28 A. Baldereschi: Phys. Rev. B **7**, 5212 (1973)
8.29 D.R. Penn: Phys. Rev. **128**, 2093 (1962)
8.30 W. Mönch: Band-structure lineup at I-III-VI_2 Schottky contacts and heterostructures. In: *Wide-Gap Chalcopyrites*, ed. by S. Siebenritt, U. Rau (Springer, Berlin, Heidelberg 2006) p. 9
8.31 A.M. Cowley, S.M. Sze: J. Appl. Phys. **36**, 3212 (1965)
8.32 W. Mönch: Appl. Surf. Sci. **92**, 367 (1996)
8.33 R.L. Anderson: Solid-State Electron. **5**, 341 (1962)
8.34 C.A. Mead: Solid-State Electron. **9**, 1023 (1969)
8.35 A.R. Miedema, P.F. de Châtel, F.R. de Boer: Physica **100B**, 1 (1980)
8.36 W. Mönch: Appl. Phys. Lett. **86**, 122101 (2005)
8.37 W.A. Harrison, E.A. Kraut, J.R. Waldrop, R.W. Grant: Phys. Rev. B **18**, 4402 (1978)
8.38 J.S. Nelson, A.F. Wright, C.Y. Fong: Phys. Rev. B **43**, 4908 (1991)
8.39 S.B. Zhang, M.L. Cohen, S.G. Louie, D. Tománek, M.S. Hybertsen: Phys. Rev. B **41**, 10058 (1990)
8.40 Y. Wang, F. Zahid, Y. Zhu, L. Liu, J. Wang, H. Guo: Appl. Phys. Lett. **102**, 132109 (2013)
8.41 Y. Hinuma, F. Oba, I. Tanaka: Phys. Rev. B **88**, 075319 (2013)
8.42 W. Mönch: Appl. Phys. Lett. **91**, 042117 (2007)
8.43 W. Mönch: J. Appl. Phys. **109**, 113724 (2011)
8.44 W. Mönch: J. Appl. Phys. **107**, 013706 (2010)
8.45 W. Mönch: Appl. Phys. Lett. **93**, 172118 (2008)
8.46 W. Mönch: J. Mater. Sci. Mater. Electron. **27**, 1444 (2016)
8.47 D. Connelly, C. Faulkner, D.E. Grupp, J.S. Harris: IEEE Trans. Nanotechnol. **3**, 98 (2004)
8.48 D. Connelly, C. Faulkner, P.A. Clifton, D.E. Grupp: Appl. Phys. Lett. **88**, 012105 (2006)
8.49 J. Hu, K.C. Saraswat, H.-S.P. Wong: J. Appl. Phys. **107**, 063712 (2010)
8.50 W. Mönch: J. Appl. Phys. **111**, 073706 (2012)
8.51 M. Aoki, H. Kawarada: Jpn. J. Appl. Phys. **33**, L708 (1994)
8.52 W. Mönch: Europhys. Lett. **27**, 479 (1994)
8.53 J.-R. Wu, Y.-H. Wu, C.-Y. Hou, M.-L. Wu, C.-C. Lin, L.-L. Chen: Appl. Phys. Lett. **99**, 253504 (2011)
8.54 Y. Zhou, W. Han, Y. Wang, F. Xiu, J. Zou, R.K. Kawakami, K.L. Wang: Appl. Phys. Lett. **96**, 102103 (2010)

8.55 W. Mönch: J. Appl. Phys. **120**, 104501 (2016)
8.56 M.S. Hybertsen, S.G. Louie: Phys. Rev. B **34**, 5390 (1986)
8.57 R.W. Godby, M. Schlüter, L.J. Sham: Phys. Rev. B **37**, 10159 (1988)
8.58 M. Lazzarino, G. Scarel, S. Rubini, G. Bratina, L. Sorba, A. Franciosi, C. Berthod, N. Binggeli, A. Baldereschi: Phys. Rev. B **57**, R9431 (1998)
8.59 S. Rubini, E. Pellucchi, M. Lazzarino, D. Kumar, A. Franciosi, C. Berthod, N. Binggeli, A. Baldereschi: Phys. Rev. B **63**, 235307 (2001)
8.60 J. Bardi, N. Binggeli, A. Baldereschi: Phys. Rev. B **54**, R11102 (1996)
8.61 S. Picozzi, A. Continenza, G. Satta, S. Massidda, A.J. Freeman: Phys. Rev. B **61**, 16736 (2000)
8.62 S. Picozzi, G. Profeta, A. Continenza, S. Massidda, A.J. Freeman: Phys. Rev. B **65**, 165316 (2002)
8.63 J. Hoekstra, M. Kohyama: Phys. Rev. B **57**, 2334 (1998)
8.64 M. Kohyama, J. Hoekstra: Phys. Rev. B **61**, 2672 (2000)
8.65 M. Städele, A.J. Majewski, P. Vogl: Phys. Rev. B **56**, 6911 (1997)
8.66 J.A. Majewski, M. Städele, P. Vogl: Mater. Res. Soc. Symp. Proc. **449**, 917 (1997)
8.67 N. Binggeli, P. Ferrara, A. Baldereschi: Phys. Rev. B **63**, 245306 (2001)
8.68 B.K. Agrawal, S. Agrawal, R. Srivastava, P. Srivastava: Physica E **11**, 27 (2001)
8.69 M.R. Laridjani, P. Masri, J.A. Majewski: Mater. Res. Soc. Symp. Proc. **639**, G11.34 (2001)

9. Charge Transport in Disordered Materials

Sergei Baranovskii, Oleg Rubel

This chapter surveys general theoretical concepts developed to qualitatively understand and to quantitatively describe the electrical conduction properties of disordered organic and inorganic materials. In particular, these concepts are applied to describe charge transport in amorphous and microcrystalline semiconductors and in conjugated and molecularly doped polymers. Electrical conduction in such systems is achieved through incoherent transitions of charge carriers between spatially localized states. Basic theoretical ideas developed to describe this type of electrical conduction are considered in detail. Particular attention is given to the way the kinetic coefficients depend on temperature, the concentration of localized states, the strength of the applied electric field, and the charge carrier localization length. Charge transport via delocalized states in disordered systems and the relationships between kinetic coefficients under the nonequilibrium conditions are also briefly reviewed.

9.1 General Remarks on Charge Transport in Disordered Materials 195
9.2 Charge Transport in Disordered Materials via Extended States 198
9.3 Hopping Charge Transport in Disordered Materials via Localized States 200
9.3.1 Nearest-Neighbor Hopping 201
9.3.2 Variable-Range Hopping 203
9.3.3 Description of Charge-Carrier Energy Relaxation and Hopping Conduction in Inorganic Noncrystalline Materials 205
9.3.4 Description of Charge-Carrier Energy Relaxation and Hopping Conduction in Organic Noncrystalline Materials 211
9.4 Concluding Remarks 215
References 216

Many characteristics of charge transport in disordered materials differ markedly from those in perfect crystalline systems. The term *disordered materials* usually refers to noncrystalline solid materials without perfect order in the spatial arrangement of atoms. One should distinguish between disordered materials with ionic conduction and those with electronic conduction. Disordered materials with ionic conduction include various glasses consisting of a *network-formers* such as SiO_2, B_2O_3 and Al_2O_3, and of *network-modifiers* such as Na_2O, K_2O and Li_2O. When an external voltage is applied, ions can drift by hopping over potential barriers in the glass matrix, contributing to the electrical conduction of the material. Several fascinating effects have been observed for this kind of electrical conduction. One is the extremely nonlinear dependence of the conductivity on the concentration of ions in the material. Another beautiful phenomenon is the so-called *mixed alkali effect*: mixing two different modifiers in one glass leads to an enormous drop in the conductivity in comparison to that of a single modifier with the same total concentration of ions. A comprehensive description of these effects can be found in the review article of *Bunde* et al. [9.1]. Although these effects sometimes appear puzzling, they can be naturally and rather trivially explained using routine classical percolation theory [9.2]. The description of ionic conduction in glasses is much simplified by the inability of ions to tunnel over large distances in the glass matrix in single transitions. Every transition occurs over a rather small interatomic distance, and it is relatively easy to describe such electrical conductivity theoretically [9.2]. On the other hand, disordered systems with electronic conduction have a much more complicated theoretical description. Transition probabilities of electrons between spatially different regions in the material significantly depend not only on the energy parameters (as in the case of ions), but also on spatial factors such as the tunnelling distance, which can be rather large. The interplay between the energy and spatial fac-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_9

tors in the transition probabilities of electrons makes the development of a theory of electronic conduction in disordered systems challenging. Since the description of electronic conduction is less clear than that of ionic conduction, and since disordered electronic materials are widely used for various device applications, in this chapter we concentrate on disordered materials with the electronic type of electrical conduction.

Semiconductor glasses form one class of such materials. This class includes amorphous selenium, a-Se and other chalcogenide glasses, such as a-As_2Se_3. These materials are usually obtained by quenching from the melt. Another broad class of disordered materials, inorganic amorphous semiconductors, includes amorphous silicon a-Si, amorphous germanium a-Ge, and their alloys. These materials are usually prepared as thin films by the deposition of atomic or molecular species. Hydrogenated amorphous silicon, a-Si:H, has attracted much attention from researchers, since incorporation of hydrogen significantly improves conduction, making it favorable for use in amorphous semiconductor devices. Many other disordered materials, such as hydrogenated amorphous carbon (a-C:H) and its alloys, polycrystalline and microcrystalline silicon are similar to a-Si:H in terms of their charge transport properties. Some crystalline materials can also be considered to be disordered systems. This is the case for doped crystals if transport phenomena within them are determined by randomly distributed impurities, and for mixed crystals with disordered arrangements of various types of atoms in the crystalline lattice. In recent years much research has also been devoted to the study of organic disordered materials, such as conjugated and molecularly doped polymers and organic glasses, since these systems have been shown to possess electronic properties similar to those of inorganic disordered materials, while they are easier to manufacture than the latter systems.

There are two reasons for the great interest of researchers in the conducting properties of disordered materials. On the one hand, disordered systems represent a challenging field in a purely academic sense. For many years the theory of how semiconductors perform charge transport was mostly confined to crystalline systems where the constituent atoms are in regular arrays. The discovery of how to make solid amorphous materials and alloys led to an explosion in measurements of the electronic properties of these new materials. However, the concepts often used in textbooks to describe charge carrier transport in crystalline semiconductors are based on an assumption of long-range order, and so they cannot be applied to electronic transport in disordered materials. It was (and still is) a highly challenging task to develop a consistent theory of charge transport in such systems. On the other hand, the explosion in research into charge transport in disordered materials is related to the various current and potential device applications of such systems. These include the application of disordered inorganic and organic materials in photovoltaics (the functioning material in solar cells), in electrophotography, in large-area displays (they are used in thin film transistors), in electrical switching threshold and memory devices, in light-emitting diodes, in linear image sensors, and in optical recording devices. Readers interested in the device applications of disordered materials should be aware that there are numerous monographs on this topic: the literature on this field is very rich. Several books are recommended [9.3–12], as are numerous review articles referred to in these books.

In this chapter we focus on disordered semiconductor materials, ignoring the broad class of disordered metals. In order to describe electronic transport in disordered metals, one can more or less successfully apply extended and modified conventional theoretical concepts developed for electron transport in ordered crystalline materials, such as the Boltzmann kinetic equation. Therefore, we do not describe electronic transport in disordered metals here. We can recommend a comprehensive monograph to interested readers [9.13], in which modern concepts about conduction in disordered metals are presented beautifully.

Several nice monographs on charge transport in disordered semiconductors are also available. Although many of them were published several years ago (some even decades ago), we can recommend them to the interested reader as a source of information on important experimental results. These results have permitted researchers the present level of understanding of transport phenomena in disordered inorganic and organic materials. A comprehensive collection of experimental data for noncrystalline materials from the books specified above would allow one to obtain a picture of the modern state of experimental research in the field.

We will focus in this chapter on the theoretical description of charge transport in disordered materials, introducing some basic concepts developed to describe electrical conduction. Several excellent books already exist in which a theoretical description of charge transport in disordered materials is the main topic. Among others we can recommend the books of *Shklovskii* and *Efros* [9.14], *Zvyagin* [9.15], *Böttger* and *Bryksin* [9.16], and *Overhof* and *Thomas* [9.17]. There appears to be a time gap in which comprehensive monographs on the theoretical description of electrical conduction in disordered materials were not published. During this period some new and rather powerful theoretical concepts were developed. We present these concepts below, along with some more traditional ones.

9.1 General Remarks on Charge Transport in Disordered Materials

Although the literature on transport phenomena in disordered materials is enormously rich, there are still many open questions in this field due to various problems specific to such materials. In contrast to ordered crystalline semiconductors with well-defined electronic energy structures consisting of energy bands and energy gaps, the electronic energy spectra of disordered materials can be treated as quasi-continuous. Instead of bands and gaps, one can distinguish between extended and localized states in disordered materials. In an extended state, the charge carrier wavefunction is spread over the whole volume of a sample, while the wavefunction of a charge carrier is localized in a spatially restricted region in a localized state, and a charge carrier present in such a state cannot spread out in a plane wave as in ordered materials. Actually, localized electron states are known in ordered systems too. Electrons and holes can be spatially localized when they occupy donors or acceptors or some other impurity states or structural defects in ordered crystalline materials. However, the localized states usually appear as δ-like discrete energy levels in the energy spectra of such materials. In disordered semiconductors, on the other hand, energy levels related to spatially localized states usually fill the energy spectrum continuously. The energy that separates the extended states from the localized ones in disordered materials is called the mobility edge. To be precise, we will mostly consider the energy states for electrons in the following. In this case, the states above the mobility edge are extended and the states below the edge are localized. The localized states lie energetically above the extended states for holes. The energy region between the mobility edges for holes and electrons is called the mobility gap. The latter is analogous to the band gap in ordered systems, although the mobility gap contains energy states, namely the spatially localized states. Since the density of states (DOS), defined as the number of states per unit energy per unit volume, usually decreases when the energy moves from the mobility edges toward the center of the mobility gap, the energy regions of localized states in the vicinity of the mobility edges are called band tails. We would like to emphasize that the charge transport properties depend significantly on the energy spectrum in the vicinity and below the mobility edge (in the band tails). Unfortunately this energy spectrum is not known for almost all disordered materials. A whole variety of optical and electrical investigation techniques have proven unable to determine this spectrum. Since the experimental information on this spectrum is rather vague, it is difficult to develop a consistent theoretical description for charge transport ab initio. The absence of reliable information on the energy spectrum and on the structures of the wavefunctions in the vicinity and below the mobility edges can be considered to be the main problem for researchers attempting to quantitatively describe the charge transport properties of disordered materials.

An overview of the energy spectrum in a disordered semiconductor is shown in Fig. 9.1. The energy levels ε_v and ε_c denote the mobility edges for the valence and conduction bands, respectively. Electron states in the mobility gap between these energies are spatially localized. The states below ε_v and above ε_c can be occupied by delocalized holes and electrons. Some peaks in the DOS are shown in the mobility gap, which can be created by some defects with particularly high concentrations. Although there is a consensus between researchers on the general view of the DOS in disordered materials, the particular structure of the energy spectrum is not known for most disordered systems. From a theoretical point of view, it is enormously difficult to calculate this spectrum.

There are several additional problems that make the study of charge transport in disordered materials more difficult than in ordered crystalline semiconductors. The particular spatial arrangements of atoms and molecules in different samples with the same chemical composition can differ from each other depending on the preparation conditions. Hence, when discussing electrical conduction in disordered materials one often should specify the preparation conditions. Another problem is related to the long-time relaxation processes in disordered systems. Usually these systems are not in thermodynamic equilibrium and the slow relaxation of the atoms toward the equilibrium arrangement can lead to some changes in electrical conduction proper-

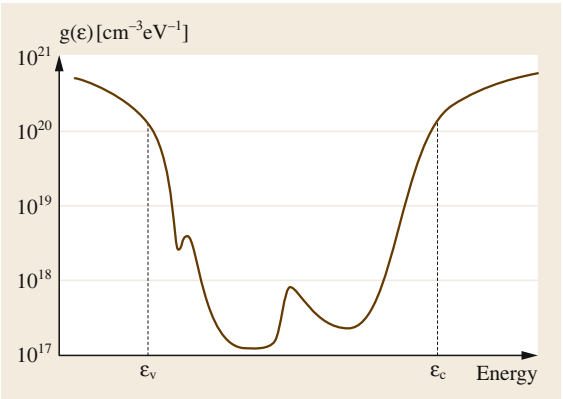


Fig. 9.1 Density of states of a noncrystalline semiconductor (schematic); ε_v and ε_c correspond to mobility edges in the conduction band and the valence band, respectively

ties. In some disordered materials a long-time electronic relaxation can affect the charge transport properties too, particularly at low temperatures, when electronic spatial rearrangements can be very slow. At low temperatures, when tunneling electron transitions between localized states dominate electrical conduction, this long-time electron relaxation can significantly affect the charge transport properties.

It is fortunate that, despite these problems, some general transport properties of disordered semiconductors have been established. Particular attention is usually paid to the temperature dependence of the electrical conductivity, since this dependence can indicate the underlying transport mechanism. Over a broad temperature range, the direct current (DC) conductivity in disordered materials takes the form

$$\sigma = \sigma_0 \exp\left[-\left(\frac{\Delta(T)}{k_B T}\right)^{\beta}\right], \tag{9.1}$$

where the pre-exponential factor σ_0 depends on the underlying system and the power exponent β depends on the material and also sometimes on the temperature range over which the conductivity is studied; $\Delta(T)$ is the activation energy. In many disordered materials, like vitreous and amorphous semiconductors, σ_0 is of the order of 10^2–$10^4\,\Omega^{-1}\,\mathrm{cm}^{-1}$. In such materials the power exponent β is close to unity at temperatures close to and higher than the room temperature, while at lower temperatures β can be significantly smaller than unity. In organic disordered materials, values of β that are larger than unity also have been reported. For such systems the value $\beta \approx 2$ is usually considered to be appropriate [9.18].

Another important characteristic of the electrical properties of a disordered material is its alternating current (AC) conductivity measured when an external alternating electric field with some frequency ω is applied. It has been established in numerous experimental studies that the real part of the AC conductivity in most disordered semiconductors depends on the frequency according to the power law

$$\mathrm{Re}\,\sigma(\omega) = C\omega^s\,, \tag{9.2}$$

where C is constant and the power s is usually smaller than unity. This power law has been observed in numerous materials at different temperatures over a wide frequency range. This frequency dependence differs drastically from that predicted by the standard kinetic theory developed for quasi-free charge carriers in crystalline systems. In the latter case, the real part of the AC conductivity has the frequency dependence

$$\mathrm{Re}\,\sigma(\omega) = \frac{ne^2}{m}\frac{\tau}{1+\omega^2\tau^2}\,, \tag{9.3}$$

where n is the concentration of charge carriers, e is the elementary charge, m is the effective mass and τ is the momentum relaxation time. Since the band electrons in crystalline semiconductors usually have rather short momentum relaxation times, $\tau \approx 10^{-14}$ s, the contribution of charge carriers in delocalized states to the AC conductivity usually does not depend on frequency at $\omega \ll \tau^{-1}$. Therefore, the observed frequency dependence described by (9.2) should be ascribed to the contribution of charge carriers in localized states.

One of the most powerful tools used to study the concentrations of charge carriers and their mobilities in crystalline semiconductors is the provided by measurements of the Hall constant, R_H. Such measurements also provide direct and reliable information about the sign of the charge carriers in crystalline materials. Unfortunately, this is not the case for disordered materials. Moreover, several anomalies have been established for Hall measurements in the latter systems. For example, the sign of the Hall constant in disordered materials sometimes differs from that of the thermoelectric power, α. This anomaly has not been observed in crystalline materials. The anomaly has been observed in liquid and solid noncrystalline semiconductors. Also, in some materials, like amorphous arsenic, a-As, $R_H > 0$, $\alpha < 0$, while in many other materials other combinations with different signs of R_H and α have been experimentally established.

In order to develop a theoretical picture of the transport properties of any material, the first issues to clarify are the spectrum of the energy states for charge carriers and the spatial structure of such states. Since these two central issues are yet to be answered properly for noncrystalline materials, the theory of charge transport in disordered systems should be considered to be still in its embryonic stage.

The problem of deducing electron properties in a random field is very complicated, and the solutions obtained so far only apply to some very simple models. One of them is the famous Anderson model that illustrates the localization phenomenon caused by random disorder [9.19]. In this model, one considers a regular system of rectangular potential wells with randomly varying depths, as shown schematically in Fig. 9.2. The ground state energies of the wells are assumed to be randomly distributed over the range with a width of W. First, one considers the ordered version of the model, with W equal to zero. According to conventional band theory, a narrow band arises in the ordered system where the energy width depends on the overlap integral I between the electron wavefunctions in the adjusting wells. The eigenstates in such a model are delocalized with wavefunctions of the Bloch type. This is trivial. The problem is to find the solution for a finite

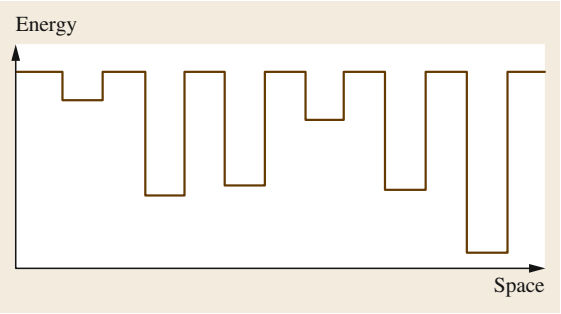


Fig. 9.2 Anderson model of disorder potential

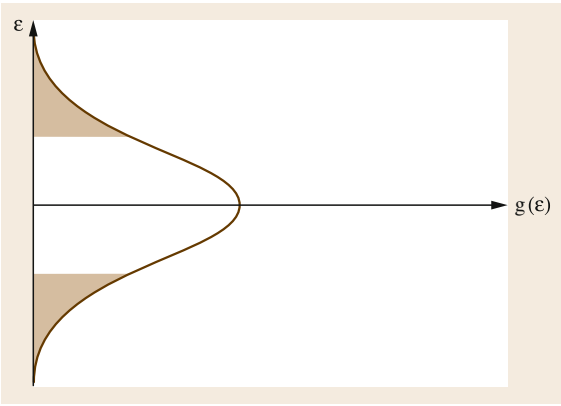


Fig. 9.3 Density of states in the Anderson model. *Hatched regions* in the tails correspond to spatially localized states

degree of disorder ($W \neq 0$). The result from the Anderson model for such a case is described as follows. At some particular value for the ratio $W/(zI)$, where z is the coordination number of the lattice, all electron states of the system are spatially localized. At smaller values of $W/(zI)$ some states in the outer regions of the DOS are localized and other states in the middle of the DOS energy distribution are spatially extended, as shown schematically in Fig. 9.3. This is one of the most famous results in the transport theory of disordered systems. When considering this result, one should note the following points. (i) It was obtained using a single-electron picture without taking into account long-range many-particle interactions. However, in disordered systems with localized electrons such interactions can lead to the localization of charge carriers and they often drastically influence the energy spectrum [9.14]. Therefore the applicability of the single-electron Anderson result to real systems is questionable. (ii) Furthermore, the energy structure of the Anderson model shown in Fig. 9.3 strongly contradicts that observed in real disordered materials. In real systems, the mobility gap is located between the mobility edges, as shown in Fig. 9.1, while in the Anderson model the energy region between the mobility edges is filled with delocalized states. Moreover, in one-dimensional and in some two-dimensional systems, the Anderson model predicts that all states are localized at any amount of disorder. These results are of little help when attempting to interpret the DOS scheme in Fig. 9.1.

A different approach to the localization problem is to try to impose a random potential $V(x)$ onto the band structure obtained in the frame of a traditional band theory. Assuming a classical smoothly varying (in space) potential $V(x)$ with a Gaussian distribution function

$$F(V) = \frac{1}{\varepsilon_0\sqrt{2\pi}} \exp\left(-\frac{V^2}{2\varepsilon_0^2}\right), \tag{9.4}$$

one can solve the localization problem using the classical percolation theory illustrated in Fig. 9.4. In Fig. 9.4a, an example of a disorder potential experienced by electrons is shown schematically. In Fig. 9.4b and Fig. 9.4c the regions below a given energy level E_c are colored black. In Fig. 9.4b this level is positioned very low, so that regions with energies below E_c do not provide a connected path through the system. In Fig. 9.4c an infinite percolation cluster consisting only of black regions exists. The E_c that corresponds to the first appearance of such a connected path is called the classical percolation level [9.14]. Mathematically soluing the percolation problem shows that the mobility edge identified with the classical percolation level in the potential $V(x)$ is shifted with respect to the band edge of the ordered system by an amount $\xi\varepsilon_0$, where $\xi \approx 0.96$ towards the center of the bandgap [9.15]. A similar result, though with a different constant ξ, can be obtained via a quantum-mechanical treatment of a short-range potential $V(x)$ of white-noise type [9.20]. As the amplitude ε_0 of the random potential increases the band gap narrows, while the conduction and valence bands become broader. Although this result is provided by both limiting models – by the classical one with a long-range smoothly varying potential $V(x)$ and by the quantum-mechanical one with a short-range white-noise potential $V(x)$ – none of the existing theories can reliably describe the energy spectrum of a disordered material and the properties of the charge carrier wavefunctions in the vicinity of the mobility edges, in other words in the energy range which is most important for charge transport.

The DC conductivity can generally be represented in the form

$$\sigma = e \int \mu(\varepsilon) n(\varepsilon) \mathrm{d}\varepsilon \,, \tag{9.5}$$

where e is the elementary charge, $n(\varepsilon)\mathrm{d}\varepsilon$ is the concentration of electrons in the energy range between ε and $\varepsilon + \mathrm{d}\varepsilon$ and $\mu(\varepsilon)$ is the mobility of these electrons.

Fig. 9.4a–c Disorder potential landscape experienced by a charge carrier **(a)**. Regions with energies below some given energy level E_c are *colored black*. In **(b)** this level is very low and there is no connected path through the system via *black regions*. In **(c)** the level E_c corresponds to the classical percolation level

The integration is carried out over all energies ε. Under equilibrium conditions, the concentration of electrons $n(\varepsilon)\mathrm{d}\varepsilon$ is determined by the density of states $g(\varepsilon)$ and the Fermi function $f(\varepsilon)$, which depends on the position of the Fermi energy ε_F (or a quasi-Fermi energy in the case of the stationary excitation of electrons)

$$n(\varepsilon) = g(\varepsilon) f(\varepsilon)\,, \tag{9.6}$$

where

$$f(\varepsilon) = \frac{1}{1+\exp\left(\frac{\varepsilon-\varepsilon_F}{k_B T}\right)}\,. \tag{9.7}$$

Here T is the temperature and k_B is the Boltzmann constant.

The Fermi level in almost all known disordered semiconductors under real conditions is situated in the mobility gap – in the energy range which corresponds to spatially localized electron states. The charge carrier mobility $\mu(\varepsilon)$ in the localized states below the mobility edge is much less than that in the extended states above the mobility edge. Therefore, at high temperatures, when a considerable fraction of electrons can be found in the delocalized states above the mobility edge, these states dominate the electrical conductivity of the system. The corresponding transport mechanism under such conditions is similar to that in ordered crystalline semiconductors. Electrons in the states within the energy range of the width, of the order $k_B T$ above the mobility edge, dominate the conductivity. In such a case the conductivity can be estimated as

$$\sigma \approx e\mu_c n(\varepsilon_c) k_B T\,, \tag{9.8}$$

where μ_c is the electron mobility in the states above the mobility edge ε_c, and $n(\varepsilon_c)k_B T$ is their concentration. This equation is valid under the assumption that the typical energy scale of the DOS function $g(\varepsilon)$ above the mobility edge is larger than $k_B T$. The position of the Fermi level in disordered materials usually depends on temperature only slightly. Combining (9.6)–(9.8), one obtains the temperature dependence of the DC conductivity in the form

$$\sigma = \sigma_0 \exp\left(-\frac{\Delta}{k_B T}\right), \tag{9.9}$$

described by (9.1) with $\beta = 1$ and constant activation energy, which is observed in most disordered semiconductors at high temperatures.

In order to obtain the numerical value of the conductivity in this high-temperature regime, one needs to know the density of states in the vicinity of the mobility edge $g(\varepsilon_c)$, and also the magnitude of the electron mobility μ_c in the delocalized states above ε_c. While the magnitude of $g(\varepsilon_c)$ is usually believed to be close to the DOS value in the vicinity of the band edge in crystalline semiconductors, there is no consensus among researchers on the magnitude of μ_c. In amorphous semiconductors μ_c is usually estimated to be in the range of $1-10\,\mathrm{cm^2/(V\,s)}$. Unfortunately, there are no reliable theoretical calculations of this quantity for most disordered materials. The only exception is provided by so-called mixed crystals, which are also sometimes called crystalline solid solutions. In the next section we describe the theoretical method which allows one to estimate μ_c in such systems. This method can be extended to other disordered materials, provided the statistical properties of the disorder potential, essential for electron scattering, are known.

9.2 Charge Transport in Disordered Materials via Extended States

The states with energies below ε_v and above ε_c in disordered materials are believed to possess similar properties to those of extended states in crystals. Various experimental data suggest that these states in disordered materials are delocalized states. However, traditional band theory is largely dependent upon the system having translational symmetry. It is the periodic atomic structure of crystals that allows one to describe

electrons and holes within such a theory as quasi-particles that exhibit behavior similar to that of free particles in vacuum, albeit with a renormalized mass (the so-called *effective mass*). The energy states of such quasi-particles can be described by their momentum values. The wavefunctions of electrons in these states (the so-called Bloch functions) are delocalized. This means that the probability of finding an electron with a given momentum is equal at corresponding points of all elementary cells of the crystal, independent on the distance between the cells.

Strictly speaking, the traditional band theory fails in the absence of translational symmetry – for disordered systems. Nevertheless, one still assumes that the charge carriers present in delocalized states in disordered materials can be approximately described by wavefunctions with a spatially homogeneous probability of finding a charge carrier with a given quasi-momentum. As for crystals, one starts from the quasi-free particle picture and considers the scattering effects in a perturbation approach following the Boltzmann kinetic description. This description is valid if the de Broglie wavelength of the charge carrier $\lambda = \hbar/p$ is much less than the mean free path $l = v\tau$, where τ is the momentum relaxation time and p and v are the characteristic values of the momentum and velocity, respectively. This validity condition for the description based on the kinetic Boltzmann equation can also be expressed as $\hbar/\tau \ll \varepsilon$, where ε is the characteristic kinetic energy of the charge carriers, which is equal to $k_B T$ for a nondegenerate electron gas and to the Fermi energy in the degenerate case. While this description seems valid for delocalized states far from the mobility edges, it fails for energy states in the vicinity of the mobility edges. So far, there has been no consensus between the theorists on how to describe charge carrier transport in the latter states. Moreover, it is not clear whether the energy at which the carrier mobility drops coincides with the mobility edge or whether it is located above the edge in the extended states. Numerous discussions of this question, mostly based on the scaling theory of localization, can be found in special review papers. For the rest of this section, we skip this rather complicated subject and instead we focus on the description of charge carrier transport in a semiconductor with a short-range random disorder potential of white-noise type. This seems to be the only disordered system where a reliable theory exists for charge carrier mobility via extended states above the mobility edge. Semiconductor solid solutions provide an example of a system with this kind of random disorder [9.20–25].

Semiconductor solid solutions A_xB_{1-x} (mixed crystals) are crystalline semiconductors in which the sites of the crystalline sublattice can be occupied by atoms of two different types, A and B. Each site can be occupied by either an A or a B atom with some given probability x between zero and unity. The value x is often called the composition of the material. Due to the random spatial distributions of the A and B atoms, local statistical fluctuations in the composition inside the sample are unavoidable, meaning that mixed crystals are disordered systems. Since the position of the band edge depends on the composition x, these fluctuations in local x values lead to the disorder potential for electrons and holes within the crystal. To be precise, we will consider the influence of the random potential on a conduction band electron. Let $E_c(x)$ be the conduction band minimum for a crystal with composition x. In Fig. 9.5 a possible schematic dependence $E_c(x)$ is shown. If the average composition for the whole sample is x_0, the local positions of the band edge $E_c(x)$ fluctuate around the average value $E_c(x_0)$ according to the fluctuations of the composition x around x_0. For small deviations in composition Δx from the average value, one can use the linear relation

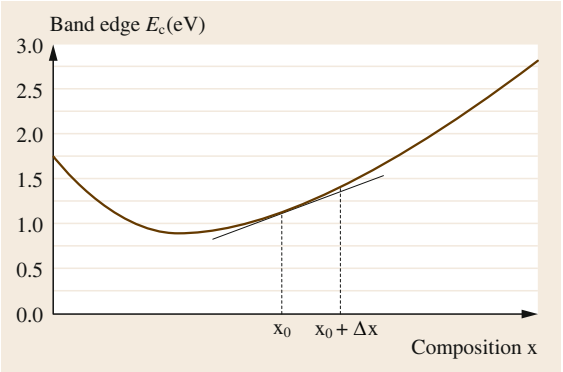


Fig. 9.5 Schematic dependence of the conduction band edge ε_c on composition x in a mixed crystal A_xB_{1-x}

$$E_c(x_0 + \Delta x) = E_c(x_0) + \alpha \Delta x \,, \tag{9.10}$$

where

$$\alpha = \left.\frac{dE_c(x)}{dx}\right|_{x=x_0} . \tag{9.11}$$

If the deviation of the concentration of A atoms from its mean value in some region of a sample is $\xi(r)$ and the total concentration of (sub)lattice sites is N, the deviation of the composition in this region is $\Delta x = \xi(r)/N$, and the potential energy of an electron at the bottom of the conduction band is

$$V(r) = \alpha \frac{\xi(r)}{N} . \tag{9.12}$$

Although one calls the disorder in such systems a *short-range* disorder, it should be noted that the consideration

is valid only for fluctuations that are much larger than the lattice constant of the material. The term *short-range* is due to the assumption that the statistical properties of the disorder are absolutely uncorrelated. This means that potential amplitudes in the adjusting spatial points are completely uncorrelated to each other. Indeed, it is usually assumed that the correlation function of the disorder in mixed crystals can be approximated by a white-noise correlation function of the form

$$\left\langle \xi(r)\xi(r')\right\rangle = x(1-x)N\delta(r-r') \,. \tag{9.13}$$

The random potential caused by such compositional fluctuations is then described by the correlation function [9.20]

$$\left\langle V(r)V(r')\right\rangle = \gamma\delta(r-r') \tag{9.14}$$

with

$$\gamma = \frac{\alpha^2}{N}x(1-x) \,. \tag{9.15}$$

Charge carriers in mixed crystals are scattered by compositional fluctuations. As is usual in kinetic descriptions of free electrons, the fluctuations on the spatial scale of the order of the electron wavelength are most efficient. Following *Shlimak* et al. [9.23], consider an isotropic quadratic energy spectrum

$$\varepsilon_p = \frac{p^2}{2m} \,, \tag{9.16}$$

where p and m are the quasi-momentum and the effective mass of an electron, respectively. The scattering rate for such an electron is

$$\nu_p = \frac{2\pi}{\hbar}\sum_q \left\langle \left|V_q\right|^2\right\rangle \left(1-\cos\vartheta_q\right)\delta\left(\varepsilon_p-\varepsilon_{p-q}\right) \,, \tag{9.17}$$

where ϑ_q is the scattering angle and

$$\left\langle \left|V_q\right|^2\right\rangle = \frac{1}{\Omega}\int \mathrm{d}^3 r \exp\left(\mathrm{i}qr\right)\left\langle V(r)V(0)\right\rangle \,. \tag{9.18}$$

The quantity Ω in this formula is the normalization volume. Using the correlation function (9.14), one obtains the relation

$$\left\langle \left|V_q\right|^2\right\rangle = \frac{\alpha^2 x(1-x)}{\Omega N} \,, \tag{9.19}$$

which shows that the scattering by compositional fluctuations is equivalent to that by a short-range potential [9.23]. Substituting (9.19) into (9.17) one obtains the following expression for the scattering rate [9.20]

$$\nu_p = \frac{\alpha^2 x(1-x)mp}{\pi\hbar^4 N} \,. \tag{9.20}$$

This formula leads to an electron mobility of the following form in the framework of the standard Drude approach [9.20, 23]

$$\mu_\mathrm{C} = \frac{\pi^{3/2}}{2\sqrt{2}}\frac{e\hbar^4 N}{\alpha^2 x(1-x)m^{5/2}(k_\mathrm{B}T)^{1/2}} \,. \tag{9.21}$$

Very similar formulae can be found in many recent publications (see for example *Fahy* and *O'Reily* [9.26]). It has also been modified and applied to two-dimensional systems [9.27] and to disordered diluted magnetic semiconductors [9.28].

It would not be difficult to apply this theoretical description to other disordered systems, provided the correlation function of the disorder potential takes the form of (9.14) with known amplitude γ. However, it is worth emphasizing that the short-range disorder of white-noise type considered here is a rather simple model that cannot be applied to most disordered materials. Therefore, we can conclude that the problem of theoretically describing charge carrier mobility via delocalized states in disordered materials is still waiting to be solved.

In the following section we present the general concepts developed to describe electrical conduction in disordered solids at temperatures where tunneling transitions of electrons between localized states significantly contribute to charge transport.

9.3 Hopping Charge Transport in Disordered Materials via Localized States

Electron transport via delocalized states above the mobility edge dominates the electrical conduction of disordered materials only at temperatures high enough to cause a significant fraction of the charge carriers fill these states. As the temperature decreases, the concentration of the electrons described by (9.9) decreases exponentially and so their contribution to electrical conductivity diminishes. Under these circumstances, tunneling transitions of electrons between localized states in the band tails dominate the charge transport in dis-

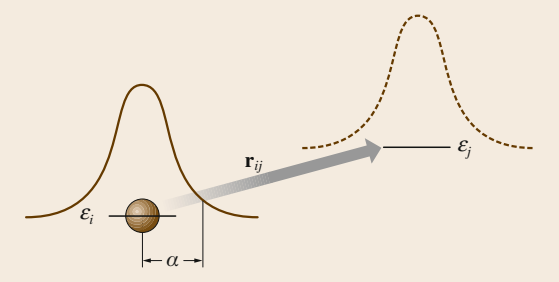


Fig. 9.6 Hopping transition between two localized states i and j with energies of ε_i and ε_j, respectively. The *solid* and *dashed lines* depict the carrier wavefunctions at sites i and j, respectively; α is the localization radius

ordered semiconductors. This transport regime is called hopping conduction, since the incoherent sequence of tunneling transitions of charge carriers resembles a series of their hops between randomly distributed sites. Each site in this picture provides a spatially localized electron state with some energy ε. In the following we will assume that the localized states for electrons (concentration N_0) are randomly distributed in space and their energy distribution is described by the DOS function $g(\varepsilon)$

$$g(\varepsilon) = \frac{N_0}{\varepsilon_0} G\left(\frac{\varepsilon}{\varepsilon_0}\right), \tag{9.22}$$

where ε_0 is the energy scale of the DOS distribution.

The tunneling transition probability of an electron from a localized state i to a localized state j that is lower in energy depends on the spatial separation r_{ij} between the sites i and j as

$$\nu_{ij}(r) = \nu_0 \exp\left(-\frac{2r_{ij}}{\alpha}\right), \tag{9.23}$$

where α is the localization length, which we assume to be equal for sites i and j. This length determines the exponential decay of the electron wavefunction in the localized states, as shown in Fig. 9.6. The pre-exponential factor ν_0 in (9.23) depends on the electron interaction mechanism that causes the transition. Usually it is assumed that electron transitions contributing to charge transport in disordered materials are caused by interactions of electrons with phonons. Often the coefficient ν_0 is simply assumed to be of the order of the phonon frequency ($\approx 10^{13}\,\mathrm{s}^{-1}$), although a more rigorous approach is in fact necessary to determine ν_0. This should take into account the particular structure of the electron localized states and also the details of the interaction mechanism [9.29, 30].

When an electron transits from a localized state i to a localized state j that is higher in energy, the transition rate depends on the energy difference between the states. This difference is compensated for by absorbing a phonon with the corresponding energy [9.31]

$$\nu_{ij}(r, \varepsilon_i, \varepsilon_j) = \nu_0 \exp\left(-\frac{2r_{ij}}{a}\right) \times \exp\left(-\frac{\varepsilon_j - \varepsilon_i + |\varepsilon_j - \varepsilon_i|}{2k_\mathrm{B}T}\right). \tag{9.24}$$

Equations (9.23) and (9.24) were written for the case in which the electron occupies site i whereas site j is empty. If the system is in thermal equilibrium, the occupation probabilities of sites with different energies are determined by Fermi statistics. This effect can be taken into account by modifying (9.24) and adding terms that account for the relative energy positions of sites i and j with respect to the Fermi energy ε_F. Taking into account these occupation probabilities, one can write the transition rate between sites i and j in the form [9.31]

$$\nu_{ij} = \nu_0 \exp\left(-\frac{2r_{ij}}{a}\right) \times \exp\left(-\frac{|\varepsilon_i - \varepsilon_\mathrm{F}| + |\varepsilon_j - \varepsilon_\mathrm{F}| + |\varepsilon_j - \varepsilon_i|}{2k_\mathrm{B}T}\right). \tag{9.25}$$

Using these formulae, the theoretical description of hopping conduction is easily formulated. One has to calculate the conductivity provided by transition events (the rates of which are described by (9.25)) in the manifold of localized states (where the DOS is described by (9.22)).

9.3.1 Nearest-Neighbor Hopping

Before presenting the correct solution to the hopping problem we would like to emphasize the following. The style of the theory for electron transport in disordered materials via localized states significantly differs from that used for theories of electron transport in ordered crystalline materials. While the description is usually based on various averaging procedures in crystalline systems, in disordered systems these averaging procedures can lead to extremely erroneous results. We believe that it is instructive to analyze some of these approaches in order to illustrate the differences between the descriptions of charge transport in ordered and disordered materials. To treat the scattering rates of electrons in ordered crystalline materials, one usually proceeds by averaging the scattering rates over the ensemble of scattering events. A similar procedure is often attempted for disordered systems too, although

various textbooks (see, for instance, *Shklovskii* and *Efros* [9.14]) illustrate how erroneous such an approach can be in the case of disordered materials.

Let us consider the simplest example of hopping processes, namely the hopping of an electron through a system of isoenergetic sites randomly distributed in space with some concentration N_0. It will be always assumed in this chapter that electron states are strongly localized and the strong inequality $N_0\alpha^3 \ll 1$ is fulfilled. In such a case the electrons prefer to hop between the spatially nearest sites and therefore this transport regime is often called nearest-neighbor hopping (NNH). This type of hopping transport takes place in many real systems at temperatures where the thermal energy k_BT is larger than the energy scale of the DOS. In such situations the energy-dependent terms in (9.24) and (9.25) do not play any significant role and the hopping rates are determined solely by the spatial terms. The rate of transition of an electron between two sites i and j is described in this case by (9.23). The average transition rate is usually obtained by weighting this expression with the probability of finding the nearest neighbor at some particular distance r_{ij}, and by integrating over all possible distances

$$\begin{aligned}\langle\nu\rangle &= \int_0^\infty \mathrm{d}r\nu_0 \\ &\quad\times \exp\left(-\frac{2r}{\alpha}\right) 4\pi r^2 N_0 \exp\left(-\frac{4\pi}{3}r^3 N_0\right) \\ &\approx \pi\nu_0 N_0\alpha^3 \,.\end{aligned} \tag{9.26}$$

Assuming that this average hopping rate describes the mobility, diffusivity and conductivity of charge carriers, one apparently comes to the conclusion that these quantities are linearly proportional to the density of localized states N_0. However, experiments evidence an exponential dependence of the transport coefficients on N_0.

Let us look therefore at the correct solution to the problem. This solution is provided in the case considered here, $N_0\alpha^3 \ll 1$, by percolation theory (see, for instance, *Shklovskii* and *Efros* [9.14]). In order to find the transport path, one connects each pair of sites if the relative separation between the sites is smaller than some given distance R, and checks whether there is a continuous path through the system via such sites. If such a path is absent, the magnitude of R is increased and the procedure is repeated. At some particular value $R = R_c$, a continuous path through the infinite system via sites with relative separations $R < R_c$ arises. Various mathematical considerations give the following relation

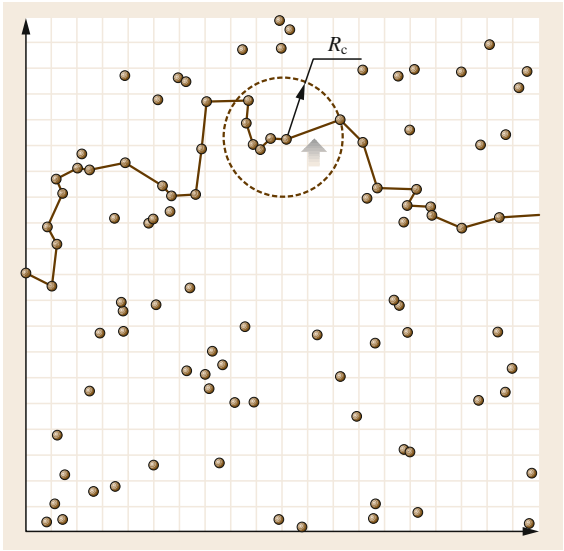


Fig. 9.7 A typical transport path with the lowest resistance. *Circles* depict localized states. The *arrow* points out the most *difficult* transition, with length R_c

for R_c [9.14]

$$\frac{4\pi}{3}N_0R_c^3 = B_c \,, \tag{9.27}$$

where $B_c = 2.7 \pm 0.1$ is the average number of neighboring sites available within a distance of less than R_c. The corresponding value of R_c should be inserted into (9.23) in order to determine kinetic coefficients such as the mobility, diffusivity and conductivity. The idea behind this procedure is as follows. Due to the exponential dependence of the transition rates on the distances between the sites, the rates for electron transitions over distances $r < R_c$ are much larger than those over distances R_c. Such fast transitions do not play any significant role as a limiting factor in electron transport and so they can be neglected in calculations of the resistivity of the system. Transitions over distances R_c are the slowest among those that are necessary for DC transport and hence such transitions determine the conductivity. The structure of the percolation cluster responsible for charge transport is shown schematically in Fig. 9.7. The transport path consists of quasi-one-dimensional segments, each containing a *difficult* transition over the distance $\approx R_c$. Using (9.23) and (9.27), one obtains the dependence of the conductivity on the concentration of localization sites in the form

$$\sigma = \sigma_0 \exp\left(-\frac{\gamma}{\alpha N_0^{1/3}}\right), \tag{9.28}$$

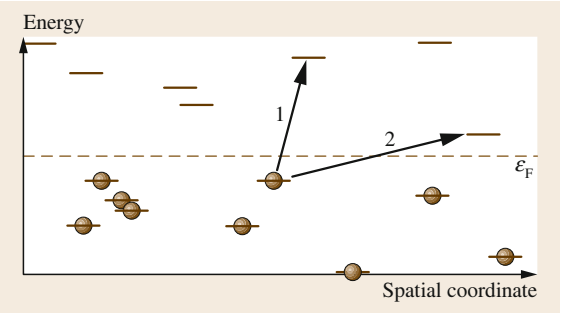


Fig. 9.8 Two alternative hopping transitions between occupied states (*filled circles*) and unoccupied states. The *dashed line* depicts the position of the Fermi level. Transitions (1) and (2) correspond to nearest-neighbor hopping and variable-range hopping regimes, respectively

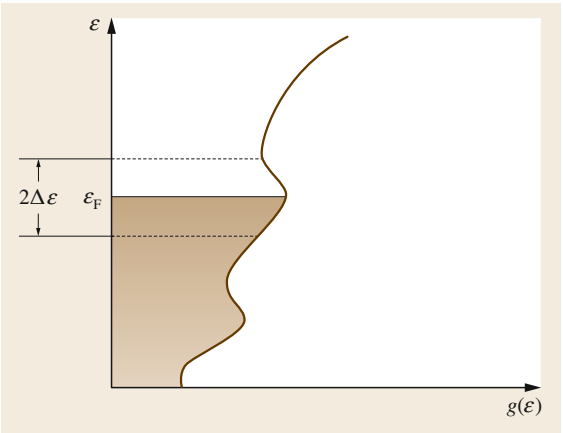


Fig. 9.9 Effective region in the vicinity of the Fermi level where charge transport takes place at low temperatures

where σ_0 is the concentration-independent pre-exponential factor and $\gamma = 1.73 \pm 0.03$. Such arguments do not allow one to determine the exponent in the kinetic coefficients with an accuracy better than a number of the order of unity [9.14]. One should note that the quantity in the exponent in (9.28) is much larger than unity for a system with strongly localized states when the inequality $N_0\alpha^3 \ll 1$ is valid. This inequality justifies the above derivation. The dependence described by (9.28) has been confirmed in numerous experimental studies of the hopping conductivity via randomly placed impurity atoms in doped crystalline semiconductors [9.14]. The drastic difference between this correct result and the erroneous one based on (9.26) is apparent. Unfortunately, the belief of many researchers in the validity of the procedure based on the averaging of hopping rates is so strong that the agreement between (9.28) and experimental data is often called occasional. We would like to emphasize once more that the ensemble averaging of hopping rates leads to erroneous results. The magnitude of the average rate in (9.26) is dominated by rare configurations of very close pairs of sites with separations of the order of the localization length α. Of course, such pairs allow very fast electron transitions, but electrons cannot move over considerable distances using only these close pairs. Therefore the magnitude of the average transition rate is irrelevant for calculations of the hopping conductivity. The correct concentration dependence of the conductivity is given by (9.28). This result was obtained under the assumption that only spatial factors determine transition rates of electrons via localized states. This regime is valid at reasonably high temperatures.

If the temperature is not as high and the thermal energy k_BT is smaller than the energy spread of the localized states involved in the charge transport process, the problem of calculating the hopping conductivity becomes much more complicated. In this case, the interplay between the energy-dependent and the distance-dependent terms in (9.24) and (9.25) determines the conductivity. The lower the temperature, the more important the energy-dependent terms in the expressions for transition probabilities of electrons in (9.24) and (9.25) become. If the spatially nearest-neighboring sites have very different energies, as shown in Fig. 9.8, the probability of an upward electron transition between these sites can be so low that it would be more favorable for this electron to hop to a more distant site at a closer energy. Hence the typical lengths of electron transitions increase with decreasing temperature. This transport regime was termed *variable-range hopping*. Next we describe several useful concepts developed to describe this transport regime.

9.3.2 Variable-Range Hopping

The concept of variable-range hopping (VRH) was put forward by Mott (see *Mott* and *Davis* [9.32]) who considered electron transport via a system of randomly distributed localized states at low temperatures. We start by presenting Mott's arguments. At low temperatures, electron transitions between states with energies in the vicinity of the Fermi level are most efficient for transport since filled and empty states with close energies can only be found in this energy range. Consider the hopping conductivity resulting from energy levels within a narrow energy strip with width $2\Delta\varepsilon$ symmetric to the Fermi level shown in Fig. 9.9. The energy width of the strip useful for electron transport can be determined from the relation

$$g(\varepsilon_F)\Delta\varepsilon r^3(\Delta\varepsilon) \approx 1 \,. \qquad (9.29)$$

This criterion is similar to that used in (9.27), although we do not care about numerical coefficients here. Here we have to consider the percolation problem in four-dimensional space since in addition to the spatial terms considered in Sect. 9.3.1 we now have to consider the energy too. The corresponding percolation problem for the transition rates described by (9.25) has not yet been solved precisely. In (9.29) it is assumed that the energy width $2\Delta\varepsilon$ is rather small and that the DOS function $g(\varepsilon)$ is almost constant in the range $\varepsilon_F \pm \Delta\varepsilon$. One can obtain the typical hopping distance from (9.29) as a function of the energy width $\Delta\varepsilon$ in the form

$$r(\Delta\varepsilon) \approx [g(\varepsilon_F)\Delta\varepsilon]^{-1/3} , \tag{9.30}$$

and substitute it into (9.24) in order to express the typical hopping rate

$$\nu = \nu_0 \exp\left(-\frac{2[g(\varepsilon_F)\Delta\varepsilon]^{-1/3}}{\alpha} - \frac{\Delta\varepsilon}{k_B T}\right) . \tag{9.31}$$

The optimal energy width $\Delta\varepsilon$ that provides the maximum hopping rate can be determined from the condition $d\nu/d\Delta\varepsilon = 0$. The result reads

$$\Delta\varepsilon = \left(\frac{2k_B T}{3g^{1/3}(\varepsilon_F)}\right)^{3/4} . \tag{9.32}$$

After substitution of (9.32) into (9.31) one obtains Mott's famous formula for temperature-dependent conductivity in the VRH regime

$$\sigma = \sigma_0 \exp\left[-\left(\frac{T_0}{T}\right)^{1/4}\right] , \tag{9.33}$$

where T_0 is the characteristic temperature

$$T_0 = \frac{\beta}{k_B g(\varepsilon_F)\alpha^3} . \tag{9.34}$$

Mott gave only a semi-quantitative derivation of (9.33), from which the exact value of the numerical constant β cannot be determined. Various theoretical studies in thee-dimensional (3-D) systems suggest values for β in the range $\beta = 10.0$ to $\beta = 37.8$. According to our computer simulations, the appropriate value is close to $\beta = 17.6$.

Mott's law implies that the density of states in the vicinity of the Fermi level is energy-independent. However, it is known that long-range electron–electron interactions in a system of localized electrons cause a gap (the so-called Coulomb gap) in the DOS in the vicinity of the Fermi energy [9.33, 34]. The gap is shown schematically in Fig. 9.10. Using simple semiquantitative arguments, *Efros* and *Shklovskii* [9.33] suggested a parabolic shape for the DOS function

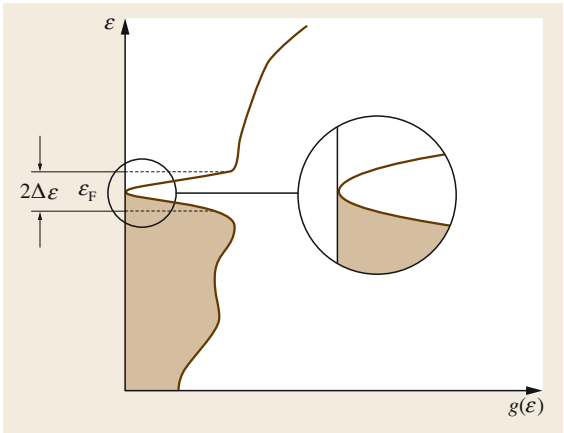


Fig. 9.10 Schematic view of the Coulomb gap. The *insert* shows the parabolic shape of the DOS near the Fermi level

$$g(\varepsilon) = \frac{\eta\,\kappa^3}{e^6}(\varepsilon - \varepsilon_F)^2 , \tag{9.35}$$

where κ is the dielectric constant, e is the elementary charge and η is a numerical coefficient. This result was later confirmed by numerous computer simulations (see, for example, *Baranovskii* et al. [9.35]). At low temperatures, the density of states near the Fermi level has a parabolic shape, and it vanishes exactly at the Fermi energy. As the temperature rises, the gap disappears (see, for example, *Shlimak* et al. [9.36]).

As we have seen above, localized states in the vicinity of the Fermi energy are the most useful for transport at low temperatures. Therefore the Coulomb gap essentially modifies the temperature dependence of the hopping conductivity in the VRH regime at low temperatures. The formal analysis of the T-dependence of the conductivity in the presence of the Coulomb gap is similar to that for the Mott's law discussed above. Using the parabolic energy dependence of the DOS function, one arrives at the result

$$\sigma = \sigma_0 \exp\left[-\left(\frac{\tilde{T}_0}{T}\right)^{1/2}\right] \tag{9.36}$$

with $\tilde{T}_0 = \tilde{\beta}e^2/(\kappa\alpha k_B)$, where $\tilde{\beta}$ is a numerical coefficient.

Equations (9.33) and (9.36) belong to the most famous theoretical results in the field of variable-range hopping conduction. However these formulae are usually of little help to researchers working with essentially

noncrystalline materials, such as amorphous, vitreous or organic semiconductors. The reason is as follows. The above formulae were derived for the cases of either constant DOS (9.33) or a parabolic DOS (9.36) in the energy range associated with hopping conduction. These conditions can usually be met in the impurity band of a lightly doped crystalline semiconductor. In the most disordered materials, however, the energy distribution of the localized states is described by a DOS function that is very strongly energy-dependent. In amorphous, vitreous and microcrystalline semiconductors, the energy dependence of the DOS function is believed to be exponential, while in organic materials it is usually assumed to be Gaussian. In these cases, new concepts are needed in order to describe the hopping conduction. In the next section we present these new concepts and calculate the way the conductivity depends on temperature and on the concentration of localized states in various significantly noncrystalline materials.

9.3.3 Description of Charge-Carrier Energy Relaxation and Hopping Conduction in Inorganic Noncrystalline Materials

In most inorganic noncrystalline materials, such as vitreous, amorphous and polycrystalline semiconductors, the localized states for electrons are distributed over a rather broad energy range with a width of the order of an electronvolt. The DOS function that describes this energy distribution in such systems is believed to have a purely exponential shape

$$g(\varepsilon) = \frac{N_0}{\varepsilon_0} \exp\left(-\frac{\varepsilon}{\varepsilon_0}\right), \tag{9.37}$$

where the energy ε is counted positive from the mobility edge towards the center of the mobility gap, N_0 is the total concentration of localized states in the band tail, and ε_0 determines the energy scale of the tail. To be precise, we consider that electrons are the charge carriers here. The result for holes can be obtained in an analogous way. Values of ε_0 in inorganic noncrystalline materials are believed to vary between 0.025 and 0.05 eV, depending on the system under consideration.

It is worth noting that arguments in favor of a purely exponential shape for the DOS in the band tails of inorganic noncrystalline materials described by (9.37) cannot be considered to be well justified. They are usually based on a rather ambiguous interpretation of experimental data. One of the strongest arguments in favor of (9.37) is the experimental observation of the exponential decay of the light absorption coefficient for photons with an energy deficit ε with respect to the energy width of the mobility gap (see, for example, *Mott* and *Davis* [9.32]). One should mention that this argument is valid only under the assumption that the energy dependence of the absorption coefficient is determined solely by the energy dependence of the DOS. However, in many cases the matrix element for electron excitation by a photon in noncrystalline materials also strongly depends on energy [9.14, 37]. Hence any argument for the shape of the DOS based on the energy dependence of the light absorption coefficient should be taken very cautiously. Another argument in favor of (9.37) comes from the measurements of dispersive transport in time-of-flight experiments. In order to interpret the observed time dependence of the mobility of charge carriers, one usually assumes that the DOS for the band tail takes the form of (9.37) (see, for example, *Orenstein* and *Kastner* [9.38]). One of the main reasons for such an assumption is probably the ability to solve the problem analytically without elaborate computer work.

In the following we start our consideration of the problem by also assuming that the DOS in a band tail of a noncrystalline material has an energy dependence that is described by (9.37). This simple function will allow us to introduce some valuable concepts that have been developed to describe dynamic effects in noncrystalline materials in the most transparent analytical form. We first present the concept of the so-called transport energy, which, in our view, provides the most transparent description of the charge transport and energy relaxation of electrons in noncrystalline materials.

The Concept of the Transport Energy

The crucial role of a particular energy level in the hopping transport of electrons via localized band-tail states with the DOS described by (9.37) was first recognized by *Grünewald* and *Thomas* [9.39] in their numerical analysis of equilibrium variable-range hopping conductivity. This problem was later considered by *Shapiro* and *Adler* [9.40], who came to the same conclusion as Grünewald and Thomas, namely that the vicinity of one particular energy level dominates the hopping transport of electrons in the band tails. In addition, they achieved an analytical formula for this level and showed that its position does not depend on the Fermi energy.

Independently, the rather different problem of nonequilibrium energy relaxation of electrons by hopping through the band tail with the DOS described by (9.37) was solved at the same time by *Monroe* [9.41]. He showed that, starting from the mobility edge, an electron most likely makes a series of hops downward in energy. The manner of the relaxation process changes at some particular energy ε_t, which Monroe called the transport energy (TE). The hopping process near and below TE resembles a multiple-trapping type

of relaxation, with the TE playing a role similar to the mobility edge. In the multiple-trapping relaxation process [9.38], only electron transitions between delocalized states above the mobility edge and the localized band-tail states are allowed, while hopping transitions between the localized tail states are neglected. Hence, every second transition brings the electron to the mobility edge. The TE of *Monroe* [9.41] coincides exactly with the energy level discovered by *Grünewald* and *Thomas* [9.39] and by *Shapiro* and *Adler* [9.40] for equilibrium hopping transport.

Shklovskii et al. [9.42] have shown that the same energy level ε_t also determines the recombination and transport of electrons in the nonequilibrium steady state under continuous photogeneration in a system with the DOS described by (9.37).

It is clear, then, that the TE determines both equilibrium and nonequilibrium and both transient and steady-state transport phenomena. The question then arises as to why this energy level is so universal that electron hopping in its vicinity dominates all transport phenomena. Below we derive the TE by considering a single hopping event for an electron localized deep in the band tail. It is the transport energy that maximizes the hopping rate as a final electron energy in the hop, independent of its initial energy [9.43]. All derivations below are carried out for the case $k_BT < \varepsilon_0$.

Consider an electron in a tail state with energy ε_i. According to (9.24), the typical rate of downward hopping of such an electron to a neighboring localized state deeper in the tail with energy $\varepsilon_j \geq \varepsilon_i$ is

$$\nu_{\downarrow}(\varepsilon_i) = \nu_0 \exp\left(-\frac{2r(\varepsilon_i)}{\alpha}\right), \tag{9.38}$$

where

$$r(\varepsilon) \approx \left[\frac{4\pi}{3}\int_{\varepsilon_i}^{\infty} g(x)\mathrm{d}x\right]^{-1/3}. \tag{9.39}$$

The typical rate of upward hopping for such an electron to a state less deep in the tail with energy $\varepsilon_j \leq \varepsilon_i$ is

$$\nu_{\uparrow}(\varepsilon_i,\, \delta) = \nu_0 \exp\left[-\frac{2r(\varepsilon_i-\delta)}{\alpha} - \frac{\delta}{k_BT}\right], \tag{9.40}$$

where $\delta = \varepsilon_i - \varepsilon_j \geq 0$. This expression is not exact. The average nearest-neighbor distance, $r(\varepsilon_i - \delta)$, is based on all states deeper than $\varepsilon_i - \delta$. For the exponential tail, this is equivalent to considering a slice of energy with a width of the order ε_0. This works for a DOS that varies slowly compared with k_BT, but not in general. It is also assumed for simplicity that the localization length, α, does not depend on energy. The latter assumption can be easily jettisoned at the cost of somewhat more complicated forms of the following equations.

We will analyze these hopping rates at a given temperature T, and try to find the energy difference δ that provides the fastest typical hopping rate for an electron placed initially at energy ε_i. The corresponding energy difference δ is determined by the condition

$$\frac{\mathrm{d}\nu_{\uparrow}(\varepsilon_i, \delta)}{\mathrm{d}\delta} = 0\,. \tag{9.41}$$

Using (9.37), (9.39) and (9.40), we find that the hopping rate in (9.40) has its maximum at

$$\delta = \varepsilon_i - 3\varepsilon_0 \ln \frac{3\varepsilon_0(4\pi/3)^{1/3}N_0^{1/3}\alpha}{2k_BT}\,. \tag{9.42}$$

The second term in the right-hand side of (9.42) is called the transport energy ε_t after *Monroe* [9.41]

$$\varepsilon_t = 3\varepsilon_0 \ln \frac{3\varepsilon_0(4\pi/3)^{1/3}N_0^{1/3}\alpha}{2k_BT}\,. \tag{9.43}$$

We see from (9.42) that the fastest hop occurs to the state in the vicinity of the TE, independent of the initial energy ε_i, provided that ε_i is deeper in the tail than ε_t; in other words, if $\delta \geq 0$. This result coincides with that of *Monroe* [9.41]. At low temperatures, the TE ε_t is situated deep in the band tail, and as the temperature rises it moves upward towards the mobility edge. At some temperature T_c, the TE merges with the mobility edge. At higher temperatures, $T > T_c$, the hopping exchange of electrons between localized band tail states becomes inefficient and the dynamic behavior of electrons is described by the well-known multiple-trapping model (see, for instance, *Orenstein* and *Kastner* [9.38]). At low temperatures, $T < T_c$, the TE replaces the mobility edge in the multiple-trapping process [9.41], as shown in Fig. 9.11. The width, W, of the maximum of the hopping rate is determined by the requirement that near ε_t the hopping rate, $\nu_{\uparrow}(\varepsilon_i, \delta)$, differs by less than a factor of e from the value $\nu_{\uparrow}(\varepsilon_i, \varepsilon_i - \varepsilon_t)$. One finds [9.42]

$$W = \sqrt{6\varepsilon_0 k_BT}\,. \tag{9.44}$$

For shallow states with $\varepsilon_i \leq \varepsilon_t$, the fastest hop (on average) is a downward hop to the nearest spatially localized state in the band tail, with the rate determined by (9.38) and (9.39). We recall that the energies of electron states are counted positive downward from the mobility edge towards the center of the mobility gap. This means that electrons in the shallow states with $\varepsilon_i \leq \varepsilon_t$ normally hop

into deeper states with $\varepsilon > \varepsilon_i$, whereas electrons in the deep states with $\varepsilon_i > \varepsilon_t$ usually hop upward in energy into states near ε_t in the energy interval W, determined by (9.44).

This shows that ε_t must play a crucial role in those phenomena, which are determined by electron hopping in the band tails. This is indeed the case, as shown in numerous review articles where comprehensive theories based on the concept of the TE can be found (see, for instance, *Shklovskii* et al. [9.42]). We will consider only one phenomenon here for illustration, namely the hopping energy relaxation of electrons in a system with the DOS described by (9.37). This problem was studied initially by *Monroe* [9.41].

Consider an electron in some localized shallow energy state close to the mobility edge. Let the temperature be low, $T < T_c$, so that the TE, ε_t, lies well below the mobility edge, which has been chosen here as a reference energy, $\varepsilon = 0$. The aim is to find the typical energy, $\varepsilon_d(t)$, of our electron as a function of time, t. At early times, as long as $\varepsilon_d(t) < \varepsilon_t$, the relaxation is governed by (9.38) and (9.39). The depth $\varepsilon_d(t)$ of an electron in the band tail is determined by the condition

$$\nu_{\downarrow}\left[\varepsilon_d(t)\right] t \approx 1 \,. \tag{9.45}$$

This leads to the double logarithmical dependence $\varepsilon_d(t) \propto \varepsilon_0 \ln[\ln(\nu_0 t)] + C$, where constant C depends on $\varepsilon_0, N_0, \alpha$ in line with (9.38) and (9.39). Indeed, (9.38) and (9.45) prescribe the logarithmic form of the time dependence of the hopping distance, $r(t)$, and (9.37) and (9.39) then lead to another logarithmic dependence $\varepsilon_d\left[r(t)\right]$ [9.41]. At the time

$$t_C \approx \nu_0^{-1} \exp\left(\frac{3\varepsilon_0}{k_B T}\right) \tag{9.46}$$

the typical electron energy, $\varepsilon_d(t)$, approaches the TE ε_t, and the style of the relaxation process changes. At $t > t_c$, every second hop brings the electron into states in the vicinity of the TE ε_t from where it falls downward in energy to the nearest (in space) localization site. In the latter relaxation process, the typical electron energy is determined by the condition [9.41]

$$\nu_{\uparrow}\left[\varepsilon_d(t), \varepsilon_t\right] t \approx 1 \,, \tag{9.47}$$

where $\nu_{\uparrow}\left[\varepsilon_d(t), \varepsilon_t\right]$ is the typical rate of electron hopping upward in energy toward the TE [9.41]. This condition leads to a typical energy position of the relaxing electron at time t of

$$\varepsilon_d(t) \approx 3\varepsilon_0 \ln\left[\ln\left(\nu_0 t\right)\right] - \varepsilon_0 \frac{8}{N_0 \alpha^3} \,. \tag{9.48}$$

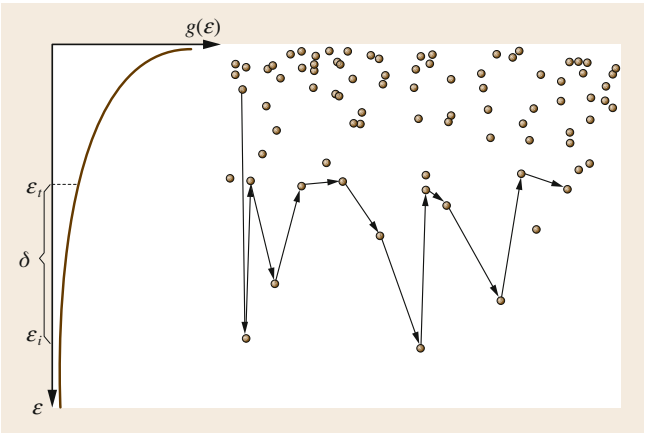


Fig. 9.11 Hopping path via the transport energy. On the *left*, the exponential DOS is shown schematically. The *right part* depicts the transport path constructed from upward and downward hops. The upward transitions bring the charge carrier to sites with energies close to the transport energy ε_t

This is a very important result, which shows that in a system where the DOS has a pure exponential energy dependence, described by (9.37), the typical energy of a set of independently relaxing electrons would drop deeper and deeper into the mobility gap with time. This result is valid as long as the electrons do not interact with each other, meaning that the occupation probabilities of the electron energy levels are not taken into account. This condition is usually met in experimental studies of transient processes, in which electrons are excited by short (in time) pulses, which are typical of time-of-flight studies of the electron mobility in various disordered materials. In this case, only a small number of electrons are present in the band tail states. Taking into account the huge number of localized band tail states in most disordered materials, one can assume that most of the states are empty and so the above formulae for the hopping rates and electron energies can be used. In this case the electron mobility is a time-dependent quantity [9.41]. A transport regime in which mobility of charge carriers is time-dependent is usually called dispersive transport (see, for example, *Mott* and *Davis* [9.32], *Orenstein* and *Kastner* [9.38], *Monroe* [9.41]). Hence we have to conclude that the transient electron mobility in inorganic noncrystalline materials with the DOS in the band tails as described by (9.37) is a time-dependent quantity and the transient electrical conductivity has dispersive character. This is due to the nonequilibrium behavior of the charge carriers. They continuously drop in energy during the course of the relaxation process.

In some theoretical studies based on the Fokker–Planck equation it has been claimed that the maximum

of the energy distribution of electrons coincides with the TE ε_t and hence it is independent of time. This statement contradicts the above result where the maximum of the distribution is at energy $\varepsilon_d(t)$, given by (9.48). The Fokker–Planck approach presumes the diffusion of charge carriers over energy. Hence it is invalid for describing the energy relaxation in the exponential tails, in which electron can move over the full energy width of the DOS (from a very deep energy state toward the TE) in a single hopping event.

In the equilibrium conditions, when electrons in the band tail states are provided by thermal excitation from the Fermi energy, a description of the electrical conductivity can easily be derived using (9.5)–(9.7) [9.39]. The maximal contribution to the integral in (9.5) comes from the electrons with energies in the vicinity of the TE ε_t, in an energy range with a width, W, described by (9.44). Neglecting the temperature dependence of the pre-exponential factor, σ_0, one arrives at the temperature dependence of the conductivity

$$\sigma \approx \sigma_0 \exp\left(-\frac{2r(\varepsilon_t)}{B_c^{-1/3}\alpha} - \frac{\varepsilon_F - \varepsilon_t}{k_B T}\right), \tag{9.49}$$

where coefficient $B_c \approx 2.7$ is inserted in order to take into account the need for a charge carrier to move over macroscopic percolation distances in order to provide low-frequency charge transport.

A very similar theory is valid for charge transport in noncrystalline materials under stationary excitation of electrons (for example by light) [9.42]. In such a case, one first needs to develop a theory for the steady state of the system under stationary excitation. This theory takes into account various recombination processes for charge carriers and provides their stationary concentration along with the position of the quasi-Fermi energy. After solving this recombination problem, one can follow the track of the theory of charge transport in quasi-thermal equilibrium [9.39] and obtain the conductivity in a form similar to (9.49), where ε_F is the position of the quasi-Fermi level. We skip the corresponding (rather sophisticated) formulae here. Interested readers can find a comprehensive description of this sort of theory for electrical conductivity in the literature (see, for instance, *Shklovskii* et al. [9.42]).

Instead, in the next section we will consider a very interesting problem related to the nonequilibrium energy relaxation of charge carriers in the band tail states. It is well known that at low temperatures, $T \leq 50$ K, the photoconductivities of various inorganic noncrystalline materials, such as amorphous and microcrystalline semiconductors, do not depend on temperature [9.44–46]. At low temperatures, the TE ε_t lies very deep in the band tail and most electrons hop downward in energy, as described by (9.38) and (9.39). In such a regime, the photoconductivity is a temperature-independent quantity determined by the loss of energy during the hopping of electrons via the band-tail states [9.47]. During this hopping relaxation, neither the diffusion coefficient D nor the mobility of the carriers μ depend on temperature, and the conventional form of Einstein's relationship $\mu = eD/(k_B T)$ cannot be valid. The question then arises as to what the relation between μ and D is for hopping relaxation. We answer this question in the following section.

Einstein's Relationship for Hopping Electrons

Let us start by considering a system of nonequilibrium electrons in the band tail states at $T = 0$. The only process that can happen with an electron is its hop downward in energy (upward hops are not possible at $T = 0$) to the nearest localized state in the tail. Such a process is described by (9.37)–(9.39). If the spatial distribution of localized tail states is isotropic, the probability of finding the nearest neighbor is also isotropic in the absence of the external electric field. In this case, the process of the hopping relaxation of electrons resembles diffusion in space. However, the median length of a hop (the distance r to the nearest available neighbor), as well as the median time, $\tau = \nu_\downarrow^{-1}(r)$, of a hop (see (9.38)) increases during the course of relaxation, since the hopping process brings electrons deeper into the tail. Nevertheless, one can ascribe a diffusion coefficient to such a process [9.42]

$$D(r) = \frac{1}{6}\nu_\downarrow(r) r^2 . \tag{9.50}$$

Here $\nu_\downarrow(r)r^2$ replaces the product of the *mean free path* r and the *velocity* $r\nu_\downarrow(r)$, and the coefficient $1/6$ accounts for the spatial symmetry of the problem. According to (9.37)–(9.39) and (9.50), this diffusion coefficient decreases exponentially with increasing r and hence with the number of successive electron hops in the relaxation process.

In order to calculate the mobility of electrons during hopping relaxation under the influence of the electric field, one should take into account the spatial asymmetry of the hopping process due to the field [9.47, 48]. Let us consider an electron in a localized state at energy ε. If an external electric field with a strength F is applied along direction x, the concentration of tail states available to this hopping electron at $T = 0$ (in other words those that have energies deeper in the tail than ε) is [9.47]

$$N(\varepsilon, x) = N(\varepsilon)\left(1 + \frac{eFx}{\varepsilon_0}\right), \tag{9.51}$$

where

$$N(\varepsilon) = \int_{\varepsilon}^{\infty} g(\varepsilon)\mathrm{d}\varepsilon = N_0 \exp\left(-\frac{\varepsilon}{\varepsilon_0}\right). \tag{9.52}$$

It was assumed in the derivation of (9.51) that $eFx \ll \varepsilon_0$.

Due to the exponential dependence of the hopping rate on the hopping length r, the electron predominantly hops to the nearest tail state among the available states if $r \gg \alpha$, which we assume to be valid. Let us calculate the average projection $\langle x\rangle$ on the field direction of the vector $\boldsymbol{r}$ from the initial states at energy ε to the nearest available neighbor among sites with a concentration $N(\varepsilon, x)$ determined by (9.51). Introducing spherical coordinates with the angle θ between r and the x-axis, we obtain [9.48]

$$\begin{aligned}\langle x\rangle = &\int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \sin\theta \\ &\times \int_0^{\infty} [\mathrm{d}r\, r^3 \cos(\theta)\, N(\varepsilon, r\cos\theta)] \\ &\times \exp\left[-\int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \sin\theta \right. \\ &\left. \times \int_0^{r} \mathrm{d}r' r'^{2N}(\varepsilon, r'\cos\theta)\right].\end{aligned} \tag{9.53}$$

Substituting (9.51) for $N(\varepsilon, r\cos\theta)$, calculating the integrals in (9.53) and omitting the second-order terms

$$\left(\frac{eN^{-1/3}(\varepsilon)F}{\varepsilon_0}\right)^2 \ll 1\,, \tag{9.54}$$

we obtain

$$\langle x\rangle = \frac{eFN^{-2/3}(\varepsilon)}{3\varepsilon_0}\frac{\Gamma(5/3)}{(4\pi/3)^{2/3}}\,, \tag{9.55}$$

where Γ is the gamma-function and $N(\varepsilon)$ is determined by (9.52). Equation (9.55) gives the average displacement in the field direction of an electron that hops downward from a state at energy ε to the nearest available neighbor in the band tail. The average length $\langle r\rangle$ of such a hop is

$$\begin{aligned}\langle r\rangle &= \int_0^{\infty} \mathrm{d}r 4\pi r^3 N(\varepsilon) \exp\left[-\frac{4\pi}{3}N(\varepsilon)r^3\right] \\ &= \left(\frac{4\pi N(\varepsilon)}{3}\right)^{-1/3} \Gamma\left(\frac{4}{3}\right).\end{aligned} \tag{9.56}$$

One can ascribe to the hopping process a mobility

$$\begin{aligned}\mu &= \frac{v}{F} = \frac{\langle x\rangle\, \nu\, (\langle r\rangle)}{F} \\ &= \frac{eN^{-2/3}(\varepsilon)\, \nu\, (\langle r\rangle)}{3\varepsilon_0}\frac{\Gamma(5/3)}{(4\pi/3)^{2/3}}\end{aligned} \tag{9.57}$$

and a diffusion coefficient

$$\begin{aligned}D &= \frac{1}{6}\langle r\rangle^2\, \nu\, (\langle r\rangle) \\ &= \frac{1}{6}N^{-2/3}(\varepsilon)\, \nu\, (\langle r\rangle)\, \frac{\Gamma^2(4/3)}{(4\pi/3)^{2/3}}\,.\end{aligned} \tag{9.58}$$

Expressions (9.57) and (9.58) lead to a relationship between μ and D of the form

$$\mu = \frac{2\Gamma(5/3)}{\Gamma^2(4/3)}\frac{e}{\varepsilon_0}D \approx 2.3\frac{e}{\varepsilon_0}D\,. \tag{9.59}$$

This formula replaces the Einstein's relationship $\mu = eD/(k_{\mathrm{B}}T)$ for electron hopping relaxation in the exponential band tail. Several points should be noted about this result. First of all, one should clearly realize that (9.59) is valid for nonequilibrium energy-loss relaxation in which only downward (in energy) transitions between localized states can occur. This regime is valid only at low temperatures when the TE ε_{t} is very deep in the band tail. As the temperature increases, the upward hops become more and more efficient for electron relaxation. Under these circumstances, the relation between μ and D evolves gradually with rising temperature from its temperature-independent form at $T = 0$ to the conventional Einstein's relationship, $\mu = eD/(k_{\mathrm{B}}T)$ [9.49, 50]. Secondly, one should realize that (9.59) was derived in the linear regime with respect to the applied field under the assumption that $eFx \ll \varepsilon_0$. According to (9.55), the quantity $\langle x\rangle$ is proportional to $N^{-2/3}(\varepsilon) = N_0^{-2/3}\exp[2\varepsilon/(3\varepsilon_0)]$, in other words it increases exponentially during the course of the relaxation toward larger localization energies ε. This means that for deep localized states in the band tail, the condition $eFx \ll \varepsilon_0$ breaks down. The boundary energy for application of the linear theory depends on the strength of the electric field, F. As F decreases, this boundary energy drops

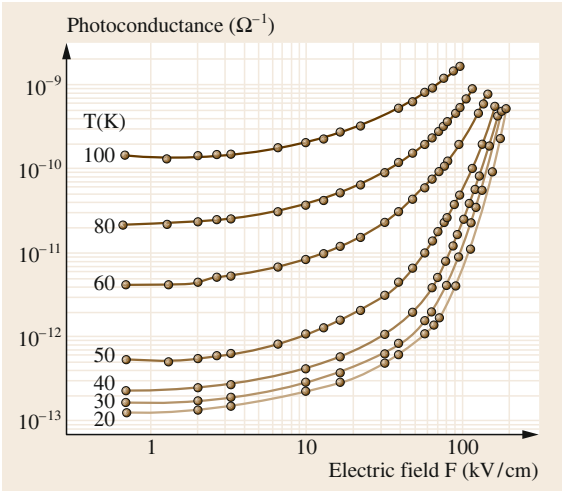


Fig. 9.12 Dependence of the photoconductivity in a-Si:H on the electric field at different temperatures. (After [9.53])

deeper into the tail. However, for any F, there is always a boundary energy in the tail below which the condition $eFx \ll \varepsilon_0$ cannot be fulfilled and where nonlinear effects play the decisive role in the hopping conduction of charge carriers. In the next section we show how one can describe these nonlinear effects with respect to the applied electric field.

Nonlinear Effects in Hopping Conduction

Transport phenomena in inorganic noncrystalline materials, such as amorphous semiconductors, under the influence of high electric fields are the foci for intensive experimental and theoretical study. This is due to observations of strong nonlinearities in the dependencies of the dark conductivity [9.11, 51, 52], the photoconductivity [9.53] and the charge carrier drift mobility [9.54–56] on the field for high electric fields. These effects are most pronounced at low temperatures, when charge transport is determined by electron hopping via localized band tail states (Fig. 9.12).

Whereas the field-dependent hopping conductivity at low temperatures has always been a challenge to describe theoretically, theories for the temperature dependence of the hopping conductivity in low electric fields have been successfully developed for all of the transport regimes discussed: for the dark conductivity [9.39], for the drift mobility [9.41], and for the photoconductivity [9.42]. In all of these theories, hopping transitions of electrons between localized states in the exponential band tails play a decisive role, as described above in (9.37)–(9.59).

Shklovskii [9.57] was the first to recognize that a strong electric field plays a similar role to that of temperature in hopping conduction. In order to obtain the field dependence of the conductivity $\sigma(F)$ at high fields, *Shklovskii* [9.57] replaced the temperature T in the well-known dependence $\sigma(T)$ for low fields by a function $T_{\text{eff}}(F)$ of the form

$$T_{\text{eff}} = \frac{eF\alpha}{2k_{\text{B}}} , \tag{9.60}$$

where e is the elementary charge, k_{B} is the Boltzmann constant, and α is the localization length of electrons in the band tail states. A very similar result was obtained later by *Grünewald* and *Movaghar* [9.58] in their study of the hopping energy relaxation of electrons through band tails at very low temperatures and high electric fields. The same idea was also used by *Shklovskii* et al. [9.42], who suggested that, at $T = 0$, one can calculate the field dependence of the stationary photoconductivity in amorphous semiconductors by replacing the laboratory temperature T in the formulae of the low-field finite-temperature theory by an effective temperature $T_{\text{eff}}(F)$ given by (9.60).

It is easy to understand why the electric field plays a role similar to that of temperature in the energy relaxation of electrons. Indeed, in the presence of the field, the number of sites available at $T = 0$ is significantly enhanced in the field direction, as shown in Fig. 9.13. Hence electrons can relax faster at higher fields. From the figure it is apparent that an electron can increase its energy with respect to the mobility edge by an amount $\varepsilon = eFx$ in a hopping event over a distance x in the direction prescribed by the electric field. The process is reminiscent of thermal activation. The analogy becomes tighter when we express the transition rate for this hop as

$$\begin{aligned} \nu &= \nu_0 \exp\left(-\frac{2x}{\alpha}\right) = \nu_0 \exp\left(-\frac{2\varepsilon}{eF\alpha}\right) \\ &= \nu_0 \exp\left(-\frac{\varepsilon}{k_{\text{B}} T_{\text{eff}}(F)}\right) , \end{aligned} \tag{9.61}$$

where $T_{\text{eff}}(F)$ is provided by (9.60).

This electric field-induced activation at $T = 0$ produces a Boltzmann tail to the energy distribution function of electrons in localized states as shown by numerical calculations [9.59, 60]. In Fig. 9.12, the field-dependent photoconductivity in a-Si:H is shown for several temperatures [9.53]. If we compare the photoconductivity at the lowest measured temperature, $T = 20\,\text{K}$ in Fig. 9.12, with the low-field photoconductivity at $T = T_{\text{eff}} = eF\alpha/(2k_{\text{B}})$ as measured by *Hoheisel* et al. [9.44] and by *Stradins* and *Fritzsche* [9.45], we come to the conclusion that the data agree quantitatively if one assumes that the localization length $\alpha = 1.05\,\text{nm}$ [9.42], which is very close to the value

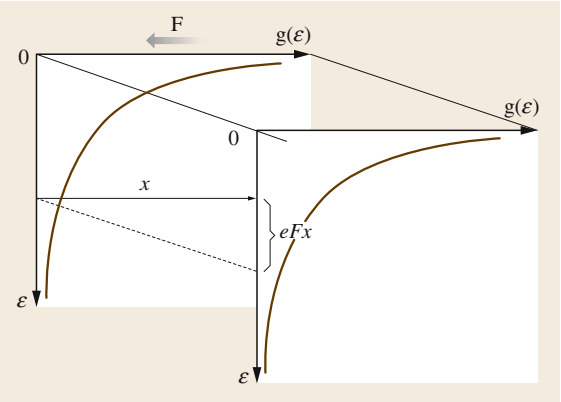


Fig. 9.13 Tunneling transition of a charge carrier in the band tail that is affected by a strong electric field. Upon traveling the distance x, the carrier acquires the energy eFx, where F is the strength of the electric field, and e is the elementary charge

$\alpha \approx 1.0\,\text{nm}$ found for a-Si:H from independent estimates [9.11]. This comparison shows that the concept of the effective temperature based on (9.60) provides a powerful tool for estimating transport coefficient nonlinearity with respect to the electric field using the low-field results for the temperature dependencies of such coefficients.

However, experiments are usually carried out not at $T = 0$ but at finite temperatures, and so the question of how to describe transport phenomena in the presence of both factors, finite T and high F, arises. By studying the steady state energy distribution of electrons in numerical calculations and computer simulations [9.59, 60], as well as straightforward computer simulations of the steady-state hopping conductivity and the transient energy relaxation of electrons [9.61], the following result was found. The whole set of transport coefficients can be represented by a function with a single parameter $T_{\text{eff}}(F, T)$

$$T_{\text{eff}}(F,T) = \left[T^{\beta} + \left(\gamma \frac{eF\alpha}{k_{\text{B}}} \right)^{\beta} \right]^{1/\beta} , \qquad (9.62)$$

where $\beta \approx 2$ and γ is between 0.5 and 0.9 depending on which transport coefficient is considered [9.61]. We are aware of no analytical theory that can support this numerical result.

To wrap up this section we would like to make the following remark. It is commonly claimed in the scientific literature that transport coefficients in the hopping regime should have a purely exponential dependence on the applied electric field. The idea behind such statements seems rather transparent. Electric field diminishes potential barriers between localized states by an amount $\Delta\varepsilon = eFx$, where x is the projection of the hopping radius on the field direction. The field should therefore diminish the activation energies in (9.24) and (9.25) by this amount, leading to the term $\exp(eFx/(k_{\text{B}}T))$ in the expressions for the charge carrier mobility, diffusivity and conductivity. One should, however, take into account that hopping transport in all real materials is essentially described by the variable-range hopping process. In such a process, as discussed above, the interplay between spatial and energy-dependent terms in the exponents of the transition probabilities determine the conduction path. Therefore it is not enough to solely take into account the influence of the strong electric field on the activation energies of single hopping transitions. One should consider the modification of the whole transport path due to the effect of the strong field. It is this VRH nature of the hopping process that leads to a more complicated field dependence for the transport coefficients expressed by (9.60)–(9.62).

We have now completed our description of electron transport in inorganic disordered materials with exponential DOS in the band tails. In the next section we tackle the problem of charge transport in organic disordered materials.

9.3.4 Description of Charge-Carrier Energy Relaxation and Hopping Conduction in Organic Noncrystalline Materials

Electron transport and energy relaxation in disordered organic solids, such as molecularly doped polymers, conjugated polymers and organic glasses, has been the subject of intensive experimental and theoretical study for more than 20 years. Although there is a wide array of different disordered organic solids, the charge transport process is similar in most of these materials. Even at the beginning of the 1980s it was well understood that the main transport mechanism in disordered organic media is the hopping of charge carriers via spatially randomly distributed localized states. Binary systems like doped polymeric matrices provide canonical examples of disordered organic materials that exhibit the hopping transport mechanism. Examples include polyvinylcarbazole (PVK) or bis-polycarbonate (Lexan) doped with either strong electron acceptors such as trinitrofluorenone acting as an electron transporting agent, or strong electron donors such as derivatives of tryphenylamine of triphenylmethane for hole transport [9.62, 63]. To avoid the need to specify whether transport is carried by electrons or holes each time, we will use a general notation of *charge carrier* below. The results are valid for both

types of carrier – electrons or holes. Charge carriers in disordered organic materials are believed to be strongly localized [9.18, 62–64]. The localization centers are molecules or molecular subunits, henceforth called sites. These sites are located in statistically different environments. As a consequence, the site energies, which are to great extent determined by electronic polarization, fluctuate from site to site. The fluctuations are typically on the order of 0.1 eV [9.65]. This is about one order of magnitude larger than the corresponding transfer integrals [9.65]. Therefore carrier wavefunctions can be considered to be strongly localized [9.65].

As discussed above, the crucial problem when developing a theoretical picture for hopping transport is the structure of the energy spectrum of localized states, DOS. It is believed that, unlike inorganic noncrystalline materials where the DOS is believed exponential, the energy dependence of the DOS in organic disordered solids is Gaussian (see *Bässler* [9.18] and references therein),

$$g(\varepsilon) = \frac{N_0}{\varepsilon_0\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2\varepsilon_0^2}\right), \tag{9.63}$$

where N_0 is the total concentration of states and ε_0 is the energy scale of the DOS. The strongest evidence in favor of such an energy spectrum in disordered organic materials is the ability to reproduce the observed experimentally temperature dependence of the carrier mobility and that of hopping conductivity assuming the Gaussian DOS in computer simulations [9.18, 66]. It has been observed in numerous experimental studies [9.67–73] that the temperature dependence of the drift mobility of charge carriers in disordered organic solids takes the form

$$\mu \propto \exp\left[-\left(\frac{T_0}{T}\right)^2\right] \tag{9.64}$$

with a characteristic temperature T_0, as shown in Fig. 9.14a. Computer simulations and theoretical calculations [9.65, 66, 74, 75] with the Gaussian DOS described by (9.63) give a dependence of the form

$$\mu \propto \exp\left[-\left(C\frac{\varepsilon_0}{k_\mathrm{B}T}\right)^2\right], \tag{9.65}$$

where C is a numerical coefficient. Computer simulations [9.65, 66] give a value $C \approx 0.69$ for this coefficient, and analytical calculations [9.74, 75] predict a similar value of $C \approx 0.64$. Equation (9.65) is often used to determine the parameter ε_0 of the DOS from experimental measurements of the $\ln(\mu)$ versus $(1/T)^2$ dependences (see, for example, *Ochse* et al. [9.71]).

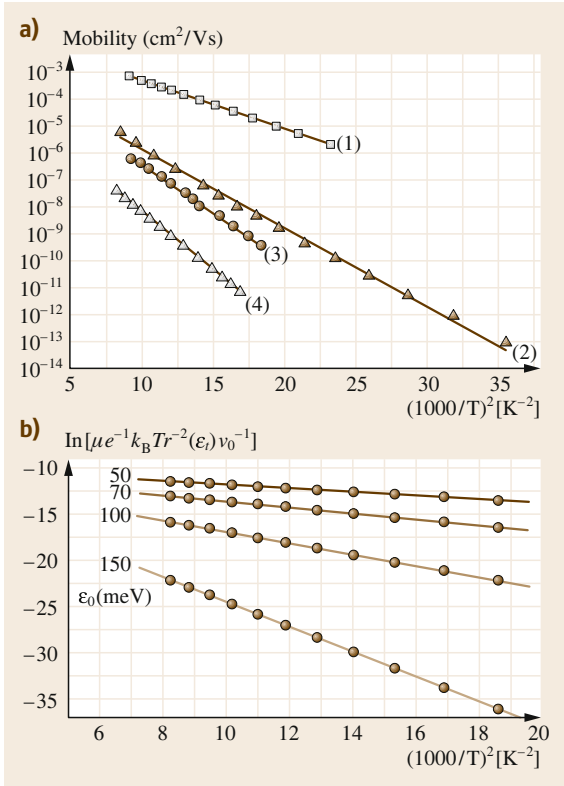


Fig. 9.14a,b Temperature dependence of the zero-field mobility in organic semiconductors. Experimental data **(a)**: (1) di-*p*-tolylphenylamine containing 1-phenyl-3-*p*-diethylaminostyryl-5-*p*-diethylamino-phenylpyrazoline (DEASP)-traps [9.69]; (2) butadiene (BD)-doped polycarbonate [9.70]; (3) naphthalenetetracarboxylic diimide (NTDI)-doped poly(styrene) [9.68]; (4) (BD)-doped TTA/polycarbonate [9.72]. Theoretical results **(b)** were obtained via (9.73)

One may wonder whether the theoretical description of hopping conduction and carrier energy relaxation in a system with a Gaussian DOS (9.63) should differ significantly from the theory described above for disordered systems with a purely exponential DOS (9.37). The answer to this question is yes. The reason becomes clear if one considers the behavior of a single charge carrier in an empty system. In an empty system with an exponential DOS, a charge carrier always (on average) falls downward in energy if $k_\mathrm{B}T < \varepsilon_0$ (see (9.45)–(9.48)), and its mobility continuously decreases with time; however, in a system with a Gaussian DOS, a particular energy level ε_∞ determines the equilibrium energy position of a charge carrier. When it is located at some site with high energy in the Gaussian DOS, the charge carrier first hops via localized states so that its average energy $\varepsilon_\mathrm{d}(t)$

decreases until it achieves the energy level ε_∞ after some typical time period τ_{rel}. At times $t < \tau_{rel}$ the behavior of the carrier qualitatively resembles that seen for the purely exponential DOS. The downward hops are then replaced by relaxation hops that send the carrier upward to the transport energy, and the carrier mobility at $t < \tau_{rel}$ decreases with time. However, in contrast with the case for the exponential DOS, in a Gaussian DOS the carrier mobility becomes time-independent after a time τ_{rel}, when the average carrier energy reaches the level ε_∞. At $t > \tau_{rel}$, the dispersive transport regime with time-dependent carrier mobility is replaced by a quasi-equilibrium so-called *Gaussian transport* regime, in which the spatial spreading of the carrier packet with time can be described by the traditional diffusion picture with a time-independent diffusion coefficient.

The peculiarity of the hopping energy relaxation of charge carriers in a system with a Gaussian DOS described above makes it easier to describe charge transport at times $t > \tau_{rel}$ than in the case of the exponential DOS. In the latter case, only the presence of a significant number of carriers in a quasi Fermi level can make kinetic coefficients such as mobility, diffusivity and conductivity time-independent and hence conventionally measurable and discussible quantities. In the case of the Gaussian DOS, these kinetic coefficients are not time-dependent at times $t > \tau_{rel}$. Moreover, in diluted systems one can calculate these coefficients by considering the behavior of a single charge carrier. This makes theoretical considerations of electrical conductivity in organic disordered solids with a Gaussian DOS much easier than when considering inorganic noncrystalline materials with an exponential DOS. Let us now calculate ε_∞, τ_{rel} and μ in disordered organic solids with a Gaussian DOS.

Computer simulations [9.66] and analytical calculations [9.74, 75] show that the mean energy of the independently hopping carriers, initially distributed randomly over all states in the Gaussian DOS, decreases with time until it approaches the thermal equilibrium value

$$\varepsilon_\infty = \frac{\left[\int_{-\infty}^{\infty} \varepsilon \exp\left(-\frac{\varepsilon}{k_B T}\right) g(\varepsilon) \mathrm{d}\varepsilon\right]}{\left[\int_{-\infty}^{\infty} \exp\left(-\frac{\varepsilon}{k_B T}\right) g(\varepsilon) \mathrm{d}\varepsilon\right]} = -\frac{\varepsilon_0^2}{k_B T} \,. \tag{9.66}$$

The time τ_{rel} required to reach this equilibrium is of key importance in the analysis of experimental data [9.65], since at $t < \tau_{rel}$ the carrier mobility decreases with time (dispersive transport) until it reaches its equilibrium, time-independent value at $t \approx \tau_{rel}$. It has been established by computer simulations that τ_{rel} strongly depends on temperature [9.18]

$$\tau_{rel} \propto \exp\left[\left(B\frac{\varepsilon_0}{k_B T}\right)^2\right] \tag{9.67}$$

with $B \approx 1.07$. Given that the same hopping processes determine both μ and τ_{rel}, researchers were puzzled for many years by the fact that they had different coefficients B and C (in other words they have different temperature dependencies) [9.65]. Below we show how to calculate both quantities – μ and τ_{rel} – easily, and we explain their temperature dependencies (obtained experimentally and by computer simulations as expressed by (9.64), (9.65) and (9.67)).

Our theoretical approach is based on the concept of transport energy (TE), introduced in Sect. 9.3.3, where it was calculated for the exponential DOS given by (9.37). Literally repeating these calculations with the Gaussian DOS, given by (9.63), we obtain the equation [9.76, 77]

$$\exp\left(\frac{x^2}{2}\right)\left[\int_{-\infty}^{\frac{x}{\sqrt{2}}} \exp(-t^2)\mathrm{d}t\right]^{4/3} = \left[9(2\pi)^{1/2} N_0 \alpha^3\right]^{-1/3} \frac{k_B T}{\varepsilon_0} \,. \tag{9.68}$$

If we denote the solution of (9.68) as $X_t(N_0\alpha^3, k_B T/\varepsilon_0)$, then the transport energy in the Gaussian DOS is equal to

$$\varepsilon_t = \varepsilon_0 X_t\left(N_0\alpha^3, \frac{k_B T}{\varepsilon_0}\right) . \tag{9.69}$$

Charge carriers perform thermally activated transitions from states with energies below the TE, ε_t, to the states with energies close to that of the TE [9.76]. Charge carriers hop downward in energy from states with energies above the TE to the spatially nearest sites with rates determined by (9.38) and (9.39).

Now that we have clarified the relaxation kinetics of charge carriers in the Gaussian DOS, it is easy to calculate the relaxation time τ_{rel} and the drift mobility μ. We consider the case $\varepsilon_\infty < \varepsilon_t < 0$, which corresponds to all reasonable values of material parameters $N_0\alpha^3$ and $k_B T/\varepsilon_0$ [9.76]. The energy relaxation of most carriers with energies ε in the interval $\varepsilon_\infty < \varepsilon < \varepsilon_t$ occurs via a multiple trapping-like process, well described in the literature (see, for example, *Orenstein* and *Kastner* [9.38] or *Marschall* [9.78]). Below ε_t the average energy of the carriers $\varepsilon(t)$ moves logarithmically downward with time t. States above $\varepsilon(t)$ achieve

thermal equilibrium with states at ε_t at time t, while states below $\varepsilon(t)$ have no chance at time t to exchange carriers with states in the vicinity of ε_t. Hence the occupation of those deep states does not correspond to the equilibrium one, being determined solely by the DOS of the deep states. The system reaches thermal equilibrium when the time-dependent average energy $\varepsilon(t)$ achieves the equilibrium level ε_∞, determined by (9.66). This happens at $t = \tau_{rel}$. Since the relaxation of carriers occurs via thermal activation to the level ε_t, the relaxation time τ_{rel} is determined by the time required for activated transitions from the equilibrium level ε_∞ to the transport energy ε_t. Hence, according to (9.40) and (9.47), τ_{rel} is determined by the expression

$$\tau_{rel} = \nu_0^{-1} \exp\left[\frac{2r(\varepsilon_t)}{\alpha} + \frac{\varepsilon_t - \varepsilon_\infty}{k_B T}\right]. \qquad (9.70)$$

From (9.68)–(9.70) it is obvious that the activation energy of the relaxation time depends on the parameters $N_0\alpha^3$ and k_BT/ε_0. Hence, generally speaking, this dependence cannot be represented by (9.67) and, if at all, the coefficient B should depend on the magnitude of the parameter $N_0\alpha^3$. However, numerically solving (9.68)–(9.70) using the value $N_0\alpha^3 = 0.001$, which was also used in computer simulations by *Bässler* [9.18, 65], confirms the validity of (9.67) with $B \approx 1.0$. This result is in agreement with the value $B \approx 1.07$ obtained from computer simulations [9.18, 65]. A way to describe the temperature dependence of the relaxation time τ_{rel} by (9.67) is provided by the strong temperature dependence of ε_∞ in the exponent in (9.70), while the temperature dependencies of the quantities ε_t and $r(\varepsilon_t)$ in (9.70) are weaker and they almost cancel each other out. However, if $N_0\alpha^3 = 0.02$, the relaxation time is described by (9.67) with $B \approx 0.9$. This shows that (9.67) can only be considered to be a good approximation.

Now we turn to the calculation of the carrier drift mobility μ. We assume that the transition time t_{tr} necessary for a carrier to travel through a sample is longer than τ_{rel}, and hence the charge transport takes place under equilibrium conditions. As described above, every second jump brings the carrier upward in energy to the vicinity of ε_t, and is then followed by a jump to the spatially nearest site with deeper energy, determined solely by the DOS. Therefore, in order to calculate the drift mobility μ, we must average the hopping transition times over energy states below ε_t, since only these states are essential to charge transport in thermal equilibrium [9.77, 79]. Hops downward in energy from the level ε_t occur exponentially faster than upward hops to-

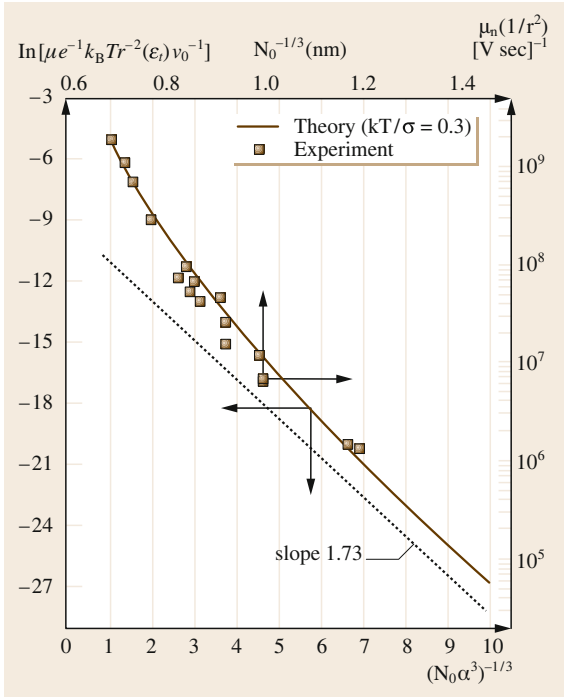


Fig. 9.15 Concentration dependence of the drift mobility evaluated from (9.73) (*solid line*), and the dependence observed experimentally (*circles*) for trinitrofluorenone with polyvinylcarbazole (TNF/PVK). (After [9.63])

wards ε_t. This means that one can neglect the former in the calculation of the average time $\langle t \rangle$. The carrier drift mobility can be evaluated as

$$\mu \approx \frac{e}{k_B T} \frac{r^2(\varepsilon_t)}{\langle t \rangle}, \qquad (9.71)$$

where $r(\varepsilon_t)$ is determined via (9.39), (9.63), (9.68) and (9.69). The average hopping time takes the form [9.79]

$$\langle t \rangle = \left[\int_{-\infty}^{\varepsilon_t} g(\varepsilon)\mathrm{d}\varepsilon\right]^{-1} \times \int_{-\infty}^{\varepsilon_t} \nu_0^{-1} g(\varepsilon) \times \exp\left(\frac{2r(\varepsilon_t)B_c^{1/3}}{a} + \frac{\varepsilon_t - \varepsilon}{k_B T}\right) \mathrm{d}\varepsilon, \qquad (9.72)$$

where $B_c \approx 2.7$ is the percolation parameter. This numerical coefficient is introduced into (9.72) in order to warrant the existence of an infinite percolation path over the states with energies below ε_t. Using (9.63), (9.68), (9.69), (9.71) and (9.72), one obtains the following relation for the exponential terms in the expression for the

carrier drift mobility

$$\ln\left[\frac{\mu}{\left(\frac{er^2(\varepsilon_t)\nu_0}{k_BT}\right)}\right] = -2\left[\frac{4\sqrt{\pi}}{3B_c}N_0\alpha^3\int_{-\infty}^{X_t/\sqrt{\pi}}\exp(-t^2)\mathrm{d}t\right]^{-1/3} - \frac{X_t\varepsilon_0}{k_BT} - \frac{1}{2}\left(\frac{\varepsilon_0}{k_BT}\right)^2 . \tag{9.73}$$

It is (9.73) that determines the dependence of the carrier drift mobility on the parameters $N_0\alpha^3$ and k_BT/ε_0.

In Fig. 9.14b, the dependence of the drift mobility on the temperature at $N_0\alpha^3 = 0.01$ is depicted for several values of ε_0. The sensitivity of the mobility to temperature is clear from this picture. Comparison of these dependencies with experimental measurements of $\ln(\mu)$ versus $(1/T)^2$ (some are shown in Fig. 9.14a) provides information on the energy scale, ε_0, of the DOS (see, for example, *Bässler* [9.18] and *Ochse* et al. [9.71]).

In Fig. 9.15, the dependence of the drift mobility on $N_0\alpha^3$ is shown for $k_BT/\varepsilon_0 = 0.3$. Experimental data from *Gill* [9.80] are also shown in the figure. It is clear that the slope of the mobility exponent as a function of $(N_0\alpha^3)^{-1/3}$ given by the theory described above agrees with the experimental data. At a very low concentration of localized states, N_0, when the probability of carrier tunneling in space dominates the transition rate in (9.24), charge carriers hop preferentially to the nearest spatial sites. In this regime of nearest-neighbor hopping, the concentration dependence of the drift mobility is described by (9.28), as illustrated by the dashed line in Fig. 9.15.

So far we have discussed the drift mobility of charge carriers under the assumption that the concentration of charge carriers is much less than that of the localized states in the energy range relevant to hopping transport. In such a case one can assume that the carriers perform independent hopping motion and so the conductivity can be calculated as the product

$$\sigma = en\mu , \tag{9.74}$$

where n is the concentration of charge carriers in the material and μ is their drift mobility. If, however, the concentration n is so large that the Fermi energy at thermal equilibrium or the quasi-Fermi energy at stationary excitation is located significantly higher (energetically) than the equilibrium energy ε_∞, a more sophisticated theory based on the percolation approach is required [9.81]. The result obtained is similar to that given by (9.49).

9.4 Concluding Remarks

Beautiful effects have been observed experimentally by studying the charge transport in disordered organic and inorganic materials. Among these, the transport coefficients in the hopping regime show enormously strong dependencies on material parameters. The dependence of the charge carrier mobility on the concentration of localized states N_0 (Fig. 9.15) spreads over many orders of magnitude, as does its dependence on the temperature T (Fig. 9.14) and on the (high) electric field strength F (Fig. 9.12). Such strong variations in physical quantities are typical, say, in astrophysics, but they are not usual in solid state physics. This makes the study of the charge transport in disordered materials absolutely fascinating. The strong dependencies of kinetic coefficients (like drift mobility, diffusivity and conductivity) in disordered materials on various material parameters makes these systems very attractive for various device applications. Since they are relatively inexpensive to manufacture too, it is then easy to understand why disordered organic and inorganic materials are of enormous interest for various technical applications.

These materials also provide a purely academic challenge with respect to their transport phenomena. While traditional kinetic theories developed for crystalline materials are largely dependent on the systems having translational symmetry, there is no such symmetry in disordered materials. However, we have shown in this chapter that it is still possible to develop a reliable theoretical approach to transport phenomena in disordered materials. Particularly interesting is the hopping transport regime. In this regime, charge carriers perform incoherent tunneling jumps between localized states distributed in space and energy. The enormously strong (exponential) dependence of the transition rates on the distances between the sites and their energies call for a completely new set of ideas compared to those for crystalline solids. Conventional transport theories based on the averaging of transition rates lead to absurd results if applied to hopping transport in disordered materials.

One can use ideas from percolation theory instead to adequately describe charge transport. One of the most important ideas in this field is so-called variable-range hopping (VRH) conduction. Although the rate of transitions between two localized states is a product of exponential terms that are separately dependent on the concentration of localized states N_0, the temperature of the system T, and also on the field strength F (for high field strengths), it is generally wrong to assume that the carrier drift mobility, diffusivity or conductivity can also be represented as the product of three functions that are separately dependent on N_0, T and F. Instead one should search for a percolation path that takes into account the exponential dependences of the hopping rates on all of these parameters simultaneously. Such a procedure, based on strong interplay between the important parameters in the exponents of the transition rates, leads to very interesting and (in some cases) unexpected results, some of which were described in this chapter. For example, it was shown that the effect of a strong electric field on transport coefficients can be accounted for by renormalizing the temperature. Most of the ideas discussed in this chapter were discussed in the early works of Mott and his coauthors (see, for example, *Mott* and *Davis* [9.32]). Unfortunately, these ideas are not yet known to the majority of researchers working in the field of disordered materials. Moreover, it is often believed that transport phenomena in different disordered materials need to be described using different ideas. Mott based his ideas, in particular the VRH, mostly on inorganic glassy semiconductors. Most of the researchers that are studying amorphous inorganic semiconductors (like a-Si:H) are aware of these ideas. However, new researchers that are working on more modern disordered materials, such as organic disordered solids and dye-sensitized materials, are often not aware of these very useful and powerful ideas developed by Mott and his followers that can be used to describe charge transport in inorganic disordered systems. In this chapter we have shown that the most pronounced charge transport effects in inorganic and organic disordered materials can be successfully described in a general manner using these ideas.

Although we have presented some useful ideas for describing charge transport in disordered systems above, it is clear that the theoretical side of this field is still embyonic. There are still no reliable theories for charge transport via extended states in disordered materials. Nor are there any reliable theoretical descriptions for the spatial structure of the localized states (DOS) in organic and inorganic noncrystalline materials. All of the theoretical concepts presented in this chapter were developed using very simple models of localization centers with a given energy spectrum that are randomly distributed in space. No correlations between the spatial positions of the sites and the energies of the electronic states at these sites were considered here. Some theoretical attempts to account for such correlations can be found in the literature, although the correlations have not been calculated ab initio: instead they are inserted into a framework of model assumptions. This shows how far the field of charge transport in disordered materials is from a desirable state. Since these materials are already widely used in various technical applications, such as field transistor manufacture, light-emitting diodes and solar cells, and since the sphere of such applications is increasing, the authors are optimistic about the future of research in this field. The study of fundamental charge transport properties in disordered materials should develop, leading us to a better understanding of the fundamental charge transport mechanisms in such systems.

Acknowledgments. The authors are indebted to numerous colleagues for stimulating and enlightening discussions. Among those are Boris Shklovski (University of Minnesota), Alexei Efros and Michael Raikh (Utah University), Hellmut Fritzsche (Chicago University), Peter Thomas, Walther Fuhs and Heinz Bässler (Philipps-University Marburg), Igor Zvyagin (Moscow State University) and many other colleagues. Financial support of the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

9.1 A. Bunde, K. Funke, M.D. Ingram: Solid State Ion. **105**, 1 (1998)
9.2 S.D. Baranovskii, H. Cordes: J. Chem. Phys. **111**, 7546 (1999)
9.3 C. Brabec, V. Dyakonov, J. Parisi, N.S. Sariciftci: *Organic Photovoltaics: Concepts and Realization* (Springer, Berlin, Heidelberg 2003)
9.4 M.H. Brodsky: *Amorphous Semiconductors* (Springer, Berlin, Heidelberg 1979)
9.5 G. Hadziioannou, P.F. van Hutten: *Semiconducting Polymers* (Wiley, New York 2000)
9.6 J.D. Joannopoulos, G. Locowsky: *The Physics of Hydrogenated Amorphous Silicon I* (Springer, Berlin, Heidelberg 1984)

9.7 J.D. Joannopoulos, G. Locowsky: *The Physics of Hydrogenated Amorphous Silicon II* (Springer, Berlin, Heidelberg 1984)
9.8 A. Madan, M.P. Shaw: *The Physics and Applications of Amorphous Semiconductors* (Academic, New York 1988)
9.9 M. Pope, C.E. Swenberg: *Electronic Processes in Organic Crystals and Polymers* (Oxford Univ. Press, Oxford 1999)
9.10 J. Singh, K. Shimakawa: *Advances in Amorphous Semiconductors* (Gordon Breach/Taylor Francis, London 2003)
9.11 R.A. Street: *Hydrogenated Amorphous Silicon*, Cambridge Solid State Science (Cambridge Univ. Press, Cambridge 1991)
9.12 K. Tanaka, E. Maruyama, T. Shimada, H. Okamoto: *Amorphous Silicon* (Wiley, New York 1999)
9.13 J.S. Dugdale: *The Electrical Properties of Disordered Metals*, Cambridge Solid State Science (Cambridge Univ. Press, Cambridge 1995)
9.14 B.I. Shklovskii, A.L. Efros: *Electronic Properties of Doped Semiconductors* (Springer, Berlin, Heidelberg 1984)
9.15 I.P. Zvyagin: *Kinetic Phenomena in Disordered Semiconductors* (Moscow Univ. Press, Moscow 1984), in Russian
9.16 H. Böttger, V.V. Bryksin: *Hopping Conduction in Solids* (Wiley, New York 1985)
9.17 H. Overhof, P. Thomas: *Electronic Transport in Hydrogenated Amorphous Semiconductors* (Springer, Berlin, Heidelberg 1989)
9.18 H. Bässler: Phys. Status Solidi B **175**, 15 (1993)
9.19 P.W. Anderson: Phys. Rev. **109**, 1492 (1958)
9.20 A.L. Efros, M.E. Raikh: Effects of composition disorder on the electronic properties of semiconducting mixed crystals. In: *Optical Properties of Mixed Crystals*, ed. by R.J. Elliott, I.P. Ipatova (Elsevier, New York 1988)
9.21 D. Chattopadhyay, B.R. Nag: Phys. Rev. B **12**, 5676 (1975)
9.22 J.W. Harrison, J.R. Hauser: Phys. Rev. B **13**, 5347 (1976)
9.23 I.S. Shlimak, A.L. Efros, I.V. Yanchev: Sov. Phys. Semicond. **11**, 149 (1977)
9.24 S.D. Baranovskii, A.L. Efros: Sov. Phys. Semicond. **12**, 1328 (1978)
9.25 P.K. Basu, K. Bhattacharyya: J. Appl. Phys. **59**, 992 (1986)
9.26 S. Fahy, E.P. O'Reily: Appl. Phys. Lett. **83**, 3731 (2003)
9.27 V. Venkataraman, C.W. Liu, J.C. Sturm: Appl. Phys. Lett. **63**, 2795 (1993)
9.28 C. Michel, P.J. Klar, S.D. Baranovskii, P. Thomas: Phys. Rev. B **69**, 165211 (2004)
9.29 T. Holstein: Philos. Mag. B **37**, 49 (1978)
9.30 H. Scher, T. Holstein: Philos. Mag. **44**, 343 (1981)
9.31 A. Miller, E. Abrahams: Phys. Rev. **120**, 745 (1960)
9.32 N.F. Mott, E.A. Davis: *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford 1971)
9.33 A.L. Efros, B.I. Shklovskii: J. Phys. C **8**, L49 (1975)
9.34 M. Pollak: Disc. Faraday Soc. **50**, 13 (1970)
9.35 S.D. Baranovskii, A.L. Efros, B.L. Gelmont, B.I. Shklovskii: J. Phys. C **12**, 1023 (1979)
9.36 I. Shlimak, M. Kaveh, R. Ussyshkin, V. Ginodman, S.D. Baranovskii, H. Vaupel, P. Thomas, R.W. van der Heijden: Phys. Rev. Lett. **75**, 4764 (1995)
9.37 S.D. Baranovskii, P. Thomas, G.J. Adriaenssens: J. Non-Cryst. Solids **190**, 283 (1995)
9.38 J. Orenstein, M.A. Kastner: Solid State Commun. **40**, 85 (1981)
9.39 M. Grünewald, P. Thomas: Phys. Status Solidi B **94**, 125 (1979)
9.40 F.R. Shapiro, D. Adler: J. Non-Cryst. Solids **74**, 189 (1985)
9.41 D. Monroe: Phys. Rev. Lett. **54**, 146 (1985)
9.42 B.I. Shklovskii, E.I. Levin, H. Fritzsche, S.D. Baranovskii: Hopping photoconductivity in amorphous semiconductors: Dependence on temperature, electric field and frequency. In: *Advances in Disordered Semiconductors*, Vol. 3, ed. by H. Fritzsche (World Scientific, Singapore 1990) p. 3161
9.43 S.D. Baranovskii, F. Hensel, K. Ruckes, P. Thomas, G.J. Adriaenssens: J. Non-Cryst. Solids **190**, 117 (1995)
9.44 M. Hoheisel, R. Carius, W. Fuhs: J. Non-Cryst. Solids **63**, 313 (1984)
9.45 P. Stradins, H. Fritzsche: Philos. Mag. **69**, 121 (1994)
9.46 J.-H. Zhou, S.D. Baranovskii, S. Yamasaki, K. Ikuta, K. Tanaka, M. Kondo, A. Matsuda, P. Thomas: Phys. Status Solidi B **205**, 147 (1998)
9.47 B.I. Shklovskii, H. Fritzsche, S.D. Baranovskii: Phys. Rev. Lett. **62**, 2989 (1989)
9.48 S.D. Baranovskii, T. Faber, F. Hensel, P. Thomas, G.J. Adriaenssens: J. Non-Cryst. Solids **198–200**, 214 (1996)
9.49 S.D. Baranovskii, T. Faber, F. Hensel, P. Thomas: Phys. Status Solidi B **205**, 87 (1998)
9.50 S.D. Baranovskii, T. Faber, F. Hensel, P. Thomas: J. Non-Cryst. Solids **227–230**, 158 (1998)
9.51 A. Nagy, M. Hundhausen, L. Ley, G. Brunst, E. Holzenkämpfer: J. Non-Cryst. Solids **164–166**, 529 (1993)
9.52 C.E. Nebel, R.A. Street, N.M. Johanson, C.C. Tsai: Phys. Rev. B **46**, 6803 (1992)
9.53 R. Stachowitz, W. Fuhs, K. Jahn: Philos. Mag. B **62**, 5 (1990)
9.54 H. Antoniadis, E.A. Schiff: Phys. Rev. B **43**, 13957 (1991)
9.55 K. Murayama, H. Oheda, S. Yamasaki, A. Matsuda: Solid State Commun. **81**, 887 (1992)
9.56 C.E. Nebel, R.A. Street, N.M. Johanson, J. Kocka: Phys. Rev. B **46**, 6789 (1992)
9.57 B.I. Shklovskii: Sov. Phys. Semicond. **6**, 1964 (1973)
9.58 M. Grünewald, B. Movaghar: J. Phys. Condens. Matter **1**, 2521 (1989)
9.59 S.D. Baranovskii, B. Cleve, R. Hess, P. Thomas: J. Non-Cryst. Solids **164–166**, 437 (1993)
9.60 S. Marianer, B.I. Shklovskii: Phys. Rev. B **46**, 13100 (1992)
9.61 B. Cleve, B. Hartenstein, S.D. Baranovskii, M. Scheidler, P. Thomas, H. Baessler: Phys. Rev. B **51**, 16705 (1995)
9.62 M. Abkowitz, M. Stolka, M. Morgan: J. Appl. Phys. **52**, 3453 (1981)
9.63 W.D. Gill: J. Appl. Phys. **43**, 5033 (1972)
9.64 S.J. Santos Lemus, J. Hirsch: Philos. Mag. B **53**, 25 (1986)

9.65 H. Bässler: Advances in disordered semiconductors. In: *Hopping and Related Phenomena*, Vol. 2, ed. by M. Pollak, H. Fritzsche (World Scientific, Singapore 1990) p. 491
9.66 G. Schönherr, H. Bässler, M. Silver: Philos. Mag. B **44**, 369 (1981)
9.67 P.M. Borsenberger, H. Bässler: J. Chem. Phys. **95**, 5327 (1991)
9.68 P.M. Borsenberger, W.T. Gruenbaum, E.H. Magin, S.A. Visser: Phys. Status Solidi A **166**, 835 (1998)
9.69 P.M. Borsenberger, W.T. Gruenbaum, E.H. Magin, S.A. Visser, D.E. Schildkraut: J. Polym. Sci. Polym. Phys. **37**, 349 (1999)
9.70 A. Nemeth-Buhin, C. Juhasz: Hole transport in 1,1-bis(4-diethylaminophenyl)-4,4-diphenyl-1,3-butadiene. In: *Hopping and Related Phenomena*, ed. by O. Millo, Z. Ovadyahu (Racah Institute of Physics, The Hebrew Univ. Jerusalem, Jerusalem 1995) p. 410
9.71 A. Ochse, A. Kettner, J. Kopitzke, J.-H. Wendorff, H. Bässler: Chem. Phys. **1**, 1757 (1999)
9.72 J. Veres, C. Juhasz: Philos. Mag. B **75**, 377 (1997)
9.73 U. Wolf, H. Bässler, P.M. Borsenberger, W.T. Gruenbaum: Chem. Phys. **222**, 259 (1997)
9.74 M. Grünewald, B. Pohlmann, B. Movaghar, D. Würtz: Philos. Mag. B **49**, 341 (1984)
9.75 B. Movaghar, M. Grünewald, B. Ries, H. Bässler, D. Würtz: Phys. Rev. B **33**, 5545 (1986)
9.76 S.D. Baranovskii, T. Faber, F. Hensel, P. Thomas: J. Phys. C **9**, 2699 (1997)
9.77 S.D. Baranovskii, H. Cordes, F. Hensel, G. Leising: Phys. Rev. B **62**, 7934 (2000)
9.78 J.M. Marshall: Rep. Prog. Phys. **46**, 1235 (1983)
9.79 O. Rubel, S.D. Baranovskii, P. Thomas, S. Yamasaki: Phys. Rev. B **69**, 014206-1 (2004)
9.80 W.D. Gill: Electron mobilities in disordred and crystalline tritrofluorenone. In: *Proc. Fifth Int. Conf. of Amorphous and Liquid Semiconductors*, ed. by J. Stuke, W. Brenig (Taylor Francis, London 1974) p. 901
9.81 S.D. Baranovskii, I.P. Zvyagin, H. Cordes, S. Yamasaki, P. Thomas: Phys. Status Solidi B **230**, 281 (2002)

10. Dielectric Response

Leonard Dissado

Nearly all materials are dielectrics, and the measurement of their dielectric response is a very common technique for their characterisation. This chapter is intended to guide scientists and engineers through the subject to the point where they can interpret their data in terms of the microscopic and atomistic dynamics responsible for the dielectric response, and hence derive useful information appropriate to their particular needs. The focus is on the physical concepts underlying the observed behaviour and is developed from material understandable by an undergraduate student. Emphasis is placed on the information content in the data, and the limits to be placed on its interpretation are clearly identified.

Generic forms of behaviour are identified using examples typical of different classes of material, rather than an exhaustive review of the literature. Limited-range charge transport is included as a special item. The theoretical concepts are developed from a basic level up to the ideas current in the field, and the points where these are controversial have been noted so that the readers can choose for themselves how far to rely on them.

10.1 **Definition of Dielectric Response** 220
10.1.1 Relationship to Capacitance 220
10.1.2 Frequency-Dependent Susceptibility ... 220
10.1.3 Relationship to Refractive Index 221
10.2 **Frequency-Dependent Linear Responses** 222
10.2.1 Resonance Response 222
10.2.2 Relaxation Response 224
10.3 **Information Contained in the Relaxation Response** 228
10.3.1 The Dielectric Increment for a Linear Response χ_0 228
10.3.2 The Characteristic Relaxation Time (Frequency) 231
10.3.3 The Relaxation Peak Shape 237
10.4 **Charge Transport** 240
10.5 **Data Presentation** 243
10.6 **A Few Final Comments** 243
References 244

Nearly all materials are dielectrics, that is they do not exhibit a direct-current (DC) conductivity on the macroscopic scale, but instead act as an electrical capacitance i. e., they store charge. The measurement of the dielectric response is noninvasive and has been used for material characterisation throughout most of the 20th century, and consequently a number of books already exist that cover the technique from various points of view. Those that have stood the test of time are *Debye* [10.1], *Smyth* [10.2], *McCrum* et al. [10.3], *Daniels* [10.4], *Bottcher* and *Bordewijk* [10.5], and *Jonscher* [10.6]. These texts cover the subject in terms of the basic physics [10.1, 5], the material properties [10.2–4], and the electrical features [10.6]. An introduction to the wide range of dielectric response measurements that are undertaken can be obtained by referring to the proceedings publication of the International Discussion Meeting on Relaxations in Complex Systems [10.7]. In view of the enormous range of properties and materials covered by the topic it is not feasible or desirable to attempt to review the whole field in a chapter such as this. Instead the topic is approached from the viewpoint of a researcher who, having measured the dielectric *spectrum* (i. e., frequency-dependent complex permittivity) of a material sample, wishes to know what information can be taken from the measurements. Along the way the limits on the information content and the problems (and controversies) associated with the microscopic and molecular-scale interpretation will be identified. Emphasis will be placed on the physical concepts involved, but inevitably there will be some mathematical expressions whose features I aim to place in as simple a physical context as possible.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_10

10.1 Definition of Dielectric Response

10.1.1 Relationship to Capacitance

Stated simply the dielectric response of a material is its response to an electric field within it. Let us start in the simplest way with some basic macroscopic definitions. Imagine that we have a flat slab of our material of thickness d with electrodes of area A on each opposing surface, and we apply a potential difference between opposing sides (Fig. 10.1). Since the material is a dielectric our system is a capacitor and electric charges $\pm Q$ will be stored on the surfaces between which the potential difference V is applied with Q given by

$$Q = CV\,, \tag{10.1}$$

where C is the capacitance of the system. For the sample geometry considered C is given by the expression

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d}\,. \tag{10.2}$$

Here the factor ε_0 is a fundamental constant termed the permittivity of free space ($\varepsilon_0 = 8.854 \times 10^{-12}\,\mathrm{F\,m^{-1}}$) and ε_r is the permittivity of the material relative to that of free space and hence is termed the relative permittivity. Of course it is a pure number, which is a material property that contains information about the way the material responds to the application of the potential difference. Since however, the replacement of our sample by a vacuum yields a capacitance given by

$$C = \frac{\varepsilon_0 A}{d}\,, \tag{10.3}$$

i. e., $\varepsilon_r = 1$, the information about the response of the material to the electric field is contained in a factor χ,

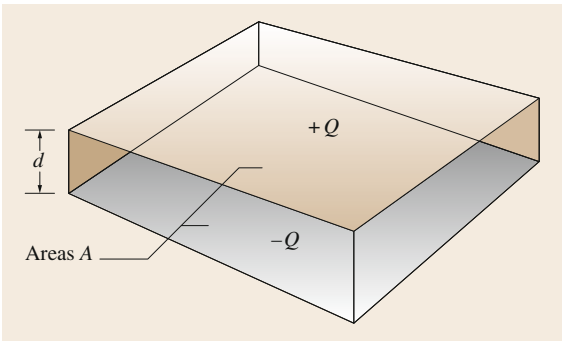


Fig. 10.1 A schematic representation of a dielectric in the form of a parallel-sided slab of surface area A and thickness d that has acquired surface charges $+Q$ and $-Q$ as a result of a potential difference applied across it. (After [10.8])

which is called the susceptibility of the material and is given by

$$\chi = \varepsilon_0(\varepsilon_r - 1)\,. \tag{10.4}$$

Noting that the electric field for the parallel electrode geometry described is given by $E = V/d$ and that the electrostatic flux density (electric displacement field) D within the material is $D = Q/A$, allows (10.1) and (10.2) to be transformed into the familiar constitutive relationship,

$$D = \varepsilon_0 \varepsilon_r E = \varepsilon_0 E + \chi E \tag{10.5}$$

which is valid irrespective of the geometry. Here the response of the material has been explicitly separated out in the form of an additive term that is called the electric polarisation P, which has dimensions of charge/area or more familiarly electric dipole moment density (electric dipole moment/volume), and is given by

$$P = \chi E\,. \tag{10.6}$$

It can be seen from (10.5) and (10.6) that the charges that correspond to P are the *additional* charges (Q_i) that are induced on the electrodes when a dielectric material replaces a vacuum with the same field applied.

An electric dipole occurs when electric charges of opposite polarity are separated in space and the magnitude of the dipole moment is given by $\boldsymbol{\mu} = \boldsymbol{r}\delta$, where δ is the magnitude of the charge and $\boldsymbol{r}$ is the separation vector. The connection between P as defined above and the dipole moment density can be easily seen once it is recognised that the separation of the charges $\pm Q_i$ by the inter-electrode separation d constitutes a macroscopic dipole moment of magnitude $Q_i d$ contained within the volume Ad. The definition of the polarisation as the dipole moment density makes it clear that the dielectric response to an electric field relates to the generation of a net dipole per unit volume in the material. In most cases the dielectric has a zero net dipole moment in the absence of an applied electric field and the action of the electric field is to produce one. Some materials such as electrets [10.9] and ferroelectrics lack a centre of symmetry, however, and possess a nonzero polarisation in the absence of an electric field. Therefore the definition of the polarisation P in (10.6) that is applicable to all cases is that P is the *change in net dipole moment density* produced by the applied field.

10.1.2 Frequency-Dependent Susceptibility

The outline given above has been written as if the electric field within the material was a static field ($f = 0$)

but (10.6) is equally valid if the field oscillates with an angular frequency ω ($\omega = 2\pi f$, where f is the frequency in Hz). In this case (10.6) becomes

$$P(\omega) = \chi(\omega)E(\omega)\,, \tag{10.7}$$

where $E(\omega) = E_0 \exp(\mathrm{i}\omega t)$.

The fundamental reason for the dependence of $P(\omega)$ upon the frequency of the alternating-current (AC) field, as in (10.7), can be envisaged by constructing a general picture of the way that a material responds to an electric field. Let us imagine that we have our material in thermal equilibrium in the absence of an electric field and we switch on a constant field at a specified time. The presence of the electric field causes the generation of a net dipole moment density (or change in one already existing). This alteration in the internal arrangement of positive and negative charges will not be instantaneous. Instead it will develop according to some equation of motion appropriate to the type of charges and dipole moments that are present. Consequently some time will be required before the system can come into equilibrium with the applied field. Formally this time will be infinity (equivalent to an AC frequency of zero), but to all intents and purposes we can regard the system as coming into equilibrium fairly rapidly after some relevant time scale, τ, with the polarisation approaching the static value $P = P(0)$ for $t \gg \tau$. If now we think of the electric field as reversing sign before equilibrium is reached, such reversal being the case for an AC field at a time $t = 1/(2f)$ after it is switched on, it is clear that the polarisation will not have reached its equilibrium value before the field is reversed and hence that $P(\omega) \lesssim P(0)$, and $\chi(\omega) \lesssim \chi(0)$. The frequency dependence of the dielectric susceptibility $\chi(\omega)$ is therefore determined by the equation of motion governing the evolution of the ensemble of electric dipole moments following excitation.

In general $\chi(\omega)$ will be a complex function with a real component $\chi'(\omega)$ defining the component of $P(\omega)$ that is in phase with the applied AC field $E(\omega) = \mathrm{Re}[E_0 \exp(\mathrm{i}\omega t)] = E_0 \cos(\omega t)$, and $\chi''(\omega)$ defining the component that is 90° out of phase. The conventional form is given by

$$\chi(\omega) = \chi'(\omega) - \mathrm{i}\chi''(\omega) \left[\mathrm{i} = \sqrt{-1}\right]\,. \tag{10.8}$$

It is easy to see that $\chi'(\omega)$ determines the net separation of charge with the dielectric in the form of a macroscopic capacitor, but the nature of $\chi''(\omega)$ is not so obvious. The answer lies in considering the *rate of change of polarisation*, $\mathrm{d}[P(\omega)]/\mathrm{d}t$. This has the dimensions of a current density (current/area), is sometimes termed the polarisation current density, and is given by,

$$\begin{aligned}\mathrm{d}[P(\omega)]/\mathrm{d}t &= \left[\chi'(\omega) - \mathrm{i}\chi''(\omega)\right]\mathrm{d}[E(\omega)]/\mathrm{d}t \\ &= [\chi'(\omega) - \mathrm{i}\chi''(\omega)]\mathrm{i}\omega[\cos(\omega t) \\ &\quad + \mathrm{i}\sin(\omega t)]E_0\,.\end{aligned} \tag{10.9}$$

Thus $\chi''(\omega)$ determines the real component of the polarisation current density that is in phase with the electric field, i. e., $J_{\mathrm{pol}}(\omega)$ given by

$$\begin{aligned}J_{\mathrm{pol}}(\omega) &= \chi''(\omega)\omega E_0 \cos(\omega t) \\ &= \sigma_{\mathrm{AC}}(\omega)E_0 \cos(\omega t)\,.\end{aligned} \tag{10.10}$$

Here $\chi''(\omega)\omega = \sigma_{\mathrm{AC}}(\omega)$ is the contribution to the AC conductivity due to the polarisation response to the electric field. If we remember Joule's Law for the power dissipated thermally by an electric current, i. e., power lost $= IV$, then we can see that $\chi''(\omega)\omega(E_0)^2$ is the power dissipated per unit volume resulting from the generation of a net polarisation by the electric field, i. e., the power dissipation density. The imaginary susceptibility $\chi''(\omega)$ is often termed the power dissipation component. It arises because the electric field has to carry out work on the dielectric in order to produce a net dipole moment density. Some of this energy is stored in the charge separations and is recoverable in an equivalent way to the elastic energy stored in a spring. The rest of the energy is used to overcome the friction opposing the establishment of the net dipole density. This energy is transferred to the dielectric in an unrecoverable way, i. e., it is dissipated within the dielectric. It can be seen that $\chi''(\omega)$ is dependent upon the form of the equations of motion governing the evolution of the net dipole moment density under the action of an electric field.

10.1.3 Relationship to Refractive Index

Equation (10.7) can be regarded as relating to the polarisation response purely to an oscillating electric field, but of course all electromagnetic waves contain such a field. In general the topic of dielectric response includes the response of the material to the electric field component of an electromagnetic field, i. e., the electromagnetic spectrum of a material is a form of dielectric response. This form of response is generally characterised by a complex frequency-dependent refractive index $n^*(f)$, with

$$n^*(f) = n(f) - \mathrm{i}\kappa(f)\,, \tag{10.11}$$

where n is the real refractive index expressing the velocity of light in the material, v, as $v = c/n$, and κ

is the absorption coefficient, which defines the reduction of intensity of light of frequency f as it passes through the medium due to absorption of the photons by the medium, i. e., it relates to energy dissipated from the electromagnetic wave. The absorption coefficient can be determined from the Beer–Lambert law, which defines the reduction in intensity as the wave travels a distance z

$$I = I_0 \exp\left(-\frac{4\pi\kappa f z}{c}\right). \tag{10.12}$$

For nonmagnetic materials an equivalence can be established between n^* and the relative permittivity (see [10.10] for example). If we think of two slabs of dielectric material placed in contact with an electromagnetic wave passing through them, the component of the field D perpendicular to the boundary is unchanged as the wave passes from one medium to another. Essentially the same electrostatic flux passes through the same perpendicular area. Equation (10.5) gives the ratio of the electric fields in the two dielectric media as the inverse of the ratio of their relative permittivities. A comparison with Snell's law [10.10, 11] then yields

$$\varepsilon_r = [n^*(f)]^2 . \tag{10.13}$$

Measurement of the dielectric response of a material involves the determination of the polarisation and its frequency dependence in some form or other. This can be carried out in a large number of ways such as the absorption spectra of electromagnetic radiation as has been described above as well as the application of an AC electric field across a sample of defined dimensions. Those readers interested in the technical details of the measurement systems are referred to the general literature, which has an enormous number of works on these experimental techniques (see for example [10.12] for bridge techniques and [10.13] for recent microwave techniques, and [10.14] for broadband dielectric spectroscopy). This chapter will have a different focus. It will in essence be a discussion of the microscopic origins of the polarisation $P(\omega)$ and the physical reasons for its variation in frequency so that data obtained from such measurements can be used to gain information relevant to the nature of the material to be studied.

10.2 Frequency-Dependent Linear Responses

In this context a linear response is one in which $P(\omega)$ is only dependent upon the first power of the electric field, i. e., $\chi(\omega)$ is independent of the electric field. The general form of frequency dependence expected for $\varepsilon(\omega) = \varepsilon_0 + \chi(\omega)$ is shown schematically in (Fig. 10.2), where it can be seen that two basic types of response can be distinguished: a resonance response at high frequencies in the quantum region, and what is termed a relaxation response at lower frequencies.

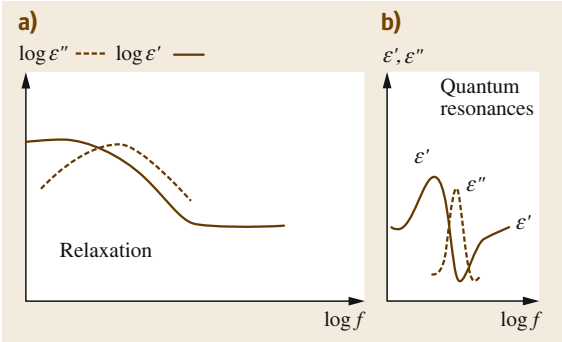


Fig. 10.2a,b Schematic representation of the frequency dependence of **(a)** relaxation and **(b)** resonance responses

10.2.1 Resonance Response

Although our main topic in this chapter will be the relaxation behaviour I will start with the resonance response as this has been dealt with extensively [10.15] from the spectroscopic viewpoint. Here we shall approach it from the perspective of its identity as a dielectric response with the intention of identifying basic features that also occur in relaxation responses. The equation of motion is familiar: this form of response relates to a *net* electric dipole moment, ϕ, that undergoes damped simple harmonic oscillation at a natural oscillation frequency $\Omega = 2\pi\nu$ in the absence of an electric field,

$$\frac{d^2\phi}{dt^2} + \frac{\gamma d\phi}{dt} + \Omega^2\phi = 0 . \tag{10.14}$$

The interpretation of (10.14) is not as obvious as it seems. We have to remember that we are always dealing with a sample of material that contains an enormous number of molecules or molecular moieties. Two types of situation may lead to this form of equation of motion. In one case a fluctuation in the positions of *groups* of positive and negative charges may form an oscillating

net dipole moment. Plasma oscillations, found in metals [10.16], are an example of this type of behaviour. In the other case, individual molecules may possess identical dipoles oscillating independently at a frequency Ω. Equation (10.14) then describes the behaviour of a *net* dipole moment produced by a fluctuation in the system of molecules independent of an applied electric field. Such fluctuations are deviations from the average equilibrium state produced by random impulses that act only for an infinitesimal time at $t = 0$, and (10.14) describes the subsequent evolution of the fluctuation. In both cases the damping (friction) term $\gamma \mathrm{d}\phi/\mathrm{d}t$, where γ is the damping factor, expresses the way in which the fluctuation dies away in the absence of an applied electric field and the net dipole moment density of the system returns to zero.

The situation applying in the case of most spectroscopic responses relates to dipole density fluctuations in which individual molecules contribute oscillating dipoles, and it is this form of behaviour that will be used to illustrate the resonance form of response. The first thing we have to do is to ask what kinds of dipole moments oscillate in our material at frequencies in the range $\nu \geq 5 \times 10^{10}$ Hz where most resonance responses are found. The answer to this question lies in the electronic structure of the atoms, ions and molecules that make up our material. The uncertainty principle states the energy of a system cannot be specified at any precise instant of time. This means that the electronic structure of atoms and molecules is allowed to continuously fluctuate between quantum states as long as the average energy over a period of time remains constant. The averaging time is determined by the uncertainty relationship between energy and time. The molecule is thus continuously moving back and forth between ground state and its excited states. These quantum fluctuations (between ground and excited states) displace the negative charge of the electron cloud with respect to the positive charge of the nucleus, and produce oscillating dipoles, termed transition dipoles (Fig. 10.3a). Their oscillation frequencies have specific values depending upon the energy difference between excited and ground quantum states of the species concerned, typically in the range $\nu \geq 4 \times 10^{14}$ Hz [10.16, 17]. These oscillating dipolar fluctuations give no net contribution to the dipole moment of the molecule or atom, i. e., no permanent dipole moment. The uncertainty principle similarly allows the nuclei of molecules, and the molecules themselves to vibrate and fluctuate between different vibration states. In many cases these nuclear fluctuations produce dipole moments [10.10, 17, 18] (Fig. 10.3b), which typically oscillate in the frequency range 4×10^{14} Hz $\geq \nu \geq 5 \times 10^{10}$ Hz. As with the electronic transition dipoles, these vibration fluctuations do not contribute to the permanent dipole moment of the molecule or system of molecules.

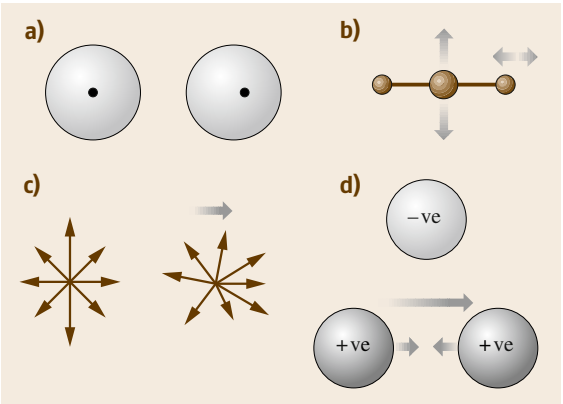


Fig. 10.3a–d Schematic representation of the various forms of dipole fluctuations: **(a)** atomic dipole, **(b)** dipoles formed by bending and stretching motion of a tri-atomic molecule of the form A–B–A, **(c)** net dipole fluctuation produced in a system of permanent dipoles, **(d)** reorienting dipole formed by transfer of an ion between two different centres of vibration. In all cases the *light arrow* shows the net dipole

When an AC electric field, such as that provided by an electromagnetic wave, is applied to the material, it couples to the transition dipoles to give a force that drives the system. As a result the population of dipoles is altered such that there is now a *net* average density of transition dipole moments. This oscillates with the frequency of the *field*, i. e., the system is polarised as defined by (10.7). Since transition dipoles do not contribute to a permanent dipole moment of a molecule, the net dipole moment density produced by the electric field is referred to as an induced dipole moment density. The response of the system to an electric field oscillating at a frequency ω is essentially obtained by adding the driving force to (10.14) and determining the solution for ϕ that oscillates with the same frequency, i. e., $\phi(t) \propto \mathrm{e}^{\mathrm{i}\omega t}$. The components of the frequency-dependent relative susceptibility, $\chi_\mathrm{r}(\omega) = \chi(\omega)/\varepsilon_0$, then take the form

$$\chi'_\mathrm{r}(\omega) = \frac{\chi_0 \Omega^2 \left(\Omega^2 - \omega^2\right)}{\left(\Omega^2 - \omega^2\right)^2 + \omega^2 \gamma^2} , \tag{10.15}$$

$$\chi''_\mathrm{r}(\omega) = \frac{\chi_0 \Omega^2 \omega \gamma}{\left(\Omega^2 - \omega^2\right)^2 + \omega^2 \gamma^2} . \tag{10.16}$$

These equations exhibit the typical signature of a resonance response such as shown in Fig. 10.2. The imaginary component $\chi''_\mathrm{r}(\omega)$ goes through a peak at $\omega = \Omega$ (the resonance frequency). The real component $\chi'_\mathrm{r}(\omega)$ exhibits a rise in positive value as the driving frequency

Ω approaches the natural frequency Ω from below, passes through zero when ω and Ω are equal, and rises towards zero from a negative value when $\omega \gg \Omega$. In some cases the exact frequency dependence of the relative susceptibility can be slightly different from that given above. The width of the peak in $\chi_r''(\omega)$ is determined by the damping constant γ. In addition the peak in $\chi_r''(\omega)$ may be broadened due to the possibility that either the resonance frequencies of different transitions of the same molecule can be close together and their responses can overlap, or that local electrical interactions between molecules cause the transition energies of individual molecules to be slightly shifted in energy giving a continuum of transitions around an average value.

As long as the transition energies are sufficiently far apart to be resolved experimental data of the resonance type will yield three pieces of information, which can be related to the electronic (or vibration) structure of the molecules. These are:

1. The natural oscillation frequency Ω
2. The damping constant γ
3. The amplitude factor χ_0.

The natural frequency ν is equal to the energy difference of the electronic states between which the fluctuation occurs, ξ, divided by Planck's constant i. e., $\nu = \xi/h$, and so provides information about the different quantum states in relation to one another. The amplitude factor χ_0 is proportional to the square of the transition dipole and therefore yields information on the relative rearrangement of positive and negative charges within the molecule by the transition fluctuation. Damping in these types of systems arises from the sharing of the transition energy between many energy states of the molecule and its vibrations. It removes energy from the specific oscillating dipoles for which the field produces a net dipole moment density. It may act through a delay in returning energy to the electromagnetic wave, i. e., incoherent reradiation, or by transferring it to other energy states where it cannot be reradiated. The damping therefore expresses the way that the energy transferred from the electromagnetic field to the molecule is absorbed and dissipated in the system. The damping factor γ often will have a complicated form. There is of course one other piece of information that is implied by data that fit (10.15)–(10.16) and that is that the equation of motion for the natural oscillating dipole moments is given by (10.14). In some cases, however, the damping factor may be frequency dependent as a result of changes in the interaction between different energy states of the system that occur on the same time scale as the relaxation time, $1/\gamma$. The equation of motion will now have a different and more complicated form than that of (10.14).

The above outline of spectroscopic responses is of necessity very sketchy as it is not the main theme of this chapter, and is dealt with in detail in many standard textbooks, for example *Heitler* [10.15]. There are however, a number of general features that can be used as a guide to what happens in the *linear* relaxation response. In the first place the dipoles involved are properties of natural fluctuations of the system, in this case quantum fluctuations in molecules. They are not produced by the electric field. In the absence of an electric field the fluctuating dipoles do not contribute to the net dipole moment of the system, in this case individual molecules. The action of the electric field in linear responses is solely to alter the population of the fluctuations such that a net dipole moment density is produced. This is achieved in the resonance cases considered above by the production of a net density of molecules in an excited state proportional to $\chi_0 E_0$. The irreversible transfer of energy from the electric field to the system relates to the sharing of this energy between the oscillating dipoles coupled to the electric field, and many equi-energetic states of the system that do not couple directly to the electric field. The energy shared in this way is dissipated among the many connected states. Dissipation is an essential consequence of natural fluctuations in an ensemble [10.19] and expresses the requirement that the fluctuation die away to zero at long times. The function $\phi(t)$ must therefore approach zero as t tends to infinity. In the absence of an electric field dipole density fluctuations utilise energy gained transiently from the ensemble and return that energy via the dissipation mechanism. When however an electric field is present, the relative number of fluctuations with dipole moments in different directions is altered and the dissipation term irreversibly transfers energy from the electric field to the ensemble.

10.2.2 Relaxation Response

We turn now to the relaxation response. The simplest way to view this behaviour is as an overdamped oscillation of the net dipole moment density, i. e., one for which $\gamma^2 > 4\Omega^2$. There are a number of ways of addressing this situation and below I shall develop the description starting from the simplest model whose behaviour is rarely found in condensed matter.

The Debye Response

In this case we can neglect the force constant term in (10.14), i. e., the term $\Omega^2\phi$. This leads to an equation of motion with the form

$$\frac{d\phi}{dt} + \gamma\phi = 0\,. \tag{10.17}$$

The solution to this equation is the very familiar exponential form, $\phi(t) \propto \mathrm{e}^{-t/\tau}$. Equation (10.17) can be interpreted by taking on board the lessons from the resonance response. As before we have to view it as describing the behaviour of a natural fluctuation in our system that produces a net dipole moment density as the result of a random impulse at $t = 0$. Now however, there is no evidence for dipole oscillation, so we are not looking at the quantum fluctuations of electronic charge clouds and nuclei positions of molecules. In this case the response originates with the permanent dipoles that many molecules possess due to the asymmetry of their atomic construction. We should also remember that, though atoms do not possess a permanent dipole moment, ion pairs in a material will act as dipoles. Such systems contain a large number (ensemble) of permanent dipoles and this ensemble will obey the laws of thermodynamics. Therefore, with the exception of such materials as electrets and ferroelectrics the orientation of the permanent dipoles will be random in the absence of an electric field, i. e., the *average* net dipole moment of the system will be zero. Thermodynamic ensembles are however described by distributions that allow for fluctuations about the defined average values, thus for example canonical ensembles allow for fluctuations in energy about a defined average energy content, and grand canonical ensembles allow for fluctuations in the number of effective units (e.g., net dipole moments) as well. In the case of dipole responses we are looking at fluctuations that involve the orientations of the permanent dipoles and hence create a net dipole density (Fig. 10.3c,d). Such fluctuations are natural to the ensemble, but are transient, i. e., as in Sect. 10.2.1 $\phi(t) \to 0$ as $t \to \infty$. Equation (10.17) describes the way in which such a local fluctuation in the dipole moment density decays (regresses) to zero, i. e., the ensemble relaxes. An applied electric field couples with the permanent dipoles to produce a torque that attempts to align the dipole with the electric field vector where its energy is lowest. Consequently the linear response of the system to the application of an electric field is an increase in the population of the permanent dipole fluctuations with a component oriented in the field direction as compared to those which have components oriented in the reverse direction. This relative change in the populations of the natural fluctuations of the system gives a net dipole moment density that is driven at the frequency of the electric field as in (10.7) [10.20].

As in the resonance case the polarisation can be obtained by adding the AC driving force oscillating at frequency ω to (10.17) and determining the solution for ϕ oscillating with the same frequency. The corresponding relative susceptibility components have the form

$$\chi'_{\mathrm{r}}(\omega) = \frac{\chi_0}{1+\omega^2/\gamma^2}\,, \tag{10.18}$$

$$\chi''_{\mathrm{r}}(\omega) = \frac{\chi_0\omega/\gamma}{1+\omega^2/\gamma^2}\,. \tag{10.19}$$

These functions show a peak in the imaginary susceptibility component, $\chi''_{\mathrm{r}}(\omega)$, at a frequency $\Omega = \gamma$, which is sometimes called the loss peak frequency since $\chi''_{\mathrm{r}}(\omega)$ is associated with the dissipation of energy, or equivalently the loss of energy from the driving electric field. The real component of the susceptibility, $\chi'_{\mathrm{r}}(\omega)$, changes monotonically from zero at high frequencies to a limiting low-frequency value of χ_0. This is termed the dielectric dispersion.

Equations (10.17)–(10.19) define what has come to be known as the Debye response after *P. Debye* who first addressed the nature of relaxation dielectric responses [10.1]. It is characterised by two pieces of information: the magnitude of the dispersion χ_0 and the damping factor γ, more usually defined via the relaxation time $\tau = 1/\gamma$ of the dipole density fluctuations. The dispersion magnitude χ_0 is a measure of the net change in dipole density fluctuations that can be produced by a unit field (i. e., $E_0 = 1\,\mathrm{V\,m^{-1}}$), and is proportional to the square of the individual permanent dipole moments. An exact fit between the relaxation response data and (10.18), (10.19) implies a specific form for the equation of motion of the dipole density fluctuations of the permanent dipole ensemble, i. e., that of (10.17) in the same way that a fit to (10.15)–(10.16) implies a fit to the damped oscillator equation of motion (10.14) in the resonance response.

Frequency-Dependent Dielectric Response in Condensed Matter

In practice the Debye response is rarely observed outside of the gas phase. Instead the experimental data can usually be characterised through fractional power laws in the frequency dependence of $\chi''_{\mathrm{r}}(\omega)$ [10.6, 8, 21] in the regions away from the peak (Fig. 10.4), i. e., for $\omega \gg \gamma$, and $\omega \ll \gamma$, giving

$$\chi'_{\mathrm{r}}(\omega) \propto \chi''_{\mathrm{r}}(\omega) \propto \omega^{n-1}, \omega \gg \gamma\,, \tag{10.20}$$

$$\chi'_{\mathrm{r}}(0) - \chi'_{\mathrm{r}}(\omega) \propto \chi''_{\mathrm{r}}(\omega) \propto \omega^{m}, \omega \ll \gamma\,. \tag{10.21}$$

Here n, and m are fractional exponents, i. e., $0 < n, m < 1$. This general form was first defined empirically as the *Havriliak–Negami* function [10.22, 23]. A number of special cases have been identified [10.5, 8]. Thus for example the Cole–Cole function is given by $n + m = 1$. When $m = 1$, and $0 < n < 1$, the Cole–Davidson form is produced, which obeys (10.20) and

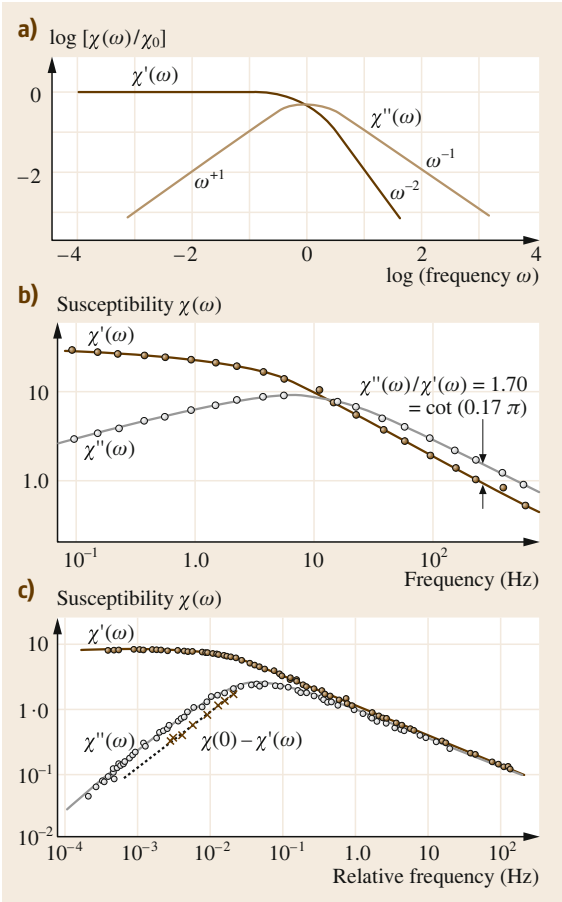


Fig. 10.4 **(a)** Schematic drawing of the Debye response. **(b)**, **(c)** examples of measured data fitted to the response function $\chi(\omega)/\chi(0) \propto [(1+\mathrm{i}\omega\tau)^{n-1}\,{}_2F_1(1-n;1-m;2-n;(1+\mathrm{i}\omega\tau)^{-1})]$ resulting from the response function $\phi(t)$ of (10.22). The function ${}_2F_1(,;;)$ is the Gaussian hypergeometric function (After [10.24]). *Plot* **(b)** data from irradiated tri-glycine sulphate (TGS); *Plot* **(c)** data from polyvinylacetate. (After [10.25])

follows $\chi''_{\mathrm{r}}(\omega) \propto \omega^m$ $(m=1)$ for $\omega \ll \gamma$ but without $\chi''_{\mathrm{r}}(\omega)$ being proportional to $\chi'_{\mathrm{r}}(0)-\chi'_{\mathrm{r}}(\omega)$ as in (10.21).

The physical theory of *Dissado* and *Hill* [10.26, 27] yields analytical expressions for $\chi'_{\mathrm{r}}(\omega)$ and $\chi''_{\mathrm{r}}(\omega)$ that contain both the Debye and Cole–Davidson functions as *exact* limiting cases, and the Cole–Cole and Havriliak–Negami functions as approximations with only minor differences in the curvature in the region of the peak in $\chi''_{\mathrm{r}}(\omega)$. The frequency-dependent susceptibility given by this theory has come to be known as the Dissado–Hill function and is defined in terms of a hypergeometric function [10.24], which is an infinite series with similarities to that of the exponential function. The reader is referred to [10.26, 27] for details.

Unlike the empirical functions the Dissado–Hill dielectric response function has a clearly defined equation of motion for $\phi(t)$ [10.27],

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}t^2} + \frac{\left[2+n+\left(\frac{t}{\tau}\right)\right]}{t}\frac{\mathrm{d}\phi}{\mathrm{d}t} + \frac{\left[n+\left(\frac{t}{\tau}\right)(1+m)\right]}{t^2}\phi = 0\,. \tag{10.22}$$

This equation has an analytical solution for $\phi(t)$ in terms of a confluent hypergeometric function [10.24], which has the limiting behaviour $\phi(t) \propto t^{-n}$ at times $t < \tau = 1/\gamma$ and $\phi(t) \propto t^{-(1+m)}$ when $t > \tau = 1/\gamma$. When $m=1$ the solution is $\phi(t) \propto t^{-n}\exp(-t/\tau)$ giving the Cole–Davidson susceptibility function, and the Debye response function $\phi(t) \propto \exp(-t/\tau)$ is obtained when $n=0$ and $m=1$. The general analytical function for $\phi(t)$ was first derived in [10.28] and later re-derived and its equation of motion discussed in [10.26, 27] to which the reader is referred for details. The form of (10.22) illustrates clearly the development of the relaxation response from short-time high-frequency damped harmonic oscillations, characterised by (10.14), as a consequence of time-dependent damping functions and oscillation frequencies. More specifically the damping function approaches a constant value $1/\tau$ (equivalent to γ) at long times, while the oscillation frequency approaches zero. In contrast the Debye model assumes that the dipole motions are overdamped motions with no time-dependent transition from the damped oscillations occurring at short times. Of course there are many ways that time-dependent damping functions and oscillation frequencies may be introduced, most of which involve making similar approximations to that of Debye as regards the ensemble forces that control these factors, but only the specific forms of (10.22) give the power-law frequency dependencies observed in the susceptibility (10.20), (10.21). These forms are produced because the dipole density fluctuations retain some memory of the restoring and damping forces that act on it over very long periods of time, i. e., these forces are not random impulses. The reader is referred to [10.27] for more detail. An alternative equation of motion for the polarisation has been obtained in terms of a fractional time derivative [10.29–31], which is a differ-integral operator involving a memory kernel with the form of a decaying fractional time power law [10.29]. Although this form of equation of motion cannot be related directly to an oscillator equation, the fractional time power law of the memory kernel must have time reversal symmetry at short time [10.19] and hence it should develop from coupled oscillators as in [10.26, 27].

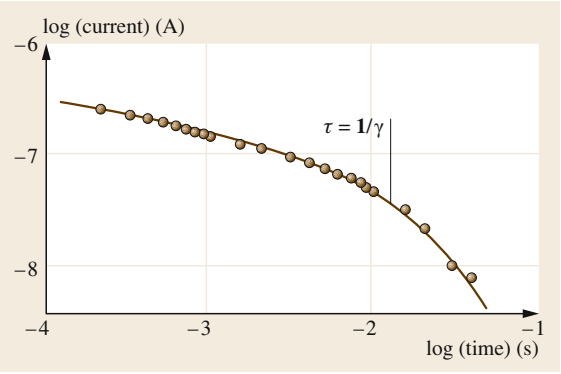


Fig. 10.5 The fit between the response function $\phi(t)/\phi(0) \propto (t/\tau)^{-n} e^{-t/\tau}\, {}_1F_1(1-m;2-n;t/\tau)$ from (10.22) (where ${}_1F_1(;;)$ is the confluent hypergeometric function [10.24]), and experimental data for amorphous gallium arsenide. (After [10.27])

The Response Function

A different approach may be taken to deriving the frequency-dependent susceptibility of linear responses. Returning to (10.14), (10.17), (10.22) it can be seen that the solutions for $\phi(t)$ in the absence of a driving field give the time evolution of the dipole density fluctuation caused by a dipole generating impulse at $t = 0$, i. e., the natural motions of the system. The resulting expression for $\phi(t)$ is called the response function [10.20], and the complex susceptibility is obtained through (10.23)

$$\chi_r(\omega) = \chi'_r(\omega) - i\chi''_r(\omega) = \mathrm{Lt}_{\varepsilon\to 0}\left[\int_0^\infty \phi(t) e^{-i\omega t-\varepsilon t} dt\right] . \qquad (10.23)$$

Here ε is an infinitesimal quantity that is taken to zero after the integral has been carried out. This equation is valid for all linear dielectric responses. The requirement that $\phi(t) \to 0$ as $t \to \infty$ for natural fluctuations ensures that $\chi'_r(\omega)$ and $\chi''_r(\omega)$ remain finite. The equations of motion, (10.14), (10.17), and (10.22) have solutions for $\phi(t)$ that can be transformed in this way to yield analytical expressions for the frequency-dependent susceptibility. An important consequence of (10.23) is that the frequency dependencies of $\chi'_r(\omega)$ and $\chi''_r(\omega)$ contain the same information since they are Laplace transforms of the same time-dependent function. This also means that a frequency-dependent conductivity $\sigma(\omega) \equiv [\omega\chi''_r(\omega)]$ contains no more information than the equivalent dielectric dispersion $\chi'_r(\omega)$. In fact the linear susceptibility components $\chi'_r(\omega)$ and $\chi''_r(\omega)$ are related to one another through the Kramers–Kronig relationships (KKR) [10.5, 6]

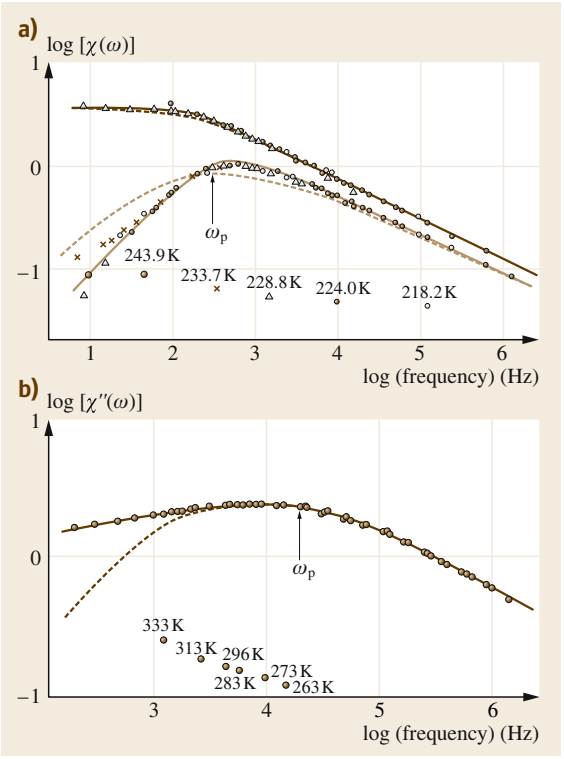


Fig. 10.6a,b Two examples of the fit between the susceptibility functions resulting from the response function $\phi(t)$ of (10.22) (*continuous line*) and (10.26) (*broken line*) and experimental master curves: **(a)** nematic form of N-(2-hydroxy-4-methoxybenzylidene)-pn-butylaniline (OHMBBA), and **(b)** polyallylbenzene. The shift of the representative point required to construct the master curve is marked on the plot. (After [10.27])

$$\chi'_r(\omega) = \frac{2}{\pi}\int_0^\infty \frac{x\chi''_r(x)}{x^2-\omega^2} dx , \qquad (10.24)$$

$$\chi''_r(\omega) = -\frac{2\omega}{\pi}\int_0^\infty \frac{\chi'_r(x)}{x^2-\omega^2} dx . \qquad (10.25)$$

The imaginary contribution from the pole at $x = \omega$ is excluded from these integrals. Although formally values of $\chi'_r(\omega)$ and $\chi''_r(\omega)$ are required over the frequency range from zero to infinity, the reciprocal relationships are adequately reproduced as long as the major part of the dispersion from any specific relaxation process is used.

Equation (10.23) means that linear dielectric responses are characterised through the time dependence of $\phi(t)$. This led *Williams* and *Watts* [10.32] to approach

the description of the frequency dependence commonly observed by proposing that the exponential behaviour of $\phi(t)$ resulting from the Debye model, (10.17), should be generalised to the form

$$\phi(t) \propto (t/\tau)^{-n} \exp\left[-(t/\tau)^{1-n}\right] . \qquad (10.26)$$

This expression is sometimes called the expanded exponential function or the Kohlrausch–Williams–Watt function, as it was later found that it was first proposed in [10.33] for mechanical responses. It is not known to possess a simple equation of motion such as (10.14), (10.17), and (10.22) but its relaxation function $\Phi(t)$, defined by $\Phi(t) = \int_t^\infty \phi(t)\mathrm{d}t$ [10.20] obeys a relaxation equation of the form of (10.17) with a time-dependent damping factor $\gamma(t) \propto t^{-n}$. The corresponding frequency-dependent susceptibility has the same power-law form as (10.20) for $\omega \gg \gamma = 1/\tau$, but exhibits a slowly varying increase of slope in a $\log[\chi_\mathrm{r}''(\omega)]$–$\log(\omega)$ plot as the frequency progressively reduces below γ that, with suitable choices for the value n, can approximate a power law for $\chi_\mathrm{r}''(\omega)$ such as is defined in (10.21). The relationship between experimental data for $\phi(t)$ and that derived in the Dissado–Hill cluster model, i. e., the solution to (10.22), is shown in Fig. 10.5. It can be seen that the data and the function for $\phi(t)$ approaches zero as t tends to infinity, with the time power law $t^{-(1+m)}$, but accurate experimental data for times several decades beyond τ is required if this behaviour is to be distinguished from that of (10.26). A better means of distinguishing the two results can often be had by recourse to their appropriate frequency-dependent susceptibilities, Fig. 10.6.

10.3 Information Contained in the Relaxation Response

As described in Sect. 10.2.2 relaxation responses contain three pieces of information. The strength of the coupling of dipole density fluctuations to the electric field characterised by χ_0, a characteristic relaxation frequency $\gamma = 1/\tau$, where τ is the characteristic relaxation time, and the relaxation dynamics characterised by the frequency dependence of $\chi_\mathrm{r}'(\omega)$ and $\chi_\mathrm{r}''(\omega)$. This latter feature is open to different interpretations, as will be discussed later.

10.3.1 The Dielectric Increment for a Linear Response χ_0

The dielectric increment is proportional to the square of the permanent dipole moments that give rise to the dipole density fluctuation. It is a feature of the dielectric response that does not usually receive the most attention, mainly because a quantitative relationship to the molecular physics of the relaxation process is often difficult to achieve. Nevertheless it has been used to determine the dipole moments of polar molecules using measurement in the gas phase or if necessary dilute solutions in a nonpolar solvent. In these cases the permanent molecular dipoles, μ, can be assumed to have the same magnitude everywhere, to be independent of one another, and to be able to adopt all orientations with equal probability in the absence of an electric field, i. e., all dipole moment orientations are at the same energy. This section starts by outlining the derivation of χ_0 for this situation even though this is not the topic area of this book and chapter. The aim is to demonstrate the procedure and bring out the assumptions involved so that the more complicated nature of dipole density fluctuations in condensed-state materials can be better appreciated.

Independent Free Dipoles

In an electric field a dipole that is at an angle θ to the field direction is at the energy $-\mu E\cos\theta$. Those molecules aligned with the electric field are therefore at the lowest energy. The thermal motions of the molecules will however tend to randomise the dipole orientations and the probability of finding a dipole with an orientation angle θ becomes $\exp(\mu E\cos\theta/k_\mathrm{B}T)$. The average value of $\mu\cos\theta$ is given by

$$M = \langle\mu\cos\theta\rangle = \frac{\int_0^\pi \mu\cos\theta\exp(\mu E\cos\theta/k_\mathrm{B}T)\sin\theta\mathrm{d}\theta}{\int_0^\pi \exp(\mu E\cos\theta/k_\mathrm{B}T)\sin\theta\mathrm{d}\theta} , \qquad (10.27)$$

and the contribution to the static polarisation is given by $N\mu < \cos\theta >$, where N is the number of permanent dipoles per unit volume. The term independent of the electric field is zero because all orientations are equally probable in the absence of the field. Equation (10.27) results in a nonlinear function in the electric field E for M, which is called the Langevin function, $L(\mu E/k_\mathrm{B}T)$, with,

$$L\left(\frac{\mu E}{k_\mathrm{B}T}\right) = \coth\left(\frac{\mu E}{k_\mathrm{B}T}\right) - \left(\frac{k_\mathrm{B}T}{\mu E}\right) . \qquad (10.28)$$

This function saturates at unity for very high values of $\mu E/k_\mathrm{B}T$, reflecting the total alignment of all the perma-

nent dipoles in the electric field. At low fields defined by $\mu E/k_B T \ll 1$, $L(\mu E/k_B T)$ is a linear function of E that gives the linear polarisation $P = N\mu < \cos\theta >$. Using $P = \varepsilon_0 \chi_r E$ gives the linear dielectric increment in χ_r as,

$$\chi_0 = \left(\frac{N}{\varepsilon_0}\right)\left(\frac{\mu^2}{3k_B T}\right) \tag{10.29}$$

Dipoles in Condensed Matter

In condensed-phase systems, particularly solids, the approximations that lead to the Langevin function and (10.29) no longer apply, and hence these expressions no longer hold. In the first place the dipoles are constrained by the local structure and in general will not be able to assume all orientations with equal probability in the absence of an electric field. In the second place we cannot assume that the dipoles are independent of one another. This dependence may arise in more than one way. For example there may be electrostatic interactions between the dipoles, such as would be responsible for the formation of ferroelectric and anti-ferroelectric states. However when the dipoles concerned are of a low concentration such as those that originate with impurities, lattice defects, interstitial ions etc. these dipole–dipole interactions may be weak. The dipoles concerned may also be arranged in such a way that even though they can adopt one or more alternative orientations their dipole–dipole interactions essentially cancel, such as might be expected in dipole glasses [10.34]. The common way to deal with this situation is to assume that a dipole representative of the average dipole in the ensemble experiences the average electric field of all the other dipoles. This is called the *mean-field* approach [10.35]. Since the mean field will be a function of the average dipole moment due to the applied electric field it is usually possible to construct an equation that can be solved to yield M and hence χ_0. Another way in which the dipoles can interact arises because permanent dipoles are part of the lattice structure of the material. Those permanent dipoles that lead to a polarisation in the presence of an electric field must have two or more local orientations available to them, i. e., they must be able to adopt a different orientation that in the presence of an electric field has a lower energy. Any such change will inevitably alter the local atomic and molecular interactions around the dipole that has moved. This effect will travel through the structure and influence other permanent dipoles through changes in atomic and molecular positions in its environment [10.36]. The strength of such interactions will vary depending on the type of dipole and the way that it is connected to the structure. For example reorientable dipoles formed by small interstitial (or substitution) ions may not interact very strongly with the surrounding lattice, whereas polar groups attached to a polymer chain will in many cases interact very strongly when they adopt a different orientation. Similarly the reorientation of polar molecules in liquids may be expected to distort their surrounding solvent cage and create a disturbance that will be transmitted to other polar molecules. The special feature of this form of interaction is that it is transmitted along specific directions depending upon the lattice structure and hence is nonisotropic.

Order–Disorder Ferroelectrics

These are materials in which the permanent dipoles possess two or a limited number of possible orientations. At high temperatures the dipoles are randomly distributed between the alternative orientations in the absence of an electric field. As the temperature is lowered the electrostatic field of the dipoles acts on any one dipole to make one of the orientations more preferable than the others. This causes the permanent dipole system to adopt a specific orientation at the Curie temperature T_c. The mean-field approach results in an expression for χ_0 that diverges at T_c, i. e.,

$$\chi_0 \propto \frac{\mu^2}{|T - T_c|} \,. \tag{10.30}$$

This expression is so common to us that it is easy to overlook the physical meaning that it contains, which is much better expressed in the renormalisation group approach [10.37]. Essentially the interactions between the dipoles cause their orientation and dynamics to become correlated to some extent. As T_c is approached from above, the dipole fluctuations in the system are correlated over increasingly long distances and involve increasingly larger groups of individual dipole moments μ. The dielectric increment increases in proportion to a power of the correlation length $\xi \propto |(T - T_c)^{-\delta}|$ and a more exact form for χ_0 is

$$\chi_0 \propto \frac{\mu^2}{|T - T_c|^{\alpha}} \,. \tag{10.31}$$

At temperatures below T_c the material will possess domains in which all the dipoles are aligned together. Dipole fluctuations in this state have the opposite orientation to that of the polarity of the domain dipole that is they are *changes in net dipole moment density*, Sect. 10.1.1. These dipole fluctuations also produce an electrostatic field that causes them to be correlated. As the temperature reduces their correlation length reduces and hence so does χ_0. These materials show that the responding dipole in condensed-phase materials will not always be that of individual molecular or ionic components of the system, they may in fact be groups of

individual dipoles that respond as a single unit. In fact the susceptibility increment may be written in terms of an effective dipole moment that is also a power of the correlation length, i. e., $\mu_{\text{eff}} \propto (\xi)^{\alpha/2\delta}$.

Kirkwood Correlation Factor

The particular feature of the ferroelectric order–disorder transition is that the correlation between dipoles in the dipole fluctuations is temperature dependent. In general this will not be the case; instead we can expect the correlation to be dependent upon the lattice structure of the material, and hence independent of temperature except for discrete step changes when the material undergoes a phase transition. One way of allowing for these correlations is to introduce a factor g, termed the *Kirkwood* factor [10.10, 38], into the expression for χ_0 such that it becomes

$$\chi_0 \propto Ng\mu^2 \,. \tag{10.32}$$

Here a value of $g = 1$ defines a system in which the dipoles are uncorrelated, i. e., they are independent of one another. Values of $g > 1$ indicate positive correlation i. e., the dipoles align in the same direction as in the ferroelectric case where g diverges as discussed. Values of $g < 1$ indicate dipoles that are anticorrelated i. e., pairs of dipoles tend to align in opposite directions. Both kinds of behaviour are known to occur.

General Features of the Temperature Dependence of the Dielectric Increment

Section 10.3.1 gives examples of behaviour in which an electric field imposes order, in terms of the polarity of the net dipole density fluctuations, upon an ensemble where the temperature acts as a disordering factor. In the case of independent free dipoles it is the applied field that attempts to align dipoles and for order–disorder ferroelectrics it is the electrostatic field of the other dipoles. It might therefore be expected that the temperature is always a factor that attempts to oppose the field and that the dielectric increment will decrease at high temperatures, but this is not always the case. For example, in situations where the dipole can adopt one of two orientations one of which is favoured by the local lattice structure, the equilibrium population will be heavily weighted towards the favoured orientation. The equilibrium value of M in such a region can be expressed as,

$$M = \mu \tanh\left(\frac{\Delta_\chi}{k_\mathrm{B}T}\right) \,, \tag{10.33}$$

where $2\Delta_\chi$ is the energy difference between the alternative orientations produced by the local structure. Here μ is the reorientable component of the dipole moment, Fig. 10.7. An applied electric field that will favour the one orientation over the other will change the energy difference between alternative orientations from $2\Delta_\chi$ to $2(\Delta_\chi - \mu E)$. Differentiation with respect to E then gives the susceptibility increment as $\chi_0 \propto (\mu^2/k_\mathrm{B}T)\cosh^{-2}(\Delta_\chi/k_\mathrm{B}T)$, with the approximately activated form

$$\chi_0 \propto \left(\frac{\mu^2}{k_\mathrm{B}T}\right)\exp\left(-\frac{2\Delta_\chi}{k_\mathrm{B}T}\right) \,, \tag{10.34}$$

holding when $\Delta_\chi/k_\mathrm{B}T \gg 1$. Although (10.33) implies that regions exist where dipoles are aligned by the structure this does not necessarily mean that the material has a net dipole in the absence of an electric field. Such regions may be local and with dipole vectors randomly arranged by the structure. This behaviour has been found in ferroelectric ceramics [10.39, 40] in both the ferroelectric and paraelectric phases outside the transition region where T approaches T_c. Expression (10.34) indicates that in the appropriate temperature range the effect of the thermal fluctuations is effectively to increase the density of dipoles that can respond to the electric field, and that this overcomes any randomising behaviour. In contrast, at high temperatures ($\Delta_\chi/k_\mathrm{B}T \ll 1$) the distribution of dipoles between the alternative orientations becomes almost random and χ_0 approaches the free-dipole result (10.29).

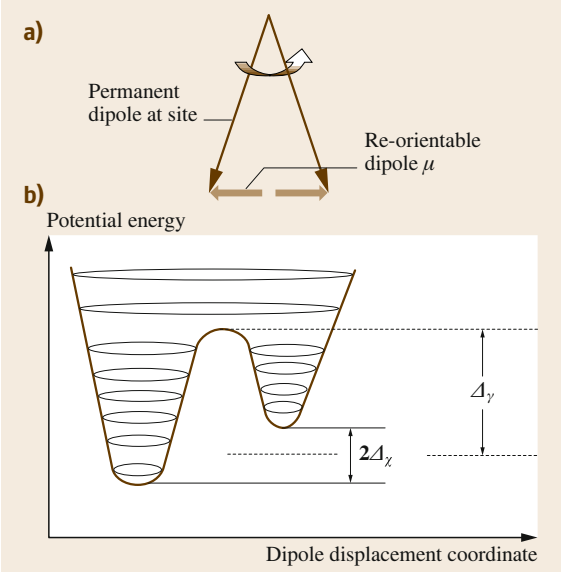


Fig. 10.7a,b Dipole reorientation between two potential wells: **(a)** shows the reorientable component of the dipole, **(b)** shows the potential-energy surface

In many experimental situations the value of the dielectric increment is essentially independent of temperature. It is difficult to see how this can occur in an ensemble where the dipole density fluctuations are produced by fluctuations in thermal energy about the average value, which couple to the electric field via changes in the heat content as in (10.27). However it may be possible to conceive of this behaviour as due to fluctuations in the configuration entropy of the molecular system, of which the dipole is a part, that take place without any change in the heat content. The effect of the field would be to change the configuration entropy S rather than the heat content H. As a result the susceptibility would be independent of temperature. This picture implies that we must think of the dipoles in this case not as local elements embedded within the material matrix moving in a fixed local potential, but as an integral part of the matrix whose dynamics is described by fluctuations in the Gibbs free energy $G = H - TS$ of the whole ensemble. In this case correlations between dipoles would be expected to occur mainly via the indirect route through their interaction with their local environment, rather than their direct electrostatic interactions.

Equation (10.31) describes the behaviour of a system undergoing an order–disorder transition *among the permanent dipole orientations*. A similar behaviour will be found for the relaxation response of a first-order ferroelectric or dipole alignment transition [10.41]. In general phase transitions will not give rise to a divergence in χ_0, which occurs because the phase transition in these cases is defined through the dipole, i. e., the dipole orientation is the order parameter. In other types of phase transition the dipoles are not the primary cause and what can be expected is an abrupt change of χ_0 as the dipoles find themselves embedded into a different lattice structure with different local potentials and orientation positions, different ensemble energies, and different correlations with one another and the material matrix.

The Information Content of the Dielectric Increment

As is clear from the above discussion it is not easy to make definite quantitative statements about the dipole system based on the dielectric increment. The basic reason for this situation is that the measurements are made on a macroscopic sample that contains an ensemble of an enormous number of dipoles, up to $\approx 10^{28}\,\mathrm{m}^{-3}$. The description of such systems can be carried through if the elements, here electric dipoles, are independent and their orientation is defined by a static local potential; as discussed above this will not be the case in general. More typically the dipoles will be correlated with the matrix in which they are embedded and/or one another. This means that the dipoles that are involved in the dipole density fluctuations are not site dipoles but groups of molecules/ions including dipoles, i. e., the responding features have a size intermediate between that of the molecule/unit cell and that of the sample. Determination of the temperature dependence will give some clues as to how to regard the dipole system through the definition of an effective dipole. The way the effective dipole changes with temperature will allow some interpretation of the kind of system that is present. Variation with other control parameters will produce more information, and systematic variation of the structure, for example replacement of side groups in polymers by longer or different side groups, or substitution of impurity ions by similar ions of different oxidation state or ionic radius, will help to identify the local dipole moment contributing the dielectric increment. However, even if the form of the site dipole is known, the quantitative evaluation of a factor such as the Kirkwood factor g (10.32) is not trivial. In condensed matter, as can be seen from Fig. 10.7, only a component of the site dipole is likely to be involved in reorientation. Unless the local structure is very well known it will not be possible to determine the actual value of the reorientable component in order to obtain a quantitative estimate of g. What can be achieved is a fingerprint of the dipole fluctuations in the material that can be used to characterise it. However at best this will be a partial picture of the dipole fluctuations in the material and information gained from the relaxation time and the relaxation dynamics should be used to enhance it further. In this way a holistic view of the dipole fluctuation can be attempted. It is important to realise that the picture obtained from these three features must be complementary. It is not acceptable to regard them as three independent features, as in fact they just yield different facets of the same process.

10.3.2 The Characteristic Relaxation Time (Frequency)

Equations (10.14), (10.17), (10.22), and (10.23) define a characteristic relaxation rate γ or relaxation time $\tau = 1/\gamma$ for the dipole density fluctuations. In the case of the Debye response, whose susceptibility functions are given by (10.18) and (10.19), γ is the frequency at which the imaginary (dielectric loss) component $\chi_\mathrm{r}''(\omega)$ exhibits a peak. It has therefore become customary to determine the dependence of the relaxation time (rate) upon the control parameters (e.g., temperature, pressure, etc.) that are varied via that of the loss peak frequency. As long as the frequency dependence of $\chi_\mathrm{r}''(\omega)$ (i. e., the loss peak shape) remains unchanged

this procedure is valid because essentially the dielectric response investigated can be treated as a single composite process, even though it has a wider frequency dependence than that predicted for the independent free dipoles. Both the theoretical equation of motion (10.22) and response function (10.26) do in fact describe the response as a single composite process with a characteristic relaxation rate (time). However the frequency of the peak in $\chi_r''(\omega)$ (loss peak frequency) predicted from (10.22) is not γ but γ multiplied by a numerical factor depending upon the power-law exponents, n and m. In many cases the loss peak is very broad ($n \to 1, m \to 0$) and it is difficult to locate the peak precisely and to be sure that the point located is at the same position with respect to the functional dependence of $\chi_r''(\omega)$ upon ω. Under these circumstances a better procedure is to construct a master curve. This is done by first plotting $\chi_r''(\omega)$ as a function of ω in log–log coordinates for each of the values of the control parameters (e.g., temperature, pressure, etc.) used in measurement, and then translating the data sets along the $\log(\omega)$ and $\log[\chi_r''(\omega)]$ axes. This will bring the data into coincidence if the susceptibility frequency dependence is unchanged. The translation required to achieve coincidence gives the dependence of the susceptibility increment ($\log(\chi'')$-axis) and characteristic relaxation frequency ($\log(\omega)$-axis) on the controlled variable. For example it gives the ratios $\gamma(T_1)/\gamma(T_2)$ (or $\chi_0(T_1)/\chi_0(T_2)$) for the temperature change T_1 to T_2 when temperature is the controlled variable. This technique also has the advantage of illustrating clearly whether or not the frequency dependence is independent of the variation in temperature (or other parameter), i. e., whether or not the different sets of data can be brought into coincidence. It can also be used to determine any relationship between χ_0 and γ. This is done by selecting a reference point (e.g., the point $\chi_0 = A, \omega = B$) and marking the position on the master curve of this point from each data set after it has been translated to achieve coincidence. A trace is formed giving the dependence of $\log[(\chi_0)^{-1}]$ as a function of $\log(\gamma^{-1})$ as shown in Fig. 10.6 where $\omega_p = \gamma$.

The relaxation rate is the dielectric response feature that shows most dependence upon the variation in the control parameters and so is the feature that is most often studied. In the following sections I will outline some of the most common types of behaviour and discuss their implications for the physics of the relaxation process.

Site Dipole Reorientation

The simplest form of relaxation rate is that described by Debye for independent molecular dipoles suspended in a viscous continuum. As described in Sect. 10.3.1 these dipoles are regarded as free to adopt any orientation in the absence of an electric field. Relaxation of a dipole density fluctuation involves the rotation of the molecular dipoles in the fluctuation to a state in which the net dipole density is zero. In such a situation the rotation of each individual dipole occurs at the same speed determined by the viscosity, η, of the medium, and the relaxation time ($\tau = 1/\gamma$) of the dipole density fluctuation is governed by that speed. For a molecular dipole whose effective length is a the relaxation time has the form [10.10],

$$\tau \propto \frac{\eta a^3}{k_B T} \,. \tag{10.35}$$

The more viscous the medium, or the bigger the molecular dipole, the slower it rotates and the slower will be the relaxation of a fluctuation, giving a net dipole moment to the ensemble.

Of course the conditions for this behaviour to be exactly applicable cannot be met except in a gaseous medium. Condensed-phase materials are not continua. Even liquids possess a local structure and molecular dipoles will either be part of that structure if they are contributed by the medium, or will be surrounded by a solvent shell if they are dissolved in the medium. In solids the molecular (or ionic) dipole is of necessity part of the structural matrix, and even though this must be irregular enough to permit rotational displacement to at least one other orientation the matrix can be expected to maintain some structural correlation to distances well away from the site of a reorientable dipole. These are the conditions that must be included in any description of the relaxation frequency (time).

The first point of departure from the picture of a free dipole in a continuum is that the dipole will possess only a limited number of orientations that it can adopt. Consequently there will be a potential barrier between these alternative equilibrium orientations. The rate of transition between alternative orientations and hence the relaxation frequency will be determined by the rate at which a dipole or, to be more specific, the atoms or ions that form the local dipole can pass over the potential barrier to switch orientation, as shown for example in Fig. 10.7. In this case the relaxation frequency will possess an activated (Arrhenius) form where the activation energy Δ_γ is the mean potential barrier height between the alternative orientations, i. e.,

$$\gamma = \frac{1}{\tau} = A \exp\left(-\frac{\Delta_\gamma}{k_B T}\right) \,. \tag{10.36}$$

The expression to be used for the pre-exponential factor A depends on the way in which the atoms/ions comprising the local dipole pass through the transition region

at the top of the barrier [10.42, 43]. In the schematic drawing of Fig. 10.7 the dipole is described as having an atom (ion) at its head that performs quantum oscillations in one of two potential wells. As long as it can be assumed that there is a thermal equilibrium between all the vibration states and that the dipole head passes into the alternative well in a single transit of the barrier region, then $A = \nu$, where ν is the frequency of the quantum vibrations at the bottom of the wells. This result continues to hold even if thermal equilibrium is established only for the states at the bottom of the well as long as the effective friction acting on the dipole head in the barrier region is weak. The type of potential surface with these properties is one that remains essentially rigid during the actual transit of the barrier region, which takes place in a time typically of $\approx 10^{-14}$ s. The activated factor in (10.36) expresses the thermal probability of finding a dipole in a quantum state at the top of the barrier. The other extreme situation occurs when the friction ς_d in the barrier region is high, for which $A \propto (1/\varsigma_d)$. This occurs when the transit of the barrier region is slow enough to allow interactions with neighbouring vibrating atoms to overdamp the motion. In this case the potential surface distorts during the transit of the barrier. Such situations can be expected when the barrier is broad and ill-defined, and correspond to local structures that are flexible, such as may be expected in viscoelastic materials. A general expression $A = \lambda_r(\nu/\nu_b)$ has been developed by *Grote* and *Hynes* [10.43] where λ_r is a function that describes the change from low to high friction and ν_b is the quantum frequency in the barrier region.

An interesting consequence of this type of potential surface is that, regardless of the magnitude of the barrier energy Δ_γ, a temperature should exist below which reorientation over the barrier would take so long that any dipole fluctuation would essentially remain unrelaxed, i. e., the dipole system becomes frozen. However when the moving atom in the permanent dipole is a hydrogen atom this is not the case; relaxation can occur by the tunnelling of the hydrogen atom through the barrier [10.44]. This has been demonstrated by experiments on deuterated oxidised polyethylene molecules at millikelvin temperatures [10.45, 46]. In this case the relaxation frequency is determined by the tunnelling probability of the deuterium/hydrogen atom through the barrier, which is dependent upon the atomic mass, the barrier width and height, but *not the temperature*, i. e., the relaxation frequency becomes temperature independent at temperatures below ≈ 100 mK.

Relaxation on a Free Energy Surface

The situation described in the previous section is one in which the dipole moves on a potential surface provided by the surrounding structural matrix. The only dynamic effect of the matrix is via elastic and inelastic interaction between the quantum vibrations of the dipole and the matrix. In many cases however, the atoms (ions) comprising the dipole will cause the displacement of the centres of motion of the surrounding atoms during its transit between alternative orientations. In this case the expression for the relaxation frequency has to refer to the group of atoms affected as a unit, and an appropriate form is that derived by *Eyring* [10.47] for chemical reactions

$$\gamma = \left(\frac{k_B T}{h}\right) \exp\left(-\frac{G^\#}{k_B T}\right) . \tag{10.37}$$

Here $G^\#$ is the change in Gibbs free energy on passing from the ground state to the transition state in the process of reorientation. The barrier is now a free energy rather than a potential barrier and reflects the need for the involvement of displacements in a number of atoms, ions or molecules in order to achieve the dipole reorientation. If we refer again to Fig. 10.7 the difference is that the normal coordinate of the quantum vibrations in the barrier region is a mixture of several different normal coordinates of the surrounding matrix as well as that of the dipole in the well. In general $G^\#$ will be composed of an activation entropy contribution $S^\#$ as well as an activation enthalpy contribution $H^\#$ with $G^\# = H^\# - TS^\#$, and both will be properties of the group of atoms/molecules involved and their structural relationship. The expression for γ therefore takes a form similar to that of (10.36)

$$\begin{aligned}\gamma &= \left[\left(\frac{k_B T}{h}\right) \exp\left(\frac{S^\#}{k}\right)\right] \exp\left(-\frac{H^\#}{k_B, T}\right) \\ &= A_{\text{eff}} \exp\left(-\frac{H^\#}{k_B T}\right) .\end{aligned} \tag{10.38}$$

The activation entropy $S^\#$ will reflect the configuration rearrangement required for the dipole to reorient. Thus for example, when reorientation requires the surrounding matrix to adopt a more irregular (disordered) arrangement so as to remove a steric hindrance to reorientation the transition region entropy will be greater than that of the dipole in the bottom of the well and $S^\#$ will be positive. Alternatively the transition region may require specific local arrangements in order that the dipole can avoid such hindrances. In this case the entropy of the transition state will be less than that of the dipole in the bottom of the well and $S^\#$ will be negative. A nonzero value of $S^\#$ will therefore lead to a relaxation frequency with an effective pre-exponential factor, A_{eff}, that is either greater or less than $k_B T/h$.

An interesting variation on this behaviour has been suggested by *Hill* and *Dissado* [10.46] who showed that several experimental relaxation frequencies could be described via what they termed an activated tunnelling expression. Here they allowed the possibility that the transfer between alternative orientations could take place by thermal activation to an energy state for which the tunnelling of hydrogen atoms through the barrier was feasible. An expression for the relaxation frequency was obtained by determining the optimum transfer rate, with all processes considered from that of an unactivated quantum tunnelling to activation over the top of the barrier. If the energy difference between the alternative orientations is zero the resulting relaxation frequency had the form

$$\gamma \propto \exp(BT)\,, \tag{10.39}$$

over a considerable temperature range, where B is a temperature-independent factor dependent upon the tunnelling distance, the barrier height and shape, and the mass of the tunnelling particle. If the *barrier* in this work is taken to be a free-energy barrier rather than a potential barrier, the analysis can be seen to be equivalent to a situation whereby the dipole relaxes by finding a route between the alternative orientations that allows for lower values of $H^{\#}$ at the expense of ordering the surrounding molecules/atoms, i. e., a negative $S^{\#}$. Routes with higher values of $H^{\#}$ require values of $S^{\#}$ that are less negative. In this case the optimum relaxation frequency will take a form such as (10.39) and only at very high temperatures will γ become purely activated with $S^{\#} \approx 0$. The picture of dielectric relaxation provided by this interpretation is consistent with the *defect diffusion* mechanism [10.48], in which defects diffusing in the structural matrix, such as a kink in a polymer chain or a dislocation in a crystal, lower or remove barriers when they reach a dipole, allowing it to take an alternative orientation of equal energy. When the defect moves on, the dipole is locked into the new position until another defect arrives.

The Glass Transition

A glass is essentially a material that has become macroscopically rigid without attaining its thermodynamically favoured crystalline state. The manner in which this occurs for a liquid or viscoelastic (rubbery) material that has been rapidly cooled has been the subject of intense investigation over a large number of years (see *Angell* [10.49]). A simple, some would say oversimple, view of the situation is to regard it as a concatenation of two effects. In the first place supercooling the liquid phase will at some temperature result in a situation where the liquid no longer possesses a heat content in excess of the crystal state at that temperature [10.50]. The liquid will be unstable with respect to a disordered solid possessing only local crystalline order. In the second place the lowering of the temperature during the supercooling will cause the lattice to contract thereby introducing and increasing barriers to local molecular/atomic motions. At some temperature the thermal fluctuations (described by the Boltzman factor in (10.36)–(10.37)) that are responsible for raising the site energy to that of the barrier become so rare that the time required for rearrangement becomes enormously long. At this temperature, termed T_g, the structure is essentially locked into a macroscopically rigid state, termed a glass. The glass transition cannot be regarded as a thermodynamic transition of state (unlike a melting/crystallisation temperature for example) and there are different ways of defining and determining T_g [10.49].

During the approach to the glass state the rate of reorientation of permanent dipoles will become slower and eventually reach zero as shown in Fig. 10.8. It is common to denote the response due to a set of dipoles that are frozen during glass formation as the α-response. A second response can be seen in the figure: termed the β-response. This is provided by dipoles that are able to reorient without requiring any substantial rearrangement of the surrounding structural matrix. It should not be expected however that this response is due just to a dipole reorientation with respect to the molecule it is attached to. In many cases the β-response involves the displacement of the molecule or part thereof as a whole [10.51]. In polymers this is a local displacement of small polymer chain sections with or without side groups (see polyethylene in [10.52] for

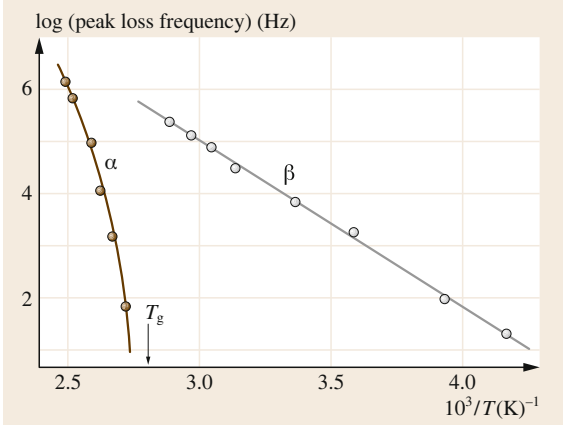


Fig. 10.8 The Arrhenius plot of the glass-forming polyvinylchloride system showing the slowing down of the α-response ($T > T_g$) as T_g is approached from above and the activated β-response. (After [10.8])

example), and either the free-energy expression (10.38) or the potential-energy expression (10.36) will apply, depending on whether the surrounding chains remain rigid during the relaxation or rearrange locally. These dipoles are active in the glass state and can be expected to have a relaxation rate of the form of (10.36), i. e., reorientation over a potential barrier. In the case of the α-response it is clear that the relaxations must involve displacements in a number of molecules/atoms other than just those comprising the permanent dipole, and hence it is instructive to discuss the behaviour in terms of the rate expression (10.37). What can be seen is that, as the temperature at which the system becomes rigid is approached from above the gradient in the Arrhenius plot gets steeper, and $H^{\#}$ therefore becomes larger. The non-thermally-activated pre-exponential factor, A_{eff}, in (10.36) is greater than $k_{\mathrm{B}}T/h$ and hence $S^{\#} > 0$. As the temperature approaches T_{g} there is an increase in A_{eff}, by many decades in frequency, which must be due to an increase in $S^{\#}$. These changes in $H^{\#}$ and $S^{\#}$ indicate that, as T_{g} is approached, dipole-orientation relaxation not only requires an increased amount of energy in order to enter the transition (barrier) region but also a larger amount of configuration disordering in the surrounding structure that makes up the molecular/atom group involved in relaxation. Although it is possible that such a situation may come about because reducing the temperature produces a local increase in density that increases steric hindrances for the same group of atoms and molecules, it is more likely that the number of molecules that are displaced in order to allow the dipole to pass through the transition region has increased. These considerations are consistent with a structure that is becoming either tangled or interlocked as the temperature decreases. Detailed expressions based on these concepts but involving macroscopic parameters have been attempted [10.30, 49, 53–56].

The glass formation discussed above has a structural basis and dipole–dipole interactions will play at most a minimal role. In some situations however, the glass is a disordered array of dipole orientations [10.34]. This sort of state is most likely to occur at very low temperatures in materials that possess dipoles occupying the sites of a regular lattice. At high temperatures the dipole orientations will be disordered but, as the temperature is reduced to low values, each individual dipole would be expected to adopt their lowest-energy orientation, resulting in a state of ordered dipole orientation. A *dipole glass* will result instead when the dipole–dipole interactions produce forces that generate barriers to the local reorientation and frustrate the ordering process at temperatures low enough that the barriers generated cannot be overcome in any conceivable time.

Ferroelectric Transition

The dielectric response of ferroelectrics at temperatures in the vicinity of their Curie (critical) temperature also exhibit relaxation frequencies that approach zero, while their dielectric increment approaches infinity (Fig. 10.9a,b) as discussed in Sect. 10.3.1. In this case both the dielectric increment and the relaxation frequency are functions of a hidden variable that characterises the system, the correlation length ξ of the dipole fluctuations. Just as the dielectric increment increases with a power of the correlation length, the relaxation frequency will decrease. Put simply the more

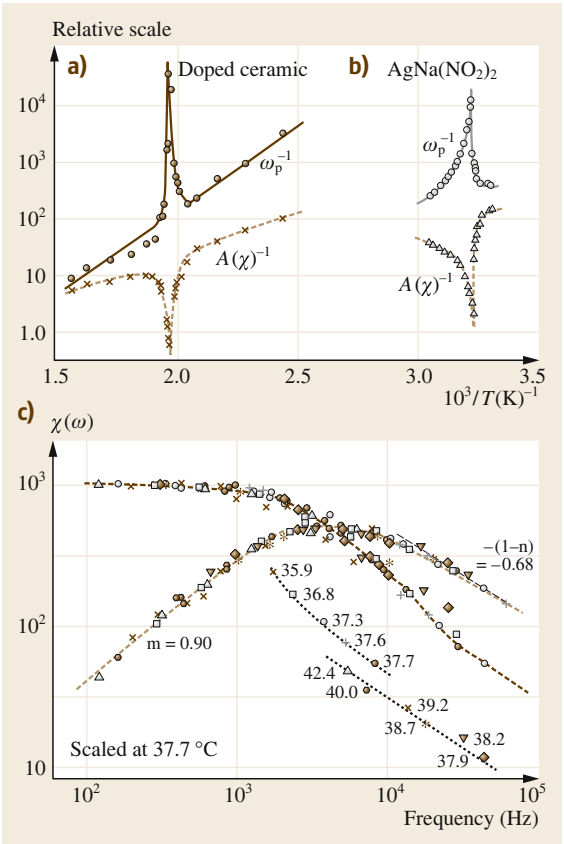


Fig. 10.9a–c An Arrhenius plot of the inverse of the susceptibility increment $A(\chi)(\equiv \chi_o)$ and loss peak frequency $\omega_{\mathrm{p}}(\equiv \gamma)$ is shown for **(a)** a ferroelectric ceramic and **(b)** the ferroelectric $AgNa(NO_2)_2$. The frequency-dependent susceptibility of $AgNa(NO_2)_2$ is shown in **(c)** (on log–log scales) together with the locus of the representative point as $T_{\mathrm{c}} = 37.8\,^{\circ}\mathrm{C}$ is approached from above and below. Note the power-law relationship between the dielectric increment and the relaxation frequency, and the parallelism between this locus and the frequency dependence for $\omega > \gamma$, demonstrating an identical power-law behaviour. (After [10.39])

dipoles are correlated in the fluctuation the longer the time that is required for its relaxation. Scaling theory [10.37] describes the system by a hierarchy of self-similar correlations. The strongest correlations are between the dipole and its nearest neighbours. This gives a local geometrical arrangement of correlations. The next-strongest correlations are between the same geometrical arrangement of groups of nearest neighbours, and the next strongest is between the same geometry of groups of groups. Eventually the whole system up to the correlation length is constructed in this way. Because the geometrical arrangement is preserved at each stage the properties for each stage have to be proportional to a power of the size. This gives

$$\gamma \propto |T - T_c|^{\beta} \propto \xi^{-\frac{\beta}{\delta}} , \tag{10.40}$$

and using (10.31) the relationship

$$\chi_0 \propto \gamma^{-\frac{\alpha}{\beta}} , \tag{10.41}$$

follows. But the theory can go further and predict the frequency dependence of $\chi_r''(\omega)$ and $\chi_r'(\omega)$ for $\omega > \gamma$. This follows because we can think of the response of the system to a field of frequency ω as being due to the correlation scale that can relax at the frequency ω, i. e., $\xi_\omega \propto \omega^{-\delta/\beta}$. The dielectric increment appropriate to this length scale can be obtained from (10.31) as $(\xi_\omega)^{\alpha/\delta}$ and hence,

$$\chi_r''(\omega) \propto \chi_r'(\omega) \propto (\xi_\omega)^{\frac{\alpha}{\delta}} \propto \omega^{-\frac{\alpha}{\beta}} = \omega^{n-1} , \tag{10.42}$$

with $n = 1 - \alpha/\beta$. These relationships describe what is called dynamic scaling. A detailed description of the derivation of the frequency dependence in (10.42) is given in [10.37] and the relationship of the power-law frequency dependence to self-similar relaxations is shown explicitly using scaled electrical circuits in [10.57]. Because of the small range of temperatures around T_c over which the power-law relationships (10.40)–(10.41) are expected to hold it is difficult to determine the exponents α and β with any accuracy, however the dynamic scaling law can be easily demonstrated by using the master-curve technique and determining the locus of the representative point (see the beginning of Sect. 10.3.2). An example is shown in Fig. 10.9c where it can be seen that the representative-point locus of χ_0 as a function of γ is a power law with the same exponent as that of the frequency dependence of $\chi_r''(\omega)$ and $\chi_r'(\omega)$ for $\omega > \gamma$.

The Information Content of the Relaxation Frequency (Time)

The typical starting point in investigating the relaxation frequency (time) is to make an Arrhenius plot, i. e., $\log(\gamma)$ is plotted as a function of $1/T$. A straight line is taken to indicate an activated process of the form of (10.36), with the gradient yielding the activation energy Δ_γ. However in most cases data is only available from a restricted temperature range and so the variation of $k_B T/h$ with temperature will be small. In this case the free energy expression (10.38) would give an equally good straight line. Since there are many situations in which the pre-exponential factor in (10.36) will be less than a quantum vibration frequency $10^{12}\,\text{Hz} \leq \nu \leq 3\times10^{13}\,\text{Hz}$, it is difficult to discriminate between the potential-barrier situation and the free-energy barrier in which $S^{\#} < 0$. Only when $A \gg 3\times10^{13}\,\text{Hz}$ will identification with relaxation via a free-energy barrier be certain. As can be seen from the foregoing it is difficult to obtain any definite information from the pre-exponential factor in an activated relaxation, except for the latter case where an evaluation of $S^{\#}$ is possible. Where $S^{\#}$ can be determined it gives us a qualitative picture of a relaxing dipole centre which has to be structurally distorted in order for the dipole to adopt a different orientation. The corresponding value of $H^{\#}$ gives the amount of energy that has to be supplied to the group in order for the reorientation to occur even allowing for the distortion. What cannot be determined from this information is the size of the group that is involved, though the behaviour and magnitude of χ_0 may give an idea as to the dipole magnitude and density. To give an idea of the difficulty of a molecular interpretation let us take the case of cyclo-hexanol [10.58]. This molecule forms a plastic crystal, i. e., the molecular centres are located on a crystal lattice but their orientations are disordered. The dipole moment is associated with the only strong polar feature, the alcohol (–OH) group, which can take one of two orientations with respect to the molecule. It would therefore be expected that the measured activation energy of $\approx 0.5\,\text{eV}$ would be the potential barrier to the transfer of the alcohol group between the two positions, with its contributions from the neighbouring molecular cage as well as the molecular energy change. However, the molecular structure of the cyclohexanol can itself exist in two conformations, the chair and the boat, and can rearrange its orientation in a lattice by passing through the alternative conformation as an intermediate. The free-energy barrier to this interconversion is also $\approx 0.5\,\text{eV}$. So we cannot decide from the relaxation frequency whether the relaxation involves just -OH group transfer or transfer via a boat-to-chair transition or a mixture of both. In this case the pre-exponential frequency $A_{\text{eff}} \approx 6\times10^{16}\,\text{Hz}$, so relaxation has to have a positive $S^{\#}$ and involve a number of atoms rather than a dipole reorienting on a rigid potential surface.

In the case of near-crystalline materials where the dipole is associated with defect centres we would expect the potential-barrier approach to be the best, but even here the fact that local reorientation is possible implies some sort of interaction between the surroundings and the moving dipole head. Calculations based on a rigid cage should [10.59] however be possible, and comparison with experiment can be expected to determine how well this represents the situation and to what extent the transit of the barrier is affected by the barrier friction. Information provided by the dielectric increment should be of help here. The calculation ought to be able to yield an estimate of the reorientable component of the dipole, and if as seems likely the alternative orientations are at different energies, the temperature dependence of χ_0 should follow (10.34).

Although it is conceptually simple to think of dipoles relaxing upon a potential surface that remains unchanged during the relaxation, this is likely to be only an approximation to reality. The fact that alternative orientations exist indicates that in most cases the surrounding structure must be modified to some extent to accommodate the change; at the very least we can expect the dipole to polarise its surroundings differently according to its orientation. The expressions in Sect. 10.3.2 for the relaxation frequency of dipoles relaxing on a free-energy surface and dipoles in ferroelectrics reflect this fact in different ways. The ferroelectric behaviour described in Sect. 10.3.2 shows that when the *dipoles* become extensively correlated the relaxation frequency reduces as an inverse power of the correlation length and the dielectric increment increases as a different power of the same correlation length. The self-similar scaling relates this behaviour to the frequency dependence of the susceptibility. The behaviour of the α-response of a glass-forming system involves dipole-structure interaction in a different way. The relaxation frequency approaches zero as T approaches T_g from above. Whatever the details of the glass-forming process this behaviour has to indicate a progressively increasing difficulty for the dipoles to reorient, which here is associated with structural ordering, densification, and atomic packing rather than long-range correlations as in ferroelectrics. This response is also one for which the dielectric increment is often insensitive to temperature. If we put the two dielectric response features together we come to a picture in which the electric field effectively modifies the configuration entropy of the system in generating a net dipole density fluctuation. The net dipole density produced is essentially the same at different temperatures, so the change in configuration density generated by the electric field does not vary, but the relaxation time increases as the activation enthalpy $H^{\#}$ and entropy $S^{\#}$ increases. Put together with the fact that in structural glass formation small local regions are attempting to adopt a crystalline structure, this data indicates that there are *local* values of ground-state configuration entropy that reduce as T_g is approached, with a transition state involving a disordering of the local regions to free the dipole enough to allow it to adopt an alternative orientation in an equally ordered but different configuration. The dipole density fluctuations that couple to the electric field seem to involve reorganisations of the structure that can occur without a change in the value of the heat content H, i. e., they are essentially configuration entropy fluctuations rather than thermal fluctuations. In contrast to the ferroelectric situation the slowing down of the relaxation is not caused by longer-range correlations but by the increasingly larger numbers of molecular adjustments required to achieve a dipole reorientation.

The message of this section is that in most cases a detailed molecular description of the dipole motions is generally not accessible just from an analysis of the dielectric response. The reason is that, in general, dipole reorientations involve adjustments in the surrounding molecules/atoms that are not easy to define in molecular terms. However by putting together the behaviour of the dielectric increment and relaxation frequency it should be possible to obtain some general idea as to the extent of the connection of the reorientation to the molecular environment and the way in which it takes place.

10.3.3 The Relaxation Peak Shape

The explanation of the frequency dependence of the susceptibility is currently the most contentious of the features of the dielectric response. Many workers are content with just defining the shape by one or other of the empirical functions mentioned in Sect. 10.2.2, or through the power-law exponents of (10.20)–(10.21). This gives a fingerprint of the dipole dynamics but no more. In particular it does not provide a description of the equation of motion of the dipole density fluctuation. Others determine what is termed a *distribution of relaxation times* for the loss peak in $\chi_r''(\omega)$. Essentially this approach is predicated on the assumption that the broadening of the loss peak compared to that of the Debye response (10.19) is the result of dipoles of the same type and dipole moment that each relax according to the Debye equation of motion (10.17) but possess different relaxation times with a distribution denoted by $g(\tau)$, which is defined via (10.43)

$$\chi_r''(\omega) = \int_0^\infty \frac{g(\tau)\chi_0\omega\tau}{1+\omega^2\tau^2}\mathrm{d}\tau\,. \tag{10.43}$$

This construction is still no more than a fingerprint unless a physical reason for the distribution $g(\tau)$ can be found. Usually this is ascribed to a distribution of local activation energies associated with dipoles that each move on their own potential surface independent of one another. The system usually quoted as an example is that of the β-response in the rigid glassy phase, which typically has a very broad loss peak. In this case it is assumed that each dipole that can reorient to contribute to the β-response is essentially trapped in a local potential surface that is held rigid in the glass state. Of course the potential surface is not truly rigid, molecular/atomic vibrations must take place, but it is assumed that their effect on the potential averages out during the relaxation and their only effect is to raise the energy state of the reorienting dipole head to the state at the top of the barrier. One problem associated with this explanation of the origin of $g(\tau)$ is that, if the *function* $g(\tau)$ is independent of temperature the values of exponents n and m (10.20), (10.21) will be temperature dependent. This does not seem to be the case in general, with these exponents usually either constant or changing at the most slowly or discretely at a transition of state [10.8, 60], but there is no real agreement on this point. Of course a temperature-dependent distribution function $g(\tau)$ may be assumed, but then the question arises as to why it *is* temperature dependent in a system that is presumed to be macroscopically rigid. Another facet of the problem associated with non-Debye loss peaks that does not seem to have received any consideration is the possibility that the magnitude of the reorientable dipole moment associated with each site of a given activation energy is also distributed. It is clear that this is highly likely even if the dipole moment that changes direction is the same everywhere, as illustrated in Fig. 10.7. Also, as described in Sects. 10.2 and 10.3, the local dipole may be correlated with other dipoles or its surroundings, and in this case we can expect the Debye rate equation not to hold. The ferroelectric result (Sect. 10.3.2) already shows that this is the case when the dipoles motions are correlated giving the system a scale relationship in its dynamics, and even correlation between the dipole and its surroundings, for which there is considerable evidence (Sects. 10.2 and 10.3) can be expected to alter the form of the equation of motion from (10.17), by for example anharmonic coupling between the various modes. Even if we assume that all the criteria for the application of (10.43) are met, the $g(\tau)$ that are required to fit the experimental form of response defined by (10.20), (10.21) (and its corresponding theoretical response function, (10.22)) have unique features that require a physical justification, i. e., there is a cusp or sharp peak at the value of τ corresponding to the characteristic frequency ($\tau_c = 1/\gamma$), and power-law wings to either side whose power exponents are $1-n$ [$g(\tau) \propto \tau^{1-n}$] for $\tau < \tau_c$, and $-m$ [$g(\tau) \propto \tau^{-m}$] for $\tau > \tau_c$. In the Debye case the distribution becomes a delta function at the characteristic relaxation time. Essentially the *distribution of relaxation times* approach is convenient but it is not as easy to justify as would seem at first sight.

The *Williams* and *Watts* response function [10.32] started life as a heuristic suggestion but has received some later theoretical support [10.61–66]. The dynamic scaling behaviour appropriate to ferroelectrics gives a clue as to the way in which a frequency-dependent susceptibility of the form of (10.20) can come about, which results from both the equation of motion (10.22) and the response function (10.26). Essentially there has to be a self-similarity (or scaling) between the relaxation frequency of subcomponents of the system and their contribution to the dielectric increment (as illustrated in the circuit model of [10.57]). The theory proposed by *Palmer* et al. [10.64] refers this scaling to the removal of a hierarchy of constraints, thus for example we may imagine that close neighbours move quickest and remove the constraints imposed on larger groups of molecules and so on. This picture would be appropriate to a system such as a glass-forming material. The assumption however is that the motions are overdamped at all levels of the hierarchy, and hence no bridge is provided to the oscillatory motions known to occur at times close to quantum vibrations. A rather different stochastic approach has been taken by *Weron* and *Jurlewicz* [10.65, 66] who assumed that the system relaxation followed a path in which the fastest dipoles out of a distribution relaxed first and then the fastest out of the residual distribution and so on. The key feature is that the relaxing dipole is the extreme fastest from the distribution existing at the time. It was argued that the extreme-value statistical distribution function then led automatically to the response function of (10.26). The choice of appropriate extreme-value distribution was made on the grounds that the relaxation time was a positive definite variable. However this is not a sufficient criterion [10.67]. In order for (10.26) to apply the continuous distribution density of relaxation times (i. e., the distribution the system would have if it were of infinite size) has to be stable to scale changes [10.68] and thus has to approach the extreme of long times as an inverse power law, otherwise a different extreme value statistic or *none at all* applies. The required form of distribution from which an extreme selection has to be made is one that applies to the size distribution of scaling systems [10.69] such as percolation clusters [10.70] for example. So even with this stochastic approach we are led back to a system for which the dynamics scale in some way.

The *Dissado–Hill* function [10.26, 27] for which the response function obeys (10.22) also has scaling features as its basis, however unlike the other approaches it starts with the vibration dynamics of the system. It is assumed that a dipole that can reorient couples local vibration modes to itself. These are no longer extended normal modes but modes centred on the dipole that reduce in frequency according to the molecular mass involved. Their frequencies lie in the region between optical modes and the relaxation frequency and have a scaling relationship one to another. In the theory of *Nigmatullin* and *Le Mehaute* [10.71, 72], the modes are impulses that are involved in the dipole relaxation process whose time of action is scaled, i. e., these are fractal time processes such that the longer the time of action the more correlated they are to the dipole motion. In general these modes are local versions of coupled optical and acoustic modes and it is not surprising that they extend to such low frequencies as those involved in relaxation, as acoustic modes essentially extend to zero frequency. Their coupling with the dipole leads to the high-frequency power law of (10.20), where n expresses the extent to which the dipole reorientation couples to the surroundings, i. e., $n = 0$ corresponds to no coupling and the dipole moves independently of its surroundings, and $n = 1$ corresponds to full coupling in which the dipole motion is just part of the local mode. In a sense the short-time development of the response function of (10.22) is that of the changes in the configuration entropy as various amounts of different local modes are progressively coupled into the dipole motion [10.26]. In this case there is no necessity for n to be temperature dependent. However the Dissado-Hill approach to the derivation of the fractional power laws of (10.20)–(10.21) can be regarded as equivalent to the development of the self-energy of a particle (the reorienting dipole) as it binds itself to the self-similar hierarchy of local modes of the medium in which it moves [10.26, 27]. Recently this parallelism between the fractional power laws and energy scaling in the coupling constant defining the self-energy of a reorienting dipole has been exploited [10.73] to derive the temperature dependence of n for aluminium oxide. At the characteristic relaxation frequency, the *characteristic* dipole group relaxes and transfers energy to the heat bath. The low-frequency behaviour of (10.21) is the result of a distribution in the ensemble of locally coupled dipole motions. This occurs because the motions of local dipole centres may be weakly coupled to one another. As a result the relaxation of the centres proceeds in a scaled or self-similar manner. First the dipole in a local centre relaxes with respect to its own environment, this leaves each dipole centre unrelaxed with respect to one another. Next groups of dipole centres, with some arrangement depending upon the specific structure involved, relax as a group. Then groups of groups relax and so on. Each level of inter-group complexity essentially has a time scale associated with its relaxation that cannot be reached until the preceding level has been completed. This is rather similar to the constraint relaxation concept of *Palmer* et al. [10.64]. The power-law exponent m expresses the way that this hierarchy of relaxing groups is scaled, by defining the power-law tail of the distribution of inter-group relaxation times in the ensemble [10.74]. A value of $m = 1$ corresponds to a sequence of inter-group relaxations with a relaxation time that is proportional to the number of groups involved in the sequence [10.74]. This implies that the sequential events are uncorrelated, i. e., the long-time relaxation is a white-noise (random) process [10.26]. When combined together with $n = 0$ the Debye response is recovered. On the other hand a value of m approaching zero corresponds to relaxation times that are a very high power of the number of groups involved [10.74] and indicates a very strong connection between groups at all levels of the hierarchy. This will spread the response to very low frequencies, as observed. Essentially m is a measure of the extent to which energy is transferred to the heat bath (dissipated) at each level of the hierarchy compared to being stored in the inter-group interactions of the next level. This scaling has also been used to derive the temperature dependence of m in the case of aluminium oxide [10.73] by means of a similar self-energy approach such as was used for n. Scaling has also been shown to yield the Cole–Cole response (Sect. 10.2.2) through a fractional Fokker–Planck equation resulting from anomalous sub-diffusion [10.75] (i. e., the compact exploration of a matrix, fractal or non-fractal) and has been demonstrated recently [10.76] to apply to the relaxation of water in porous silica glasses.

The common feature of all theories other than the *distribution of relaxation times* approach is that self-similar scaling is at the basis of the fractional power law responses. The Dissado-Hill theory differs from the others in that there are two different ways in which scaling can be involved. This theory is not generally accepted. The controversial parts of the theory are firstly the coupling of the dipole motions with vibration modes, which modifies the oscillator behaviour towards an overdamped form, and secondly the hierarchy of relaxations whereby energy is transferred to the heat bath. However it should be noted that the susceptibility function that results has a general form that agrees well with experiment. In addition the concepts are reasonable given the complexity that is likely to occur in the internal motions once an ideal crystalline regularity is ruled out by the possibility of dipole reorientation. Thus

for example this concept would apply to the dipoles involved in the β-response of the glass state as well as correlated motions of dipoles over long distances, since even in a macroscopically rigid material local vibrations take place. In fact the limited regions of local order in a glass phase can be expected to favour such local modes and increase the coupling of the dipole motions with them, as observed.

The Information Content of the Loss Peak Shape

It is clear from the foregoing discussion that for all theoretical models of the loss peak shape the characteristic or loss peak frequency is but the culmination of a process in which subsections of the dipole and environment (with or without dipoles) are mixed into the motion of the dipole centre. In these models the dipole is not an independent entity, but rather an entity that is connected to its surroundings to some extent over a region that may be small or large. This implies that the dipole is not a particle that relaxes on a rigid potential surface independently of its environment. Only the *distribution of relaxation times* approach preserves the latter concept. If the theoretical models are correct they reflect the fact that we are looking at entities that are not truly of molecular scale but are of a mesoscopic nature. The correlations noted to occur in χ_0 and the need to use free energy rather than potential surfaces in describing the relaxation frequency support this view. The local entities involved are however not rigid features like permanent dipoles, and for this reason we should expect there to be weak connections between them that can be expected to relate to the way in which the relaxation of the whole system takes place. That is, not all entities relax at the characteristic time. As one entity relaxes its neighbours have to come into equilibrium with its new orientation and the system approaches equilibrium more slowly (i. e., as the time power law $t^{-(1+m)}$) than the exponential behaviour of the Debye response function or the expanded exponential function. The information contained in the loss peak shape indicates the way in which a dipole density fluctuation evolves from its state when initially created to an ensemble of mesoscopic dipole centres. The broadening of the peak from that of a Debye peak indicates the involvement of faster and slower processes as part of the overall mechanism, whatever their detailed origin, and in particular processes that have a scale relationship to one another. This must apply even to a distribution of relaxation times because of the unique form required for that distribution. Equation (10.22) implies an equivalent description that refers the overall relaxation process to a conversion of the vibration oscillation at short times to an overdamped motion as the dipole density fluctuation dissipates its energy irreversibly. In this sense evaluation of the shape parameters n and m give a means of describing this conversion process. At the very least they give a sense of the scaling involved in spreading the relaxation process around the characteristic relaxation frequency or equivalently the characteristic relaxation time.

10.4 Charge Transport

All dielectrics possess a constant (DC) conductivity (σ_{DC}), although usually it is very weak. Since $\chi''(\omega) = \sigma(\omega)/\omega$ as demonstrated in Sect. 10.1 (10.10), it would be expected that a dielectric response at low frequencies ($f \lesssim 10^{-2}$ Hz) would take a form in which $\chi''(\omega) = \sigma_{DC}/\omega$ and χ' is independent of frequency. In many cases however the conduction process is blocked at the electrodes or internal interfaces. In this case the DC conduction charges the interface, which behaves as a capacitor, and the whole system behaves as a single dipole. As long as the interface does not possess relaxation dynamics of its own, the response that would be observed is that given by the Debye response of (10.18)–(10.19), with $\tau = 1/\gamma = RC_i$, where R is the resistance of the body of the material and C_i is the capacitance of the interface. The measured dielectric increment $\chi_0 = dC_i/A$, where A is the electrode area and d is the sample thickness, and can be very large depending upon the ratio of the sample thickness to that of the interface. The situation where the interface has a frequency-dependent capacitance has been thoroughly discussed by *Jonscher* [10.6] who has shown that $\chi'_r(\omega)$ is modified from $\chi'_r(\omega) \propto \omega^{-2}$ $(\omega > \gamma)$ to $\chi'_r(\omega) \propto \omega^{-q}$, while $\chi''_r(\omega) \propto (1/\omega)$ $(\omega > \gamma)$ as in (10.19). The value of q lies in the range $1 < q < 2$ with its value depending upon the frequency dependence of the interface capacitance.

The bulk DC conductance arises from charged particles whose movements are not bound to a charge of the opposite polarity as in dipoles but are free to move independently of their countercharge, resulting in a net charge displacement in the same way that a liquid flows. However the transport of charged particles within the body of the sample can give rise to a very different form of response when their movement lies along defined paths such that the longer the displacement of the charge the lower the number of paths or equivalently the more difficult the transport becomes. This behaviour was called low-frequency dispersion by *Jonscher* [10.6,

21] and quasi-DC conduction (q-DC) in the theoretical model of *Dissado* and *Hill* [10.77] who wished to distinguish it from low-frequency dipole responses. At frequencies below some characteristic value ω_c this form of response takes the form,

$$\sigma(\omega) \propto \omega^{1-p}$$
$$[\text{i.e., } \chi_r''(\omega) \propto \chi_r'(\omega) \propto \omega^{-p}], \quad \omega < \omega_c, \quad (10.44)$$

and at frequencies above ω_c,

$$\sigma(\omega) \propto \omega^{n}$$
$$[\text{i.e., } \chi_r''(\omega) \propto \chi_r'(\omega) \propto \omega^{n-1}], \quad \omega > \omega_c. \quad (10.45)$$

The power-law exponents p in (10.44) and n in (10.45) have positive fractional values near to unity. It is obviously difficult to identify a value of p close to unity from measurements of $\chi_r''(\omega)$ (or equivalently $\sigma(\omega)$) and in many cases it is assumed that the measured behaviour shows a static (DC) conductivity. It is then common to subtract its supposed value from the measured data for $\sigma(\omega)$ to obtain an expression for the dipole relaxation response supposedly responsible for the behaviour at $\omega > \omega_c$. The values obtained in this way for $\chi_r''(\omega)$ at frequencies $\omega < \omega_c$ will not be zero as $\sigma(\omega)$ is not in fact constant, instead they will reduce as the frequency is reduced. This procedure yields a spurious loss peak in $\chi_r''(\omega)$ if the response is actually due to a q-DC mechanism since the high-frequency behaviour is an essential component of the whole and can never be resolved as a separate peak in $\chi_r''(\omega)$. The way to be certain that the response is really of the q-DC form is to measure the frequency response for $\chi_r'(\omega)$ and show that it takes the same frequency dependence. A convenient check is to determine the ratio of $\chi_r''(\omega)$ to $\chi_r'(\omega)$ (i.e., $\tan\delta$) which will have a constant value [10.6, 21] given by

$$\frac{\chi_r''(\omega)}{\chi_r'(\omega)} = \tan\delta = \cot\left[\frac{(1-p)\,\pi}{2}\right]. \quad (10.46)$$

Here $\tan\delta$ is called the loss tangent and δ is the phase angle between the real and imaginary components of the susceptibility. This relationship holds for pairs of values of $\chi_r'(\omega)$ and $\chi_r''(\omega)$ at the same frequency even if the measurements are noisy and so make it difficult to determine accurately the value of p from the frequency dependence. Another situation where it is difficult to detect the q-DC behaviour occurs in heterogeneous materials when one component has a low DC conductivity. This will add to the AC component, (10.44) and obscure the q-DC behaviour. In this case the DC conductivity can be eliminated from the data, if it is available over a large enough frequency range, by applying the Kramers–Kronig transform of (10.25) to obtain the function $\chi_r''(\omega)$ without the DC component (σ_{DC}/ω). The validity of the procedure can be checked by applying the inverse transform (10.24) to the measured data for $\chi_r''(\omega)$. This should yield the measured $\chi_r'(\omega)$ since the DC conductivity does not contribute to the real component of the susceptibility.

The q-DC behaviour, (Fig. 10.10a), is most often found in materials that are heterogeneous on a mesoscopic scale such as ceramics [10.79], rocks [10.80], porous structures [10.81], and biological systems [10.82]. In these materials charged particles are transported via structured paths over some finite range. The transported charge and its countercharge give rise to an effective dipole with a large dipole moment. However the q-DC behaviour rarely appears as an isolated response. Because of the heterogeneous nature of the materials it is usually found to be electrically in series with other dielectric response elements such as interface capacitances, and electrically in parallel with a ca-

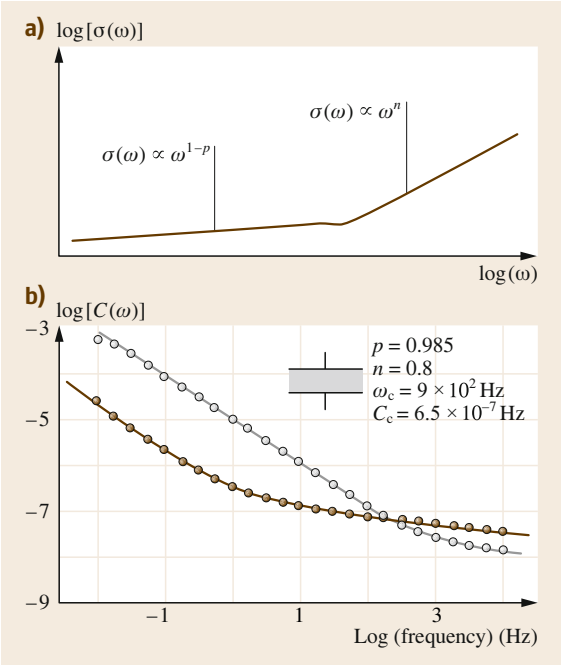


Fig. 10.10 **(a)** A schematic representation of the frequency dependence of $\sigma(\omega)$ for the q-DC process, **(b)** the equivalent q-DC susceptibility function $\chi(\omega)/\chi_c \propto (1 + i\omega\tau)^{n-1}\,{}_2F_1[1-n; 1+p; 2-n; (1+i\omega\tau)^{-1}]$ derived from the response function of (10.47) fitted to data from the leaf of *Nicotania* (Solanaceae), the *inset* gives the fitting parameters, with $\omega_c \equiv 1/\tau$, and $C_c \equiv \chi_c$. This data was chosen as it exhibits an isolated q-DC process. As before ${}_2F_1(,;;)$ is the Gaussian hypergeometric function. (After [10.24, 78])

pacitive circuit element. In this case the impedances of the series components ($Z(\omega) = [i\omega C(\omega)]^{-1}$) are added together, inverted to obtain the equivalent capacitance, and added to any capacitances in parallel. As a result identifying the response that can be assigned to any particular process requires the use of an appropriate equivalent circuit. In contrast, materials where the responses arise from different permanent dipoles embedded in the same structure but with different relaxation timescales give capacitance contributions in parallel that are simply added to yield the total capacitance, and hence are easy to separate. The reader is referred to [10.6] for examples of how to identify and separate the various processes in each situation.

The origin of the q-DC behaviour lies in a hidden scale relationship, with the dipole contribution to the susceptibility increment and its relaxation time both being a power of the length over which the transport takes place. The circuit models of *Dissado* and *Hill* et al. [10.57, 83] show how this behaviour can be produced when the system is represented by a geometrically self-similar arrangement of transport paths and blocking capacitive regions. Such geometrical regularity is not essential however [10.84]; a random arrangement of conductors (transport paths) in a dielectric (i. e., residual set of capacitances) will also result in the q-DC behaviour emerging as the total number of components increases. It is clear that this construction yields percolation clusters below the size necessary to span the material, and these sub-percolation clusters will of necessity possess scaling relationships dependent on their size and the number of paths within them, and between clusters of different sizes. Such percolation systems also show q-DC behaviour when below their critical limit [10.85]. Self-similarity in transport processes giving what has been termed a *fractal* current has also been used recently [10.86] to derive a generalised version of the low frequency power law of (10.44) and verify it for some electrode polarization processes, but it should be noted that many examples of q-DC behaviour have been demonstrated to be bulk processes [10.87, 88]. The theory proposed in [10.77] described the q-DC mechanism in similar terms to that used in percolation theory. At high frequencies they assumed that the charged particles had a short range of motion restricted to local clusters in which they were bound to their countercharges. The power-law exponent n reflected the extent of binding of the motions of the charges in the cluster and the local motions in the same way as for the dipole motions. Its value close to one indicates that the binding is strong and that the polarisation of the local clusters increases only to a small extent as the frequency is lowered, i. e., the net displacement of positive and negative charge is local and small. At the characteristic frequency the charged particles become free of their locality and are able to move over much longer distances. This frequency can be expected to correlate with the Debye screening length or an equivalent feature in the heterogeneous medium. The transport range allowed by the AC frequency increases as the frequency reduces, but the number of effective routes decreases as the distance becomes longer and hence the current becomes smaller. This leads to the weak dependence of the conductivity upon the AC frequency defined in (10.44). In the zero-frequency limit $\sigma(\omega) = 0$, which means that there are no paths across a system of infinite length, i. e., the system is formally below the percolation limit. Of course in reality the sample is of finite size and some routes may cross the sample, leading to a termination of the q-DC behaviour at low frequencies, either by a blocking capacitance at the electrodes or a weak DC current [10.89]. The theory leads to an analytical expression for the q-DC susceptibility in terms of a confluent hypergeometric function [10.24] (see [10.77] for the detailed expression), which is obtained from a response function that obeys an equation of motion similar to that of the permanent dipoles

$$\frac{d^2\phi}{dt^2} + \frac{\left(2+n+\frac{t}{\tau}\right)}{t}\frac{d\phi}{dt} + \frac{\left[n+\left(\frac{t}{\tau}\right)(1-p)\right]}{t^2}\phi = 0\,. \tag{10.47}$$

At times $t < \tau = 1/\omega_c$ the behaviour of the dipole density fluctuation is the same as that appropriate for permanent dipoles (10.22) with the same limiting solution, i. e., $\phi(t) \propto t^{-n}$ at $t < \tau$. At long times $t > \tau = 1/\omega_c$, however, it takes a different form in which the dipolar fluctuation relaxation has only a weak dependence on t, i. e., $\phi(t) \propto t^{p-1}$ at $t > \tau$. An example of the q-DC behaviour in frequency is given in Fig. 10.10b. The reader is referred to [10.78, 82, 87] for other examples.

In the limit of $p = 1$ ($\sigma(\omega) = \text{constant}; \omega < \omega_c$), the last term of (10.47) is zero in the limit $t > \tau$ at which time the decay of ϕ is governed by the first two terms. The solution for $\phi(t)$ now takes the form

$$\begin{aligned} &\phi(t) \propto t^{-n},\ t < \tau\,, \\ &\phi(t) \propto \exp\left(-\frac{t}{\tau}\right),\ t > \tau\,. \end{aligned} \tag{10.48}$$

In this case we can see that the dipole density fluctuations are produced within the local clusters and that they relax by freeing the charges to move independently at $t = \tau$. At this time there is a density of completely free charged particles with charge and countercharge cancelling throughout the system giving a net dipole of zero as in a true DC-conduction process.

10.5 Data Presentation

Throughout this chapter the dielectric data has been given in the log-log representation of a Bode plot. The reason has been to emphasise the power law wings of the dielectric loss peaks and thus demonstrate the non-exponential time dependence of the response function. There are many different representations that can be used (the reader is referred to [10.6] for details). Each of these representations contain the same information but emphasise different features of the response, thereby allowing them to be more accurately and easily evaluated. Researchers will of course use a representation that emphasises the features of most importance for their own work, but they should be aware that each representation may obscure other features to the detriment of a full interpretation of the data. Even if a computer fitting to an equivalent circuit containing frequency dependent components is used, the fitting can be misleading if the representation chosen obscures some of the components. A definite description can only be obtained if the same fitted components can reproduce the data in a number of different representations. A case in point is the use of the complex dielectric modulus

$$M^*(\omega) = \frac{1}{C^*(\omega)} = M'(\omega) - \mathrm{i}M''(\omega)\,.$$

This representation will emphasise weak loss peaks in $\chi_\mathrm{r}''(\omega)$ and allow a better identification of the loss peak frequency. It will also convert the q-DC response of Fig. 10.10b to a form in which $M''(\omega)$ has a peak at $\omega \approx \omega_\mathrm{c}$, which has been interpreted as the frequency at which ion hopping between local sites goes over to ion transport. However unless a Bode plot has been previously obtained showing clearly that a q-DC process is present rather than a DC conductivity in parallel with a constant capacitance the modulus plot is fraught with ambiguity because the modulus plot will show a peak in $M''(\omega)$ and an associated increase in $M'(\omega)$ from low frequencies that is very similar in shape and magnitude in both cases. Although an increase of several orders of magnitude in $M'(\omega)$ and $M''(\omega)$ as the frequency is raised from low values is indicative of either a parallel conductance or a q-DC process it is also possible that the peak may be confused with that arising from a dipole relaxation process if the measured frequency range does not extend very far towards low frequencies. The difficulty in ascertaining the nature of the processes is exacerbated if $M''(\omega)$ is plotted on its own as then the power law dependence of $M'(\omega)$ at low frequency cannot be easily shown to be the same as that of $M''(\omega)$ in the case of a q-DC process.

In order to avoid ambiguity it is therefore necessary to make a Bode-plot and other representations first and identify the various contributing processes, before making use of the modulus representation to emphasise the characteristic frequencies in order to evaluate them more accurately for use in analysis.

10.6 A Few Final Comments

The basic difficulty associated with the interpretation of dielectric responses is that they are of necessity macroscopic measurements made on samples that contain enormous numbers of atoms and molecules. In condensed-phase materials it is not possible to consider these systems as made up of local entities each moving independently of one another. All entities that contribute a permanent dipole are part of the condensed-phase structure, and even though they have a degree of freedom associated with the possibility of dipole reorientation, they will have motions that are correlated or connected to some extent to the molecules/atoms in their environment. This means that dipole reorientation is not that of a bare entity; instead it involves to some extent a local region. These regions will behave differently in different kinds of material and their definition and the way of describing their behaviour has not yet been established with any sort of rigour. Since we are dealing with a macroscopic measurement, there will of necessity be an ensemble of the local entities. This will result in a distribution of entities, but since these are part of the structure there will be some sort of connection between them unless the structure itself is disconnected dynamically. This means that fluctuations will take place among our entities, and perhaps even dissociation and amalgamation of the dynamically connected groups [10.90]. These effects will also have an influence on the relaxation dynamics. In the foregoing I have tried to give some simple pictures as to what is happening and to do so in a holistic way by correlating information from different facets of the measurement. What is abundantly clear is that the dynamics of such systems are very complicated in detail, but I hope that I have done enough to convince you that there are some basic features of the relaxation process that are common to all systems of this type, even though a full understanding of their nature does not yet exist.

References

10.1 P. Debye: *Polar Molecules* (Dover, New York 1945)
10.2 C.P. Smyth: *Dielectric Behaviour and Structure* (McGraw–Hill, New York 1955)
10.3 N.G. McCrum, B.E. Read, G. Williams: *Anelastic and Dielectric Effects in Polymeric Solids* (Wiley, New York 1967)
10.4 V.V. Daniels: *Dielectric Relaxation* (Academic, New York 1967)
10.5 C.J.F. Bottcher, P. Bordewijk: *Theory of Electric Polarisation*, Vol. I, II (Elesvier, Amsterdam 1978)
10.6 A.K. Jonscher: *Dielectric Relaxation in Solids* (Chelsea Dielectric, London 1983)
10.7 A.K. Jonscher: Relaxations in complex systems. In: *Proc. Int. Discuss. Meet. Relax. Complex Syst*, ed. by K.L. Ngai, G.B. Wright (Elsevier, Amsterdam 1991)
10.8 R.M. Hill: Dielectric properties and materials. In: *Electronic Materials: From Silicon to Organics*, ed. by L.S. Miller, J.B. Mullin (Plenum, New York 1991) p. 253
10.9 T. Furukawa: IEEE Trans. Electr. Insul. **24**, 375 (1989)
10.10 C.J.F. Bottcher: *Theory of Electric Polarisation* (Elsevier, Amsterdam 1952) p. 206
10.11 M. Born, E. Wolf: *Principles of Optics* (Pergamon, Oxford 1965)
10.12 A.R. von Hippel (Ed.): *Dielectric Materials and Applications* (Wiley, New York 1958)
10.13 R.N. Clarke, A. Gregory, D. Cannell, M. Patrick, S. Wylie, I. Youngs, G. Hill: Guide to the characterisation of dielectric materials at RF and microwave frequencies. In: *NPL Good Practice Guide* (Pub. Inst. Measurement and Control, London 2003)
10.14 F. Kremer, A. Schonhals: *Broadband Dielectric Spectroscopy* (Springer, Berlin, Heidelberg 2003)
10.15 W. Heitler: *The Quantum Theory of Radiation*, 3rd edn. (Dover, London 1984)
10.16 D. Bohm, D. Pines: Phys. Rev. **82**, 625 (1951)
10.17 P. Wheatley: *The Determination of Molecular Structure* (Clarendon, Oxford 1959)
10.18 H. Eyring, J. Walter, G.E. Kimball: *Quantum Chemistry* (Wiley, New York 1960)
10.19 R. Kubo: Rep. Prog. Phys. **29**, 255 (1966)
10.20 R. Kubo: J. Phys. Soc. Jpn. **12**, 570 (1957)
10.21 A.K. Jonscher: J. Phys. D Appl. Phys. **32**, R57 (1999)
10.22 S. Havriliak, S. Negami: J. Polym. Sci. **C**(14), 99 (1966)
10.23 S. Havriliak Jr., S.J. Havriliak: *Dielectric and Mechanical Relaxation in Materials* (Hanser, New York 1997)
10.24 M. Abramowitz, I.A. Stegun: *Handbook of Mathematical Functions* (Dover, New York 1965)
10.25 L.A. Dissado, R.M. Hill: Chem. Phys. **111**, 193 (1987)
10.26 L.A. Dissado, R.M. Hill: Proc. R. Soc. London **390**(A), 131 (1983)
10.27 L.A. Dissado, R.R. Nigmatullin, R.M. Hill: The fading of memory during the regression of structural fluctuations. In: *Dynamical Processes in Condensed Matter, Advances in Chemical Physics Ser*, Vol. LXIII, ed. by R.M. Evans (Wiley, New York 1985) p. 253
10.28 L.A. Dissado, R.M. Hill: Nature **279**, 685 (1979)
10.29 V.V. Novikov, V.P. Privalko: Phys. Rev. E **64**, 031504 (2001)
10.30 R.R. Nigmatullin: Physica B **358**, 201 (2005)
10.31 R.R. Nigmatullin, A.A. Arbuzov, F. Salehli, A. Giz, I. Bayrak, H. Catalgil-Giz: Physica B **388**, 418 (2007)
10.32 G. Williams, D.C. Watts: Trans. Faraday Soc. **66**, 80 (1970)
10.33 R. Kohlrausch: Pogg. Ann. Phys. **91**, 198 (1854)
10.34 A.K. Loidl, J. Knorr, R. Hessinger, I. Fehst, U.T. Hochli: J. Non-Cryst. Solids **269**, 131 (1991)
10.35 C. Kittel: *Introduction to Solid State Physics* (Wiley, New York 1966)
10.36 J. Joffrin, A. Levelut: J. Phys. **36**, 811 (1975)
10.37 P.C. Hohenberg, B.I. Halperin: Rev. Mod. Phys. **49**, 435 (1977)
10.38 J.G. Kirkwood: J. Chem. Phys. **7**, 911 (1939)
10.39 L.A. Dissado, R.M. Hill: Philos. Mag. B **41**, 625 (1980)
10.40 L.A. Dissado, M.E. Brown, R.M. Hill: J. Phys. C **16**, 4041 (1983)
10.41 L.A. Dissado, R.M. Hill: J. Phys. C **16**, 4023 (1983)
10.42 H.A. Kramers: Physica **7**(4), 284 (1940)
10.43 R.F. Grote, J.T. Hynes: J. Chem. Phys. **73**, 2715 (1980)
10.44 W.A. Phillips: Proc. R. Soc. London A **319**, 535 (1970)
10.45 J.G. le Gilchrist: Cryogenics **19**, 281 (1979)
10.46 R.M. Hill, L.A. Dissado: J. Phys. C **15**, 5171 (1982)
10.47 H. Eyring: J. Chem. Phys. **4**, 283 (1936)
10.48 S.H. Glarum: J. Chem. Phys. **33**, 1371 (1960)
10.49 C.A. Angell: Encyclopedia of materials. In: *Science and Technology*, Vol. 4, ed. by K.H.J. Buschow, R.W. Cahn, M.C. Fleming, B. Ilschner, E.J. Kramer, S. Mahajan (Elsevier, New York 2001) p. 3565
10.50 W. Kauzmann: Chem. Rev. **43**, 219 (1948)
10.51 G.P. Johari, M. Goldstein: J. Chem. Phys. **53**, 2372 (1970)
10.52 R.H. Boyd: Polymer **5**, 1123 (1985)
10.53 M.L. Williams, R.F. Landel, J.D. Ferry: J. Am. Chem. Soc. **77**, 3701 (1955)
10.54 M. Goldstein: J. Chem. Phys. **39**, 3369 (1963)
10.55 D. Turnbull, M.H. Cohen: J. Chem. Phys. **14**, 120 (1961)
10.56 R.R. Nigmatullin, S.I. Osokin, G. Smith: J. Phys. Condens. Matter **15**, 1 (2003)
10.57 R.M. Hill, L.A. Dissado, R.R. Nigmatullin: J. Phys. Condens. Matter **3**, 9773 (1991)
10.58 M. Shablakh, L.A. Dissado, R.M. Hill: J. Chem. Soc. Faraday Trans. **2**(79), 369 (1983)
10.59 R. Pirc, B. Zeks, P. Goshar: Phys. Chem. Solids **27**, 1219 (1966)
10.60 K. Pathmanathan, L.A. Dissado, R.M. Hill: Mol. Cryst. Liq. Cryst. **135**, 65 (1986)
10.61 K.L. Ngai, A.K. Jonscher, C.T. White: Nature **277**, 185 (1979)
10.62 K.L. Ngai, A.K. Rajgopal, S. Tietler: J. Phys. C **17**, 6611 (1984)
10.63 K.L. Ngai, R.W. Rendell, A.K. Rajgopal, S. Tietler: Ann. N. Y. Acad. Sci. **484**, 150 (1986)
10.64 R.G. Palmer, D. Stein, E.S. Abrahams, P.W. Anderson: Phys. Rev. Lett. **53**, 958 (1984)
10.65 K. Weron: J. Phys. Condens. Matter **4**, 10507 (1992)

10.66 K. Weron, A. Jurlewicz: J. Phys. A **26**, 395 (1993)
10.67 E.J. Gumbel: *Statistics of Extremes* (Columbia Univ. Press, New York 1958)
10.68 J.T. Bendler: J. Stat. Phys. **36**, 625 (1984)
10.69 J. Klafter, M.F. Schlesinger: Proc. Natl. Acad. Sci. **83**, 848 (1986)
10.70 D. Stauffer: *Introduction to Percolation Theory* (Taylor Francis, London 1985)
10.71 R.R. Nigmatullin: Theor. Math. Phys. **90**, 354 (1992)
10.72 R.R. Nigmatullin, A. Le Mehaute: Int. J. Sci. Geores. **8**, 2 (2003)
10.73 A. Doff, J.C. Gentilini, O. Cambruzzi: Mod. Phys. Lett. **B27**, 1350040 (2013)
10.74 L.A. Dissado, R.M. Hill: J. Appl. Phys. **66**, 2511 (1989)
10.75 R. Metzler, J. Klafter: Phys. Rep. **339**, 1 (2000)
10.76 A. Puzenko, P.B. Ishai, Y. Feldman: Phys. Rev. Lett. **105**, 037601 (2010)
10.77 L.A. Dissado, R.M. Hill: J. Chem. Soc. Faraday Trans. **2**(80), 291 (1984)
10.78 R.M. Hill, L.A. Dissado, K. Pathmanathan: J. Biol. Phys. **15**, 2 (1987)
10.79 T. Ramdeen, L.A. Dissado, R.M. Hill: J. Chem. Soc. Faraday Trans. **1**(80), 325 (1984)
10.80 R.R. Nigmatullin, L.A. Dissado, N.N. Soutougin: J. Phys. D **25**, 113 (1992)
10.81 A. Puzenko, N. Kozlovich, A. Gutina, Y. Feldman: Phys. Rev. B **60**, 14348 (1999)
10.82 L.A. Dissado: Phys. Med. Biol. **35**, 1487 (1990)
10.83 L.A. Dissado, R.M. Hill: Phys. Rev. B **37**, 3434 (1988)
10.84 D.P. Almond, C.R. Brown: Phys. Rev. Lett. **92**, 157601 (2004)
10.85 Y. Feldman, N. Kozlovich, Y. Alexandrov, R. Nigmatullin, Y. Ryabov: Phys. Rev. E **54**, 20 (1996)
10.86 I.I. Popov, R.R. Nigmatullin, A. Khamzin, I.V. Lounev: J. Appl. Phys. **112**, 094107 (2012)
10.87 M. Shablakh, L.A. Dissado, R.M. Hill: J. Biol. Phys. **12**, 63 (1984)
10.88 N.M. Chalashkanov, S.J. Dodd, L.A. Dissado, J.C. Fothergill: IEEE Trans. Dielectr. Electr. Insul. **21**, 1330 (2014)
10.89 L.A. Dissado, R.M. Hill, C. Pickup, S.H. Zaidi: Appl. Phys. Commun. **5**, 13 (1985)
10.90 L.A. Dissado, K.L. Khoo: J. Phys. D **39**, 3882 (2006)

11. Ionic Conduction and Applications

Harry Tuller

Solid state ionic conductors are crucial to a number of major technological developments, notably in the domains of energy storage and conversion and in environmental monitoring (such as battery, fuel cell and sensor technologies). Solid state ionic membranes based on *fast ion conductors* potentially provide important advantages over liquid electrolytes, including the elimination of sealing problems improved stability and the ability to miniaturize electrochemical devices using thin films. This chapter reviews methods of optimizing ionic conduction in solids and controlling the ratio of ionic to electronic conductivity in mixed conductors. Materials are distinguished based on whether they are characterized by intrinsic versus extrinsic disorder, amorphous versus crystalline structure, bulk versus interfacial control, cation versus anion conduction and ionic versus mixed ionic–electronic conduction. Data for representative conductors are tabulated.

A number of applications that rely on solid state electrolytes and/or mixed ionic–electronic conductors are considered, and the criteria used to choose such materials are reviewed. Emphasis is placed on fuel cells, sensors and batteries, where there is strong scientific and technological interest. The chapter concludes by considering how solid state ionic materials are likely to be used in the future, particularly in light of the trend for miniaturizing sensors and power sources and the interest in alternative memory devices based on memristors.

11.1 **Conduction in Ionic Solids** 248
11.2 **Fast Ion Conduction** 251
11.2.1 Structurally Disordered Crystalline Solids 251
11.2.2 Amorphous Solids 254
11.2.3 Heavily Doped Defective Solids 254
11.2.4 Interfacial Ionic Conduction and Nanostructural Effects 255
11.3 **Mixed Ionic–Electronic Conduction** 256
11.3.1 Defect Equilibria 256
11.3.2 Electrolytic Domain Boundaries 257
11.4 **Applications** 258
11.4.1 Sensors 258
11.4.2 Solid Oxide Fuel Cells (SOFC) 260
11.4.3 Membranes 261
11.4.4 Batteries 261
11.4.5 Electrochromic Windows 261
11.5 **Future Trends** 262
References 263

The ionic bonding of many refractory compounds allows for ionic diffusion and correspondingly, under the influence of an electric field, ionic conduction. This contribution to electrical conduction, for many years, was ignored as being inconsequential. However, over the past three to four decades, an increasing number of solids that support anomalously high levels of ionic conductivity have been identified. Indeed, some solids exhibit levels of ionic conductivity comparable to those of liquids. Such materials are termed *fast ion conductors*. Like solid state electronics, progress in *solid state ionics* has been driven by major technological developments, notably in the domains of energy storage and conversion and environmental monitoring, based on ongoing developments in battery, fuel cell and sensor technologies. More recently, this is being expanded to the memory device sphere as well. Some of the most important applications of solid state electronics and solid state ionics, and their categorization by type and magnitude of conductivity (such as dielectric, semiconducting, metallic and superconducting), are illustrated in Fig. 11.1 [11.1]. This figure also emphasizes that solids need not be strictly ionic or electronic, but may and often do exhibit mixed ionic–electronic conductivity. These *mixed conductors* play a critical role – particularly as electrodes – in solid state ionics, and

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_11

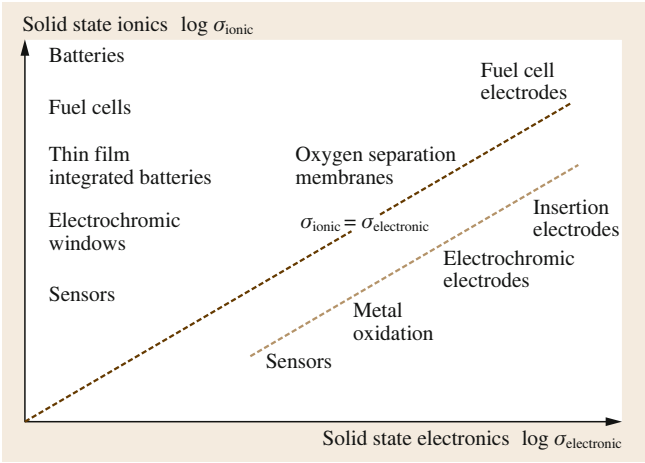


Fig. 11.1 Illustration of typical applications of ionic and electronic conductors as a function of the magnitude of electrical conductivity. Applications requiring mixed ionic electronic conductivity fall within the quadrant bounded by the two axes. (After [11.1])

are receiving comparable if not more attention than solid electrolytes at the present. Such solids which rely for their development on the intersection of a number of related fields including *solid state ionics*, *solid state electronics* and *solid state electrochemistry*, have grown in importance as our society has become more acutely concerned with efficient and environmentally clean methods for energy conversion, conservation and storage [11.2].

Solid state ionic membranes provide important potential advantages over liquids. The most important of these include:

1. Elimination of sealing problems associated with chemically reactive liquid or molten electrolytes
2. Minimization of discharge under open circuit conditions
3. Improved chemical stability under highly reactive conditions
4. The ability to miniaturize electrochemical devices through the use of thin films.

In the following, we begin by discussing methods of optimizing ionic conduction in solids and controlling the ratio of ionic to electronic conductivity. We then consider a number of applications that rely on solid state electrolytes and/or mixed ionic–electronic conductors and the criteria that should be used when selecting materials. We conclude by considering how solid state ionic materials are likely to be used in the future, particularly in light of trends related to the miniaturization of sensors, memory devices and power sources.

11.1 Conduction in Ionic Solids

The electrical conductivity, σ, the proportionality constant between the current density j and the electric field E, is given by

$$\frac{j}{E} = \sigma = \sum_i c_i Z_i q \mu_i \,, \tag{11.1}$$

where c_i is the carrier density (number/cm^3), μ_i the mobility (cm^2/Vs), and $Z_i q$ the charge ($q = 1.6 \times 10^{-19}$ C) of the ith charge carrier. The huge (many orders of magnitude) differences in σ between metals, semiconductors and insulators generally result from differences in c rather then μ. On the other hand, the higher conductivities of electronic versus ionic conductors are generally due to the much higher mobilities of electronic versus ionic species [11.3].

Optimized ionic conduction is a well-known characteristic of molten salts and aqueous electrolytes wherein all ions move with little hindrance within their surroundings. This leads to ionic conductivities as high as 10^{-1}–10^{1} S/cm in molten salts at temperatures of 400–900 °C [11.4]. Typical ionic solids, in contrast, possess limited numbers of mobile ions, hindered in their motion by virtue of being trapped in relatively stable potential wells. Ionic conduction in such solids easily falls below 10^{-10} S/cm for temperatures between room temperature and 200 °C. In the following sections, we examine the circumstances under which the magnitude of ionic conduction in solids approaches or even surpasses that found in liquid electrolytes.

The motion of ions is described by an activated jump process, for which the diffusion coefficient is given by [11.5]

$$\begin{aligned} D &= D_0 \exp\left(\frac{-\Delta G}{k_B T}\right) \\ &= \gamma(1-c) Z a^2 \nu_0 \exp\left(\frac{\Delta S}{k_B}\right) \exp\left(\frac{-E_m}{k_B T}\right), \end{aligned} \tag{11.2}$$

where a is the jump distance, ν_0 the attempt frequency, and E_m the migration energy. The factor $(1-c)Z$ defines the number of neighboring unoccupied sites, while

γ includes geometric and correlation factors. Note that the fractional occupation c here should not be confused with c_i, the charge carrier concentration, nor should the number of nearest neighbors Z be confused with Z_i, the number of charges per carrier defined in (11.1). Since the ion mobility is defined by $\mu_i = Z_i q D_i / k_B T$, where D_i and k_B are the diffusivity and Boltzmann constant respectively, and the density of carriers of charge $Z_i q$ is Nc, where N is the density of ion sites in the sublattice of interest, the ionic conductivity becomes

$$\begin{aligned}\sigma_{\text{ion}} &= \gamma \left[\frac{N(Z_i q)^2}{k_B T}\right] c(1-c) Z a^2 \nu_0 \\ &\quad \times \exp\left(\frac{\Delta S}{k_B}\right) \exp\left(\frac{-E_m}{k_B T}\right) \\ &= \left(\frac{\sigma_0}{T}\right) \exp\left(\frac{-E}{k_B T}\right) \end{aligned} \tag{11.3}$$

or

$$\begin{aligned}\sigma_{\text{ion}} &= \gamma N (Z_i q)^2 c(1-c) \frac{Z a^2 \gamma_0}{k_B T} \\ &\quad \times \exp\left(\frac{\Delta S}{k_B}\right) \exp\left(\frac{-E_m}{k_B T}\right). \end{aligned} \tag{11.4}$$

This expression shows that σ_{ion} is nonzero only when the product $c(1-c)$ is nonzero. Since all normal sites are fully occupied ($c = 1$) and all interstitial sites are empty ($c = 0$) in a perfect classical crystal, this is expected to lead to highly insulating characteristics. The classical theory of ionic conduction in solids is thus described in terms of the creation and motion of atomic defects, notably vacancies and interstitials.

Three mechanisms for ionic defect formation in oxides should be considered. These are:

1. Thermally induced intrinsic ionic disorder (such as Schottky and Frenkel defect pairs)
2. Redox-induced defects
3. Impurity-induced defects.

The first two categories of defects are predicted from statistical thermodynamics [11.6], and the latter form to satisfy electroneutrality. Examples of typical defect reactions in the three categories, representative of an ionically bonded binary metal oxide, are given in Table 11.1, in which the $K_i(T)$s represent the respective equilibrium constant and $a_{N_2O_3}$ the activity of the dopant oxide N_2O_3 added to the host oxide MO_2. Schottky and Frenkel disorder (1, 2) leave the stoichiometric balance intact. Reduction–oxidation behavior, as represented by (3), results in an imbalance in the ideal cation-to-anion ratio and thus leads to *nonstoichiometry*. Note that equilibration with the gas phase,

Table 11.1 Typical defect reaction

Defect reactions	Mass action relations	
$MO \leftrightarrow V''_M + V^{\cdot\cdot}_O$	$[V''_M][V^{\cdot\cdot}_O] = K_S(T)$	(1)
$O_O \leftrightarrow V^{\cdot\cdot}_O + O''_i$	$[V^{\cdot\cdot}_O][O''_i] = K_F(T)$	(2)
$O_O \leftrightarrow V^{\cdot\cdot}_O + 2e' + \frac{1}{2}O_2$	$[V^{\cdot\cdot}_O] n^2 = K_R(T) P_{O_2}^{-1/2}$	(3)
$O \leftrightarrow e' + h^{\cdot}$	$np = K_e(T)$	(4)
$N_2O_3(MO_2) \leftrightarrow 2N'_M + 3O_O + V^{\cdot\cdot}_O$	$[N'_M]^2 \cdot [V^{\cdot\cdot}_O] / a_{N_2O_3} = K_N(T)$	(5)

by the exchange of oxygen between the crystal lattice and the gas phase, generally results in the simultaneous generation of both ionic and electronic carriers. For completeness, the equilibrium between electrons and holes, via excitation across the band gap, is given in (4).

Altervalent impurities (for example N^{3+} substituted for the host cation M^{4+} – see (5)) also contribute to the generation of ionic carriers, commonly more than intrinsic levels do. This follows from the considerably reduced ionization energies required to dissociate impurity-defect pairs as compared to intrinsic defect generation. For example, E_A might correspond to the energy required to dissociate an acceptor–anion vacancy pair or E_D to the energy needed to dissociate a donor–anion interstitial pair. Such dissociative effects have been extensively reported in both the halide and oxide literature [11.7]. A more detailed discussion is provided below in the context of achieving high oxygen ion conductivity in solid oxide electrolytes.

The oxygen ion conductivity σ_i is given by the sum of the oxygen vacancy and interstitial partial conductivities. In all oxygen ion electrolytes of interest (in some mixed ionic conductors, such as La_2NiO_4 of interest as solid oxide fel cell cathodes, ionic conduction is largely via oxygen interstitials [11.8]), the interstitial does not appear to make significant contributions to the ionic conductivity, and so it is the product of the oxygen vacancy concentration $\left[V^{\cdot\cdot}_0\right]$, the charge $2q$, and the mobility (μ_v)

$$\sigma_i \approx \left[V^{\cdot\cdot}_0\right] 2q\mu_v \tag{11.5}$$

Optimized levels of σ_i obviously require a combination of high charge carrier density and mobility. Classically, high charge carrier densities have been induced in solids by substituting lower valent cations for the host cations [11.2]. Implicit in the requirement for high carrier densities are:

1. High solid solubility of the substituent with the lower valency
2. Low association energies between the oxygen vacancy and dopant
3. No long-range ordering of defects.

Additives which induce minimal strain tend to exhibit higher levels of solubility. The fluorite structure is the most well-known of these structures, with stabilized zirconia the best-known example. In this case, Y^{3+} substitutes for approximately 10% of Zr in $Zr_{1-x}Y_xO_{2-x/2}$, leading to $\sigma_i \approx 10^{-1}$ S/cm at 1000 °C and an activation energy of ≈ 1 eV. Other examples include CeO_2 [11.9], other fluorite-related structures such as the pyrochlores $A_2B_2O_7$ [11.10], and perovskites such as $La_{1-x}Sr_xGa_{1-y}Mg_yO_{3-\delta}$ (LSGM) [11.11].

Since the dopant and vacancy are of opposite charge (for example, Y'_{Zr} and $V_0^{\cdot\cdot}$), they tend to associate. With cations being much less mobile than oxygen ions, this serves to trap the charge carrier. It is of interest to examine how the concentration of *free* mobile carriers depends on the dopant concentration and the association energy. Consider the neutrality relation representing vacancy compensation of acceptor impurities by

$$N_V = \beta N_I \,, \tag{11.6}$$

where N_V and N_I are the vacancy and impurity densities while β reflects the relative charges of the two species and normally takes on values of 1 (for A''_M) and $\frac{1}{2}$ (for A'_M). The association reaction is given by

$$(I-V)^{x-y} \Leftrightarrow I^x + V^y \,, \quad \left(\beta = \frac{x}{y}\right), \tag{11.7}$$

where x and y are the relative charges of the impurity and vacancy, respectively. The corresponding mass action relation is then

$$\frac{N_I N_V}{N_{Dim}} = K_A^\circ \exp\left(\frac{-\Delta H_A}{k_B T}\right), \tag{11.8}$$

where N_{Dim} is the concentration of dimers and N_I and N_V are the corresponding defects remaining outside the complexes. It is straightforward to show that for weak dissociation (low temperatures or high association energies) one obtains the following solutions

$$\beta = 1: \quad N_V = \left(N_I K_A^0\right)^{\frac{1}{2}} \exp\left(\frac{-\Delta H_A}{2k_B T}\right), \tag{11.9}$$

$$\beta < 1: \quad N_V = \left(\frac{1-\beta}{\beta}\right) K_A^0 \exp\left(\frac{-\Delta H_A}{k_B T}\right). \tag{11.10}$$

The solution for condition $\beta = 1$ is the more familiar one. As in semiconductor physics [11.12], the number of free electrons or holes is proportional to the square root of the dopant density at reduced temperature, and it exhibits an Arrhenius dependence with activation energy that is equal to one half of the association or ionization energy. The solution for condition $\beta < 1$ is more unusual. Here one predicts that N_V is independent of dopant density! Also, the activation energy is predicted to be equal to the association energy.

At sufficiently high temperatures or low association energies, essentially all of the dimers are dissociated and

$$N_V = \beta N_I = \beta N_I(\text{total}) \,. \tag{11.11}$$

In general, therefore, two energies contribute to ionic conduction: a defect energy, E_D (which may either be related to the Frenkel or Schottky formation energy, or to a dissociation energy), and a migration energy E_m. The value of E in (11.3) therefore takes on different values in three characteristic temperature regimes. These include:

1. $E = E_m + E_A/2$: extrinsic associated regime at low T
2. $E = E_m$: extrinsic fully dissociated regime at intermediate T
3. $E = E_m + E_s/2$: intrinsic defect regime at elevated T (for instance, for Schottky equilibrium).

Such phenomena are clearly illustrated in recent publications dealing with donor and acceptor doping of TlBr, a gamma ray detector candidate material [11.13, 14]. By application of the appropriate defect model, values for $E_S = 0.91 \pm 0.03$ eV, $E_m(\text{Br}) = 0.28 \pm 0.05$ eV, $E_m(\text{Tl}) = 0.51 \pm 0.03$ eV and $E_A = 0.42 \pm 0.07$ eV ($Se'_{Br} - V^{\cdot}_{Br}$ associate) were extracted from the ionic conductivity data.

For optimized ionic conduction to exist, two criteria must be satisfied simultaneously. First the term c in (11.3) must approach 1/2. This corresponds to nearly all of the ions on a given sublattice being mobile. Second, the crystal structure must be arranged so as to enable easy motion of ions from one equivalent site to the next. This is reflected in exceptionally low values for the migration energy E_m. In the next section we discuss the conditions under which these criteria are satisfied.

11.2 Fast Ion Conduction

A number of routes leading to exceptionally high ion carrier densities in solids have been identified over the last few decades. These are subdivided into two major categories below (structurally disordered solids and highly defective solids). An important new development in recent years is the focus on the role of interfaces in creating ionic disorder localized in the vicinity of the boundaries. For nanosized structures, these disordered regimes may represent a large fraction of the overall volume of the material. Whatever the source of the enhanced ionic conductivity, such solids are commonly designated as *fast ion conductors* (FIC).

11.2.1 Structurally Disordered Crystalline Solids

In contrast to the idealized picture of crystal structures, many solids exist in which a sublattice of sites is only partially occupied. *Strock* [11.15, 16] already came to this conclusion in the 1930s in relation to the Ag sublattice in the high-temperature form (α-phase) of AgI. More recent neutron diffraction studies [11.17] differ with regard to the number of equivalent Ag sites. The special feature of partial occupancy of sites is nevertheless sustained.

Other notable systems characterized by sublattice disorder include Nasicon ($Na_3Zr_2PSi_2O_{12}$), sodium beta alumina (1.2 Na_2O–0.11 Al_2O_3) and $LiAlSiO_4$, which exhibit fast ion transport in three, two and one dimensions, respectively. Hundreds of other structurally disordered conductors may be found listed in review articles on the subject [11.2, 18–20]. Figure 11.2 illustrates the $\log \sigma - 1/T$ relations for representative FICs, while Table 11.2 summarizes data on representative materials in tabular form.

The rapid growth in the lithium battery industry has understandardly focused a great deal of attention on lithium ion solid electrolytes, given potentially lower capacity loss, improved cycle life, broader operation temperatures, enhanced safety and reliability, simplicity of design, and absence of leakage compared to liquid electrolytes [11.5]. The key factor limiting the application of lithium solid electrolytes has been their relatively low room temperature lithium ion conductivity. In recent years, new material compositions have been identified that exhibit much improved conductivities. A number of these are included in Fig. 11.3 along with representative organic liquid, ionic liquid and gel electrolytes. The thio-LISICON material $Li_{10}GeP_2S_{12}$ achieves a room temperature conductivity of 1.2×10^{-2} S/cm, surpassing the conductivity of many of the liquid and gel electrolytes. Several others are listed in Table 11.2.

Similarities between FICs and liquid electrolytes are often noted; the most important of these is that the disordered sublattice in the solid resembles the disordered nature of ions in a liquid. For this reason one often hears the term *lattice melting* used to describe phase transitions in solids which relate to the conversion of a conventional ionic conductor to a FIC (such as β to α transition in AgI at approximately 150 °C). Nevertheless, most investigators now believe that transport in FICs occurs via correlated jumps between well-defined sites rather than the liquid-like motion characteristic of aqueous or molten salt electrolytes.

The major structural characteristics of FICs include: (a) highly ordered, immobile or framework sublattice providing continuous open channels for ion transport, and (b) a mobile carrier sublattice which supports a random distribution of carriers over an excess number of equipotential sites. FICs exhibit framework sublattices that minimize strain, electrostatic and polarization contributions to the migration energy while offering high carrier concentrations within the mobile carrier sublattice. Given the high concentration of carriers, correlation effects between carriers must be taken into account. Calculations by *Wang* et al. [11.37], for ex-

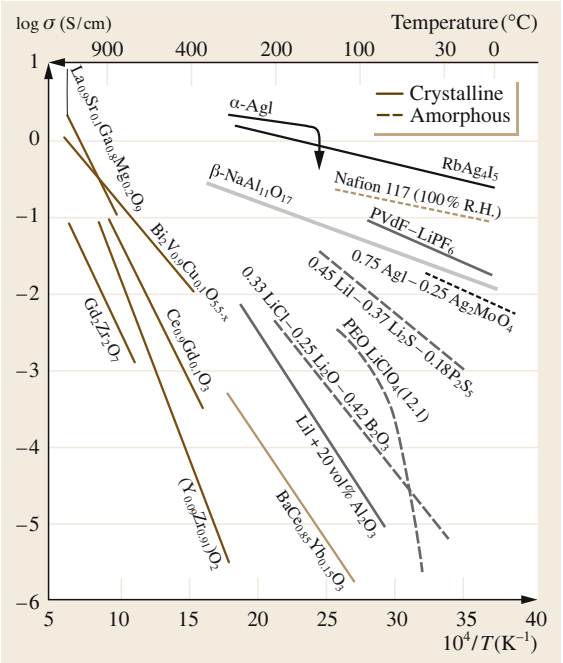


Fig. 11.2 The temperature dependences of representative FICs, including cation and anion conductors, crystalline and amorphous conductors, and inorganic and organic conductors. (After [11.3])

Table 11.2 Representative solid electrolytes and mixed conductors: mobile ions, electrical properties and apllications

Material	Mobile	Properties	Remarks and applications
$Zr_{0.85}Y_{0.15}O_{2-x}$ [11.21]	O^{2-}	$\sigma_{O^{2-}}$ (1000 °C) = 0.12 S/cm E = 0.8 eV	Material of choice in auto exhaust sensors; prime solid electrolyte candidate for solid oxide fuel cells (SOFCs)
$Ce_{0.95}Y_{0.05}O_{2-x}$ [11.22] $Ce_{0.1}Gd_{0.9}O_{2-x}$ [11.23, 24]	O^{2-}	$\sigma_{O^{2-}}$ (1000 °C) = 0.15 S/cm E = 0.76 eV $\sigma_{O^{2-}}$ (800 °C) = 0.10 S/cm E = 0.56 eV	Semiconducting at low PO_2; prime candidate solid electrolyte for intermediate temperature SOFCs
$La_{0.8}Sr_{0.2}Ga_{0.8}Mg_{0.2}O_{3-x}$ [11.25]	O^{2-}	$\sigma_{O^{2-}}$ (750 °C) = 0.35 S/cm E = 0.55 eV	Readily converts to mixed conductor by addition of transition metals; candidate solid electrolyte for intermediate temperature SOFCs
$Gd_{1.8}Ca_{1.2}Ti_2O_{7-x}$ [11.26]	O^{2-}	$\sigma_{O^{2-}}$ (1000 °C) = 0.05 S/cm E = 0.67 eV	The related pyrochlore, $Gd_2Zr_2O_7$, is an intrinsic FIC
$Bi_2V_{0.9}Cu_{0.1}O_{5.5-x}$ [11.27]	O^{2-}	$\sigma_{O^{2-}}$ (700 °C) = 0.15 S/cm E = 0.47 eV	Mixed ionic–electronic conductor, $t_{O^{2-}} \approx 0.9$ at 900 K; of interest as a permeation membrane
$La_{0.8}Sr_{0.2}MnO_{3+x}$ [11.28]	O^{2-}	$\sigma_{O^{2-}}$ (1000 °C) $\approx 3\times10^{-7}$ S/cm E = 2.81 eV	This material is largely an electronic conductor with σ_e (1000 °C) ≈ 100 S/cm; as a cathode in prime candidate SOFC
$BaCe_{0.85}Yb_{0.15}O_{3-x}$ [11.29]	H^+	σ_{H^+} (300 °C) = 7×10^{-4} S/cm E = 0.54 eV	Of interest in SOFC based on protonic conduction
Nafion [11.30]	H^+	σ_{H^+} (75 °C) $\approx 10^{-2}$ S/cm (at 20% relative humidity)	Organic; prime candidate for low-temperature solid state fuel cell based on protonic conduction
α-AgI [11.31]	Ag^+	σ_{Ag^+} (200 °C) = 1.6 S/cm E = 0.1 eV	Phase transition at 146 °C; first recognized fast ion conductor
$RbAg_4I_5$ [11.31]	Ag^+	σ_{Ag^+} (30 °C) = 0.3 S/cm E = 0.09 eV	One of the most conductive ionic conductors at room temperature
α-CuI [11.32]	Cu^+	σ_{Cu^+} (450 °C) = 10^{-1} S/cm E = 0.15 eV	Phase transition at 407 °C
Na β-alumina [11.33]	Na^+	σ_{Na^+} (300 °C) = 0.13 S/cm $E \approx$ 0.3 ev	Stoichiometry varies between Al_2O_3/Na_2O = 5.3–8.5
$60Li_2S$-$40SiS_2$ [11.34]	Li^+	σ_{Li^+} (25 °C) = 5×10^{-4} S/cm E = 0.25 eV	Amorphous
Poly(vinylidene fluoride) (PVdF) – Propylene carbonate (PC) – Li salt (LiX=$LiSO_3CF_3$ $LiPF_6$ or $LiN(SO_2CF_3)_2$ [11.35]	Li^+	σ_{Li^+} (20 °C) $\approx 10^{-3}$ S/cm	Organic conductor; of interest for lithium batteries
$Li_{10}GeP_2S_{12}$-*thio-LISICON*	Li^+	σ_{Li^+} (25 °C) = 1.2×10^{-2} S/cm E = 0.25 eV	Promising for lithium batteries; more conductive than Li ion liquid electrolytes
$Li_{3x}La_{(2/3)-x}TiO_3$ (x = 0.11)-perovskite	Li^+	σ_{Li^+} (25 °C) = 1×10^{-3} S/cm	Promising for lithium batteries; Stable to > 8 V but not in contact with Li metal
β-PbF_2 [11.36]	F^-	σ_{F^-} (100 °C) = 10^{-4} S/cm E = 0.48 eV	Basis of a variety of gas sensors

ample, have demonstrated that cooperative motions of ions can lead to significantly lower calculated migration energies than those based on consideration of isolated jumps alone. Although no precise criterion now exists for categorizing FICs, they normally exhibit unusually high ionic conductivities ($\sigma > 10^{-2}$ S/cm), well below their melting points, and generally low activation energies (commonly $E \approx 0.05$–0.5 eV, but they can be as high as ≈ 1.0 eV).

Few intrinsically disordered oxygen ion conducting FICs are known. The pyrochlores, with general formula $A_2^{3+}B_2^{4+}O_7$, represent a particularly interesting system, given that the degree of disorder can be varied almost continuously from low to high values within select solid-solution systems. The pyrochlore crystal structure is a superstructure of the defect fluorite lattice with twice the lattice parameter and one out of eight oxygens missing. This can be viewed as resulting from the need

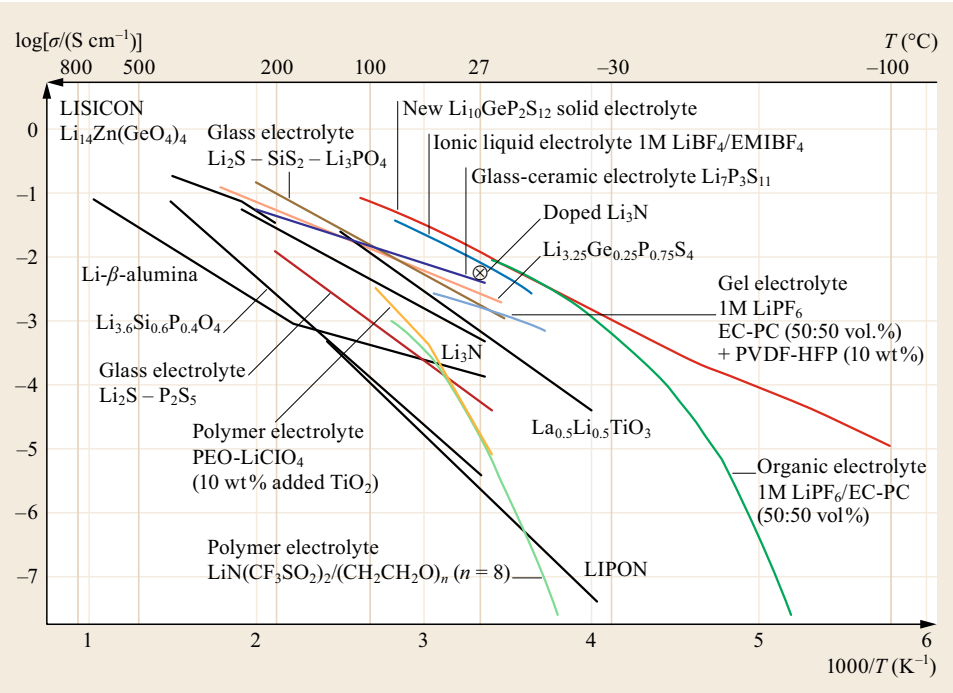


Fig. 11.3 The temperature dependences solid lithium ion conductors, together with representative organic liquid, ionic liquid and get electrolytes. (After [11.19])

to maintain charge neutrality after substituting trivalent ions for 50% of the quadravalent ions in the fluorite structure, as in $Gd_2Zr_2O_7$. Although oxygen vacancies occur at random throughout the anion sublattice in an ideal defect fluorite (such as yttria-stabilized zirconia – YSZ; Table 11.2), they are ordered onto particular sites in the pyrochlore structure. Thus, one properly views these as empty interstitial oxygen sites rather than oxygen vacancies. As a consequence, nearly ideal pyrochlore oxides, such as $Gd_2Ti_2O_7$, are ionic insulators [11.26]. Figure 11.4 illustrates the large increases in ionic conductivity induced by systematically substituting zirconium for titanium. *Moon* and *Tuller* [11.38] explain this on the basis of increased A and B cation antisite disorder as the radius of the B ion approaches that of the A ion. Thus, as the cation environments of the oxygen ions becoming more homogeneous, exchange between regular and interstitial sites also becomes more favorable, leading to increased Frenkel disorder. This interpretation has been confirmed by neutron diffraction studies on a closely related system [11.39].

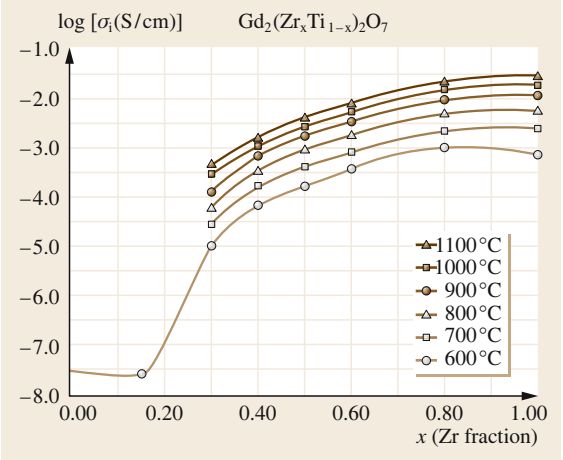


Fig. 11.4 Log ionic conductivity versus mole fraction Zr in $Gd_2(Zr_xTi_{1-x})_2O_7$ at a series of temperatures. (After [11.38])

Other important intrinsically disordered oxygen ion conductors are based on Bi_2O_3. At 730 °C [11.40], the low-temperature semiconducting modification transforms to the δ phase, which is accompanied by an oxide-ion conductivity jump of nearly three orders of magnitude. This is tied to the highly disordered fluorite-type structure, where a quarter of the oxygen sites are intrinsically empty, and to the high polarizability of the bismuth cation. *Takahashi* and *Iwahara* [11.41] succeeded in stabilizing the high-temperature δ phase to well below the transition temperature by doping with various oxides, including rare-earth oxides such as Y_2O_3. High oxide ion conductivity was also discovered above 570 °C in the Aurivillius-type γ phase of $Bi_4V_2O_{11}$ [11.42, 43], where one quarter of the oxygen sites coordinating V^{5+} are empty. Partial substitution of vanadium by lower valence cations, such as cop-

per, nickel or cobalt, led to a new family of so-called BIMEVOX compounds [11.44], with a remarkably high oxygen ion conductivity at moderate temperatures. The copper-substituted compound has an oxide ion conductivity above 200 °C which is ≈ 2 orders of magnitude higher than other oxide ion conductors. The bismuth-based electrolytes unfortunately suffer from instability under reducing conditions, which is a limitation for some applications, such as in the solid oxide fuel cells discussed below. However, *Wachsman* and *Lee* [11.45] demonstrate the ability to utilize the exceptional oxygen ion conductivity of bismuth oxide by pairing it with Gd doped ceria to form a bi-layer solid electrolyte in which the ceria exposed to the fuel environment protects the bismuth oxide from being too heavily reduced. Going one step further, *Sanna* et al. [11.46] report the ability to stabilize the δ-Bi_2O_3 phase to room temperature under both oxidizing and reducing conditions by creating layered superlattices with nano-dimensioned alternating layers of $Ce_{0.8}Gd_{0.2}O_{2-\delta}$ (gadolinium-doped ceria; CGO) and $Er_{0.4}Bi_{1.6}O_3$ (erbia-stabilized bismuth; ESB) on MgO single crystals. Interface and nanostructure effects are discussed in more detail below.

11.2.2 Amorphous Solids

One of the oft-mentioned criteria for FIC in solids (see before) is the existence of a highly ordered framework which provides channels for the ready motion of ions in the complementary, disordered sublattice. Reports of FIC in inorganic glasses [11.47] raised serious doubts concerning the relevance of this feature. The amorphous state, viewed as being liquid-like, is known to lack long-range order, with short-range order typically extending to, at most, a few atom spacings. Although highly oriented channels may be helpful in FIC, they are not essential, as demonstrated by the existence of FIC in glasses.

Fast ionic conductivity is observed in many glasses containing smaller cations with mole fractions greater than about 0.20, such as silver, copper, lithium and sodium [11.38, 47]. These glasses typically contain one or more network formers (such as SiO_2, B_2O_3, P_2O_5 or GeS_2), network modifiers (such as Ag_2O, Li_2O, Cu_2O or Ag_2S), and dopant compounds, largely halides (such as AgI, CuI and LiCl). The network structure and therefore its physical and chemical properties can be substantially modified by addition of the modifier. Dopant salts, on the other hand, do not strongly interact with the network, but *dissolve* into the interstices of the glass structure. A number of phenomenological trends have been noted including, ion conduction increases:

1. In the order K, Na, Li, Ag
2. With increasing modifier concentration
3. With halide additions in the order Cl, Br, I
4. In sulfide versus oxide glasses
5. In correlation with decreasing density and glass transition temperature.

Some authors speculate that ionic transport in glasses is enhanced upon addition of the halide anions by lowering the association energy between the mobile charge carriers and the network and thereby increasing the free carrier density [11.48]. An alternate model attributes the increased conductivity to major changes induced in the glass structure by the additives, as reflected in changes in glass transition temperature T_g and the density ρ. In this latter model, a large fraction of the carriers are already assumed to be unassociated and free to move, but with increased ionic mobility driven by structural changes. Here [11.49–52], the predominant influence of the halide addition is believed to impact the strain component of the migration energy.

Another important class of amorphous fast ion conductors is those based on organic or polymer electrolytes, which (analogous to the inorganic systems) are composed of a backbone polymer and a salt complex in which the counter-ion is covalently bound to the backbone. A classic example is the one based on polypropylene oxide $(CH_2CH(CH_3)O)_n$ (PPO) complexed with $LiCF_3SO_3$ to form $PPO_n \cdot LiCF_3SO_3$. Upon forming the complex, the Li ion conductivity increases by as much as a factor of $\approx 10^5$ [11.53]. In contrast to the inorganic glasses, which exhibit an Arrhenius temperature dependence, however, these polymers follow a curved dependence best expressed by

$$\sigma = \sigma_0 \exp\left(-\frac{B}{T - T_0}\right), \tag{11.12}$$

where T_0 is the glass transition temperature. This suggests a coupling between transport and network relaxation, a situation more closely coupled to transport in a liquid than in a solid, albeit a highly viscous liquid. Polymer electrolytes are now materials of choice for Li batteries and proton-based solid electrolytes given their attractive mechanical properties (ability to relax elastically upon stresses induced by volume changes related to charge–discharge of adjacent electrodes) and ease of processing [11.54].

11.2.3 Heavily Doped Defective Solids

Anomalously high concentrations of ionic carriers may also be induced in intrinsically insulating solids. In the following we briefly discuss two approaches for generating such *highly defective solids*.

We already know that ionic defect densities may be greatly enhanced above intrinsic levels by doping with altervalent impurities. However, the solubility limit of such impurities is often limited to only tens or hundreds of ppm. This corresponds to roughly $10^{17}-10^{18}$ defects/cm^3, a value 10^3-10^4 times smaller than in typical FICs. Compounds do exist, however, in which the solubility limit is extensive, reaching the 10–20% level even at reduced temperatures. Perhaps the most familiar example of such a system is stabilized zirconia, which due to its wide solid solubility with cations of lower valency such as Ca^{2+} and Y^{3+}, exhibits exceptionally high oxygen ion conductivity ($\sigma \approx 10^{-1}$ S/cm) at temperatures approaching 1000 °C.

As discussed above, high carrier densities must be coupled with high ion mobilities in order to attain high magnitudes of ionic conduction. The cubic fluorite structure, exhibited by stabilized zirconia (ZrO_2) and ceria (CeO_2), for example, supports high oxygen ion mobility due to the low four-fold coordination of cations around the oxygens, coupled with the interconnected nature of the face-shared polyhedra which surround the oxygen sites. Migration energies as low as ≈ 0.6 eV are reported for oxygen vacancy motion in ceria-based solid solutions [11.55]. High fluorine ion mobility is also observed in fluorite CaF_2 and related crystal systems.

More recently Ishihara demonstrated that very high oxygen conductivity can be achieved in the perovskite $LaGaO_3$ by accepter doping on both the La and Ga sites [11.11]. The solid solution $(La_{1-x}Sr_x)(Ga_{1-y}Mg_y)O_3$ exhibits ionic conductivity levels above that of ZrO_2 and CeO_2, for example 3×10^{-1} S/cm at 850 °C. Perovskites also support some of the highest proton conductivities at elevated temperatures. The most popular of these are ABO_3-type compounds with A = Ba, Sr, and B = Ce or Zr. Upon acceptor doping, as in $SrCe_{0.95}Yb_{0.05}O_3$, oxygen vacancies are generated as in the gallate above. However, in the presence of moisture, water is adsorbed and protons are generated [11.56]

$$H_2O + V_O^{\cdot\cdot} + O_O \Leftrightarrow 2OH^{\cdot} \,. \tag{11.13}$$

Given the high proton mobility, this is sufficient to induce large proton conductivity. Perovskite-related structures with the general formula $A_3B'B''O_9$ also exhibit high protonic conductivity [11.57]. Atomistic calculations simulating proton diffusion in numerous perovskite-type oxides are reported by the group of *Catlow* [11.58]. Ionic conductivities do not generally increase linearly with foreign atom additions. At the levels of defects being discussed here, defect–defect interactions become important, generally leading to defect ordering. This results in a maximum in ionic conductivity at some level of doping that depends on the particular system being investigated. *Nowick* et al. [11.55, 59] have demonstrated, in a series of studies, that the deviations from ideality are caused initially by composition-dependent activation energies rather than pre-exponentials (11.9), a feature also observed in a number of FIC glasses.

The formation of ionic defects that accompany excursions in composition away from stoichiometry due to redox reactions (Table 11.1, reaction 3) may also be large. CeO_2, for example, may be readily reduced to $CeO_{1.8}$ at 1000 °C and low PO_2s [11.60], resulting in oxygen vacancy concentrations of 5×10^{21} cm^{-3} ($c = 0.9$). It should be noted, however, that comparable concentrations of electrons are also formed during such stoichiometry excursions.

11.2.4 Interfacial Ionic Conduction and Nanostructural Effects

Interfaces can significantly modify the ionic conductivities of polycrystalline or composite materials and thin films. Modified levels of ionic conductivity near interfaces may result from space-charge regions formed near interfaces to compensate for charged defects and impurities segregated to surfaces, grain and phase boundaries. Grain boundaries, for example, serve as source and sink for impurities and point defects and thus often take on a net negative or positive charge relative to the grains. To maintain overall charge neutrality, a space charge of opposite charge forms in the grains adjacent to the grain boundaries with a width related to the Debye length L_D given by

$$L_D = \left(\frac{\varepsilon_r\varepsilon_0 k_B}{Tq^2 n_b}\right)^{\frac{1}{2}} \tag{11.14}$$

in which n_b is the majority charge carrier concentration within the grain, $\varepsilon_r\varepsilon_0$ the dielectric constant, k_B the Boltzmann constant, T the temperature and q is the electron charge. Depending on the sign of the charge at the interface, a depletion or accumulation of mobile ions in the vicinity of the boundary will form. Liang provided one of the first demonstrations of enhancement in the LiI : Al_2O_3 system [11.61].

The defect concentration profile in the space-charge region can be expressed as [11.62]

$$\frac{c_i}{c_i^\infty} = \exp\left[\frac{-q_i(\phi-\phi^\infty)}{k_B T}\right]. \tag{11.15}$$

The bulk concentration (c_i^∞) is a function of temperature, chemical potential and doping. The local concentration in the space-charge region (c_i) depends on

the difference between the bulk and the local electrical potential (ϕ^{∞} and ϕ). For positive values of ϕ, the concentrations of all negative defects are increased by the exponential factor, while those of the positive defects are decreased by the same factor and vice versa for negative values. When the mobile ion is depressed in vicinity of the grain boundary, this results in a reduction in ionic conductance due to blocking of ionic motion at the grain boundaries in polycrystalline specimens. This has been identified as a major obstacle in utilizing the perovskite based proton conductors described above in solid oxide fuel cells [11.63].

Films and/or polycrystalline materials with very small lateral dimensions can be expected to exhibit particularly strong space-charge effects on ionic conduction. This follows from the fact that the space-charge width approaches the dimensions of the film or grain. In this case, the space-charge regions overlap, and the defect densities no longer reach bulk values, even at the center of the particles [11.62]. In the limit of very small grains, local charge neutrality is not satisfied anywhere, and a full depletion (or accumulation) of charge carriers can occur with major consequences for ionic and electronic conductivity. Strong nanoscale effects on ionic and mixed ionic conductivity have been demonstrated for artificially modulated heterolayers of the solid ionic conductors CaF_2/BaF_2 [11.64] and nanocrystalline CeO_2 [11.65] and $Zr_{1-x}Y_xO_2/SrTiO_3$ superlattices [11.66] In the latter, as much as 8 orders of magnitude increase in the ionic conductivity of $Zr_{1-x}Y_xO_2$ was reported, although this interpretation has been questioned [11.67]. Figure 11.5 illustrates the orders of magnitude increase in fluorine ion conductivity possible with space-charge accumulation of mobile carriers in nanoscale CaF_2/BaF_2 multilayers.

In addition to the impact that interfaces may have on defect concentrations in their vicinity, others have investigated the potential role that interface-induced strain may have on defect mobility. In a series of studies by *Korte* et al. [11.67, 68], sandwiched $Zr_{1-x}Y_xO_2$ (YSZ) between insulating oxide layers with lattice mismatches with YSZ, thereby inducing both compressive and tensile stresses in the YSZ layers. As expected they found that dilatative strain increased the conductivity while compressive strain decreased the ionic conductivity, also consistent with a molecular dynamics study of YSZ that predicted considerably enhanced oxygen-ion diffusion for films subjected to dilatative strains [11.69]. These and other aspects of the roles of interfaces on impacting ionic conduction in heterostuctures are reviewed by *Fabbri* et al. [11.70]

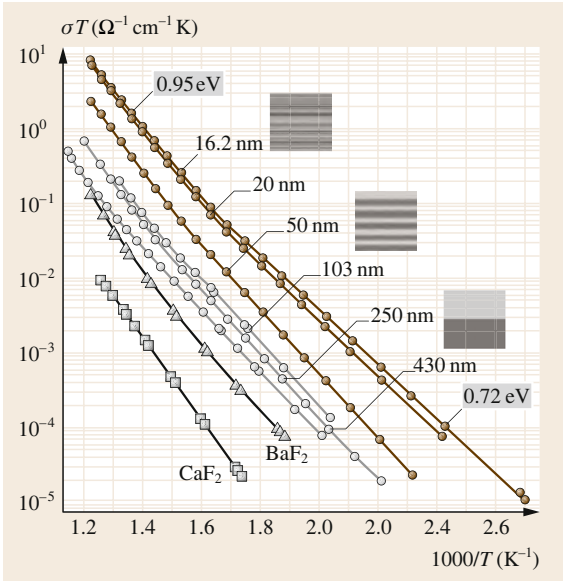


Fig. 11.5 Parallel ionic conductivity of CaF_2/BaF_2 nanometer-scale, artificially modulated heterolayers, with various periods and interfacial densities in the 430 to 16 nm range. (After [11.64])

11.3 Mixed Ionic–Electronic Conduction

11.3.1 Defect Equilibria

Deviations from stoichiometry in the direction of oxygen excess (MO_{1+x}) or deficiency (MO_{1-x}) form defect states that act identically in every way to impurity-related acceptor or donor states, respectively. In general, the electrical behavior of solids depends on defects formed in response to both impurities and deviations from stoichiometry. At or near stoichiometry, impurities predominate, while under strongly reducing or oxidizing conditions, defects associated with deviations from stoichiometry often take control. To characterize the electrical response of a metal oxide to temperature and atmosphere excursions, a series of simultaneous reactions of the form represented by (Table 11.1, reactions 1–5) must be considered. Furthermore, a representative electroneutrality equation for the case considered in Table11.1 would be

$$2\left[O_i''\right] + \left[N_M'\right] + n = 2\left[V_O^{\cdot\cdot}\right] + p\,. \qquad (11.16)$$

Note that: (1) intrinsic Frenkel disorder is assumed to predominate, so that (Table 11.1, reaction 1) may be ignored in subsequent discussions; (2) $a_{N_2O_3}$ is generally selected to be sufficiently low that all of N goes into solid solution. However, if exsolution of the dopant occurs e.g., upon cooling, it can also be treated in the framework of defect equilibria [11.14].

A piecewise solution to such problems is commonly attempted by sequentially choosing conditions for which only one term on either side of (11.16) need be considered. The region corresponding to mixed ionic conductivity is where the predominant charge carrier is an ion. In acceptor-doped material (A' being a generic acceptor), this corresponds to the condition (region II in Fig. 11.6) for which (11.16) may be simplified to read

$$[N'_M] = [A'_c] = 2[V^{\cdot\cdot}_O] \,. \tag{11.17}$$

Combining this with (Table 11.1, reaction 3) one obtains

$$n = \left[\frac{2K_R(T)}{[N'_M]}\right]^{\frac{1}{2}} P_{O_2}^{-\frac{1}{4}} \tag{11.18}$$

and from (Table 11.1, reactions 2 and 4),

$$p = K_e(T)\left(\frac{2K_R(T)}{[N'_M]}\right)^{-\frac{1}{2}} P_{O_2}^{\frac{1}{4}} \,, \tag{11.19}$$

$$[O''_i] = 2[A'_c]^{-1} K_F(T) \,. \tag{11.20}$$

Note that, in this defect regime, the ionic defects are pO_2-independent while the electronic species exhibit a $pO_2^{\pm 1/4}$ dependence. One obtains predictions for the corresponding dependencies of the partial conductivities of each of these charged species by multiplying carrier concentration by the respective charge and mobility. Experimentally, one normally observes the same pO_2 dependence of the partial conductivity as that predicted for the defect concentration, demonstrating that the mobility is pO_2-independent. One then uses the predicted pO_2 dependencies of the partial conductivities to deconvolute the ionic and electronic contributions to the electrical conductivity as discussed below.

The three other defect regimes most likely to occur, beginning at low P_{O_2} and moving on to increasing P_{O_2}, are depicted in Fig. 11.6 and include $n = 2[V^{\cdot\cdot}_O]$ (Region I), $p = [N'_M]$ (Region III) and $p = 2[O''_i]$ (Region IV).

In the case where μ_n, μ_p are sufficiently greater than $\mu(V^{\cdot\cdot}_O)$, then even in the defect regime where $V^{\cdot\cdot}_O$ is the predominant defect (so that $2[V^{\cdot\cdot}_O] = [N'_M]$), the total conductivity remains electronic. When the carrier mobility inequality is not nearly so pronounced, so that at the pO_2 at which electronic defects are at a minimum ($n = p$), conduction is predominantly ionic. Under these circumstances the oxide acts as a solid electrolyte, and in this regime of temperature and pO_2, one designates this as the *electrolytic domain*. Aside from the electrolytic domain, the neighboring zones on either side are designated as mixed zones within which both ionic and electronic conductivities are of comparable magnitude.

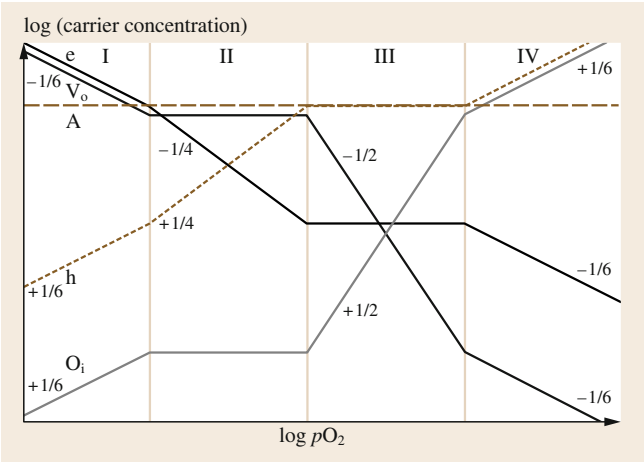


Fig. 11.6 Defect diagram for acceptor-doped oxide. (After [11.71])

11.3.2 Electrolytic Domain Boundaries

In applications where solid electrolytes are to be utilized, it is essential to know a priori under which conditions the material is likely to exhibit largely electrolytic characteristics. Expressions for the *electrolytic domain boundaries* can be obtained by first writing down general expressions for the partial conductivity (11.17)–(11.19)

$$\sigma_i = \sigma_i^\circ \exp\left(\frac{-E_i}{k_B T}\right), \tag{11.21}$$

$$\sigma_p = \sigma_p^\circ P_{o_2}^{+\frac{1}{4}} \exp\left(\frac{-E_p}{k_B T}\right), \tag{11.22}$$

$$\sigma_n = \sigma_n^\circ P_{o_2}^{-\frac{1}{4}} \exp\left(\frac{-E_n}{k_B T}\right). \tag{11.23}$$

One commonly defines the electrolytic domain boundary as that condition of T and pO_2 for which the ionic conductivity drops to 0.5 of the total conductivity. Under reducing conditions, this pO_2 is designated by P_n and under oxidizing conditions by P_p. Consequently,

one equates σ_i and σ_n or σ_i and σ_p to solve for P_n and P_p, respectively. These are given by

$$\ln P_n = \frac{-4(E_n - E_i)}{k_B}\frac{1}{T} + 4\ln\left(\frac{\sigma_n^0}{\sigma_i^0}\right), \quad (11.24)$$

$$\ln P_p = \frac{-4(E_i - E_p)}{k_B}\frac{1}{T} + 4\ln\left(\frac{\sigma_i^0}{\sigma_p^0}\right). \quad (11.25)$$

Note that since the mobilities of vacancies in such oxides have been found to be much greater than those of interstitials [11.38], we ignore the latter's contributions. The domain boundaries for stabilized zirconia are shown plotted in Fig. 11.7 [11.72]. Note that, as commonly observed, the electrolytic domain shrinks with increasing temperature due to the fact that E_n and E_p are typically greater in magnitude than E_i.

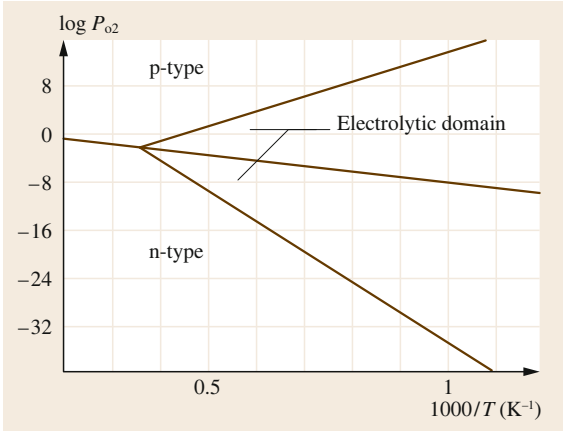


Fig. 11.7 Domain boundaries of stabilized zirconia as projected onto the log P_{o_2}–1000 T^{-1} plane. (After [11.72])

11.4 Applications

Fast ionic conducting ceramics and MIECs are finding extensive application in various solid state electrochemical devices. Some of these include fuel cells and electrolyzers [11.2, 45, 73–77] high energy density Li batteries [11.78–80], electrochromic windows [11.81] and auto exhaust sensors [11.82]. A very recent development is the proposed use of ionic or mixed ionic conductors as memory devices. These so call memristors can be switched from an insulating to a conductive state upon the application of a high electric field and switched back to the insulating state upon application of a suffiently high field of opposite polarity. The application of smaller fields are used to *read* the state of the memristor. Switching, whose speed can be below 10 ns, is believed to be driven by electromigration of either cations or anions. Such memristors are being designed to exceed 10^7 write cycles and retention times of > 10 yr [11.83].

11.4.1 Sensors

The monitoring of our environment has become essential for effective emissions control. Likewise, monitoring of chemical processes in real time enables closer quality control of products. Electrochemical sensors transform a chemical signal into an electrical signal, which is easy to measure, monitor and process [11.84]. Ionic and mixed conducting solids are basic materials for this development, because they can be easily miniaturized, for instance in thin film form, and they can often be operated at elevated temperatures or in an aggressive environment. The three major types of sensors utilizing ionic or mixed conducting materials are potentiometric amperometric and semiconducting, and are summarized below.

Potentiometric Sensors

In a potentiometric gas sensor, the concentration or partial pressure of a species is determined by measuring the emf of a solid electrolyte concentration cell. The most successful commercial sensor is the oxygen sensor [11.75], which uses stabilized zirconia as the solid oxygen ion electrolyte (Fig.11.8). The emf of the cell, $O_{2,ref}$, Pt/YSZ/Pt, O_2, can be written according to the Nernst equation

$$E = t_i\left(\frac{k_B T}{4q}\right)\ln\left[\frac{P(O_2)}{P(O_2)_{ref}}\right]. \quad (11.26)$$

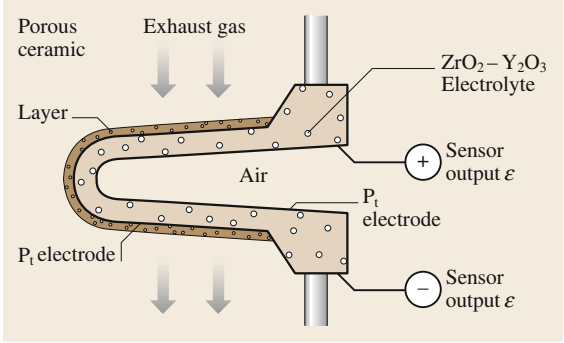


Fig. 11.8 Schematic of auto exhaust sensor based on the Nernst equation. (After [11.75])

$P(O_2)_{ref}$ is the oxygen partial pressure of the reference gas, generally air, and t_i is the ionic transference number. For proper operation, t_i needs to be kept very close to unity i. e., well within the electrolytic domain. All other terms have their common meanings.

The zirconia auto exhaust sensor monitors the air-to-fuel ratio, which is maintained within close limits for optimum operating efficiency of the three-way exhaust catalyst that serves to reduce the amount of pollutants, including unburned hydrocarbons, CO and NO_x. Such sensors are designed to provide response times on the order of tens of milliseconds. A voltage near to zero corresponds to an oxygen-rich *lean* mixture, and a voltage nearer to 1 V to an oxygen-poor *rich* mixture. The fuel injector of the engine is controlled via a closed-loop system. The zirconia oxygen sensor sees a wide range of exhaust temperatures, up to values as high as $\approx 900\,^{\circ}C$. Fortunately, the large step in voltage in going from lean to rich conditions can be easily detected by the on-board circuitry at all temperatures. The molecular mechanisms operating at the electrolyte/platinum interface have been examined in detail [11.85].

Another example of a potentiometric sensor is the one reported by *Maier* et al., of the type [11.86]

$$\mathrm{Au}, \mathrm{O_2}, \mathrm{CO_2}, \mathrm{Na_2CO_3}/\mathrm{Na^+}\text{-conductor}/\\ \mathrm{Na_2ZrO_3}, \mathrm{ZrO_2}, \mathrm{CO_2}, \mathrm{O_2}, \mathrm{Au}\,. \qquad (11.27)$$

This is used to monitor pCO_2; it eliminates the need for a gas-tight reference electrode. Here the two-phase reference electrode Na_2ZrO_3, ZrO_2, which fixes the Na activity on the right side of the cell, is insensitive to CO_2, while the change in Na activity in the Na_2CO_3 electrode on the left side can be sensed by the Na^+-conductor, typically NASICON or β-alumina. *Yamazoe* and *Miura* have reviewed the possible different types of potentiometric sensors by using single or multicomponent auxiliary phases [11.87].

Amperometric Sensors

By applying a voltage across an electrolyte, it is possible to electrochemically pump chemical species from one chamber to the other. Amperometric sensors rely on limiting current due to diffusion or interfacial phenomena at the electrode, which is linearly dependent on the partial pressure of the gas constituent [11.88, 89]. These become particularly important in so-called lean burn engines. Here the partial pressure of oxygen does not strongly vary with the air-to-fuel ratio, in contrast to engines operating at or near the stoichiometric air-to-fuel ratio. Under these circumstances, sensors are needed which have a stronger than logarithmic sensitivity to oxygen partial pressure variations. Sensors based on this principle are also being developed to detect other gases including NO_x.

Semiconducting Sensors

One can take advantage of the strong pO_2 dependence exhibited by n and p (see (11.22), (11.23)) in nonstoichiometric oxides as well to monitor the pO_2 in the surrounding gas phase by monitoring changes in their conductivity. Besides higher sensitivity, such resistive sensors offer low cost fabrication, potential for miniaturization and no need for seals and reference atmospheres, as required by zirconia-based sensors. A major limitation, however, is the significant cross sensitivity to temperature variations via the material's exponential dependence on temperature ((11.22), (11.23) large E_n and E_p). For $SrTiO_3$, $E_p = 1.3$ eV which would result in 2% change in resistance per degree Celsius, equivalent to an 8% change in pO_2 [11.90]. This temperature cross sensitivity renders $SrTiO_3$, and most other semiconducting oxides, unsuitable for this purpose. A solid solution between strontium titanate and ferrite, given by $Sr(Ti_{0.65}Fe_{0.35})O_{3-\delta}$ (STF35), on the other hand, has been found to exhibit a near zero temperature coefficient of resistance (ZTCR) [11.91]. This can be understood by examining the source of the temperature dependence of the electrical conductivity of metal oxide semiconductors given by

$$\sigma = \sigma_0 T^{(1.5-m)} \exp\left(\frac{-E_F}{k_B T}\right), \qquad (11.28)$$

where m reflects the power law dependence of mobility on temperature and E_F is the Fermi energy measured relative to the top of the valence band. Normally the first term in (11.28), which contributes to a positive coefficient of resistance PTCR, is overwhelmed by the second term, largely deriving from the exponential term, leading to a negative temperature coefficient of resistance (NTCR). However, with the reduced band gap resulting from the introduction of Fe and its lower lying 3d energy levels, the NTCR term becomes comparable to that of the PTCR term in (11.28), leading to a temperature insensitive oxygen sensor operation [11.91–93].

In addition to temperature sensitivity, for the oxygen sensor to be able to provide adequate feedback to the engine control unit, it must respond quickly (typically 10 ms) to changes in pO_2. For this to be possible, diffusion times given by $\tau_D \propto \ell^2/D$ must be exceptionally low, where D is the chemical diffusivity and ℓ the diffusion length which is the thickness for dense films and the grain radius for porous films. Screen printed STF35 thick film sensors are reported to satisfy these criteria, given the high oxygen diffusivity of this highly oxygen deficient material [11.94]. The response

time of thin films of STF, on the other hand, are limited instead by slow surface oxygen exchange kinetics. Controversy remains concerning the source(s) of this surface rate limiting step in mixed conducting oxides. For STF35, for example, surface chemistry, electronic structure and defect structure, all appear to play important roles [11.95].

11.4.2 Solid Oxide Fuel Cells (SOFC)

Solid oxide fuel cells (SOFC) provide many advantages over traditional energy conversion systems, including high energy conversion efficiency, fuel flexibility (due to internal reforming), low levels of NO_x and SO_x emissions, versatile plant size and long lifetimes [11.96]. Quiet, vibration-free operation also eliminates noise associated with conventional power-generation systems. However, operation at elevated temperature is necessary given relatively low ionic conductivities and slow electrode processes at temperatures below 800 °C. Recent progress with thinner electrolytes and advanced electrodes holds promise for reducting operating temperatures by as much as several hundred degrees.

The three major components of the elemental solid oxide fuel cell (SOFC) include the cathode, electrolyte and anode. While the solid electrolyte is selected so that it only conducts ions to ensure a Nernst open circuit potential that is as close to ideal as possible, the electrodes must support the reduction/oxidation reactions that occur at the electrolyte/electrode/gas interfaces. For example, when current is being drawn, the following reaction occurs at the cathode

$$\frac{1}{2}O_2 + 2e' + V_O^{\cdot\cdot} \Leftrightarrow O_O \,. \qquad (11.29)$$

This reaction is accelerated if the cathode can provide both electrons, as in a typical current collector, as well as oxygen vacancies. An example of such a mixed conducting cathode is $La_{1-x}Sr_xCoO_3$ (LSCO), which has an electronic conductivity of $> 100\,S/cm$ and an oxygen ion conductivity of $> 1\,S/cm$ at temperatures above 800 °C [11.97]. Given its importance to performance, modeling of the electrode processes has also received a great deal of attention recently [11.98, 99]. Unfortunately, while exhibiting highly attractive mixed conducting properties, LSCO is unstable in contact with yttria-stabilized zirconia, the electrolyte of choice.

In an attempt to take advantage of LSCO's attractive features, there is growing interest in marrying this electrode with doped ceria electrolytes, such as $Ce_{1-x}M_xO_2 : M = Gd$ or Sm, for operation at reduced temperatures of 550–750 °C, given ceria's higher ionic conductivity (albeit higher mixed ionic electronic conductivity at the anode) and stability in contact with LSCO. Electronic conduction degrades solid electrolyte performance in several ways. Because electronic conduction serves as an alternate path for charged species through the electrolyte, it decreases the power that can be dissipated through the load. Further, the short circuiting factor also serves to allow permeation of gaseous species through the electrolyte, even under open circuit conditions (see the section on membranes below). These primary figures of merit are summarized in the context of a solid oxide fuel cell (SOFC) in Fig. 11.9. E represents the potential induced across the cell under open circuit conditions for a given Po_2 gradient, t_i is the ionic transference number, E_N is the Nernst potential, R_{INT}, R_C, R_{SE} and R_A are the internal cell, cathode, solid electrolyte and anode resistances, respectively, Jo_2 is the oxygen permeation flux, and L the thickness across which the Po_2 gradient is imposed. All other terms have their normal meanings. Fortunately, the electronic conductivity in ceria electrolytes drops exponentially with decreasing temperature, and the overall power output exceeds that of zirconia-based systems at reduced temperatures, so it can be used with LSCO and other electrodes incompatible with YSZ.

The system $(La_{1-x}Sr_x)(Ga_{1-y}Mg_y)O_3$ (LSGM) was mentioned above as exhibiting one of the highest oxygen ion conductivities ($\approx 3\times10^{-1}\,S/cm$ at 850 °C) due to high levels of acceptor doping (Sr and Mg) on both the La and Ga sites. As a consequence, it is now being considered as one of several candidates for the electrolyte in solid oxide fuel cells. Experiments have shown that mixed ionic electronic conduction in a fuel cell electrode contributes to reduced overpotentials [11.74]. It has also been recognized that a single-phase *monolithic* fuel cell structure would benefit from the minimization of chemical

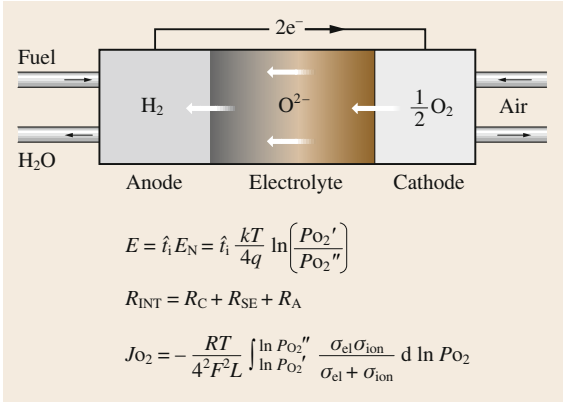


Fig. 11.9 Schematic of a solid oxide fuel cell and the primary figures of merit. (After [11.74])

and thermomechanical degradation [11.100]. Consequently, an electrode based on LSGM would satisfy all requirements. *Long* et al. proposed to add a transition metal in solution, which would introduce an additional 3d *conducting impurity* band within the wide band-gap of the initially electronically insulating gallate [11.101]. As expected, as the Ni content in the system $La_{0.9}Sr_{0.1}Ga_{1-x}Ni_xO_3$ (LSGN) increased, the electronic conductivity increased, finally reaching $\approx 50\,S/cm$ without decreasing the already high ionic conductivity. Improved electrode performance was indeed observed with the LSGM/LSGN interface [11.102]. Alternative options for a single-phase *monolithic* fuel cell structure could also, in principle, be based on fluorite structured ceria, with mixed conducting $Pr_xCe_{1-x}O_{2-\delta}$ cathode [11.103], the solid electrolyte $Gd_{0.2}Ce_{0.8}O_{1.9}$ and porous $Sm_{0.2}Ce_{0.8}O_{1.9}$ as the anode [11.104].

11.4.3 Membranes

Oxygen-permeable ceramic membranes are used for the separation of oxygen from air or for industrial-scale oxygen separation in the conversion of natural gas to syngas ($CO + H_2$) for example [11.105]. They are made from mixed conducting oxides in which ambipolar diffusion of ionic and electronic charge carriers in an oxygen potential gradient assures a high oxygen permeation flux through the membrane (Fig. 11.8 for an expression for the permeation current). High oxygen permeation rates were obtained with the system (La, Sr)$MO_{3-\delta}$ (M=Fe, Co, Cr) [11.97], but some deterioration over time was noticed. Research continues into this class of materials with regard to long-term ordering of defects, surface exchange kinetics, optimization of oxygen conduction and phase stability under steep oxygen activity gradients. One of the best materials developed to date is the BICUVOX compound, with composition $Bi_2V_{0.9}Cu_{0.1}O_{5.35}$, which shows a particularly large mixed conductivity that enables high oxygen permeation rates at moderate temperature, such as 700 K, at high and intermediate oxygen partial pressures [11.43]. Also receiving a great deal of attention is the perovskite $Ba_{0.5}Sr_{0.5}Co_{0.8}Fe_{0.2}O_{3-\delta}$ (BSCF) given its exceptionally high permeation rates, albeit with concerns about phase stability and reactivity with CO_2 in the gas phase [11.106].

Mixed oxide ion and electronic conductivity is also observed in composites of a solid oxide ion electrolyte and a noble metal, if percolating pathways exist for each component. These mixed conducting oxide ceramic–metal composites (cermets), including Y-stabilized ZrO_2 with Pd [11.107], Sm-doped CeO_2 with Pd [11.108], and rare-earth doped Bi_2O_3 with Ag [11.109], have an appreciable oxygen permeation rate at elevated temperature without degradation and are considered attractive for industrial applications, although they are relatively expensive. Recent work by *Takamura* et al. [11.110] shows promising results based on ceramic/ceramic composites.

11.4.4 Batteries

Power storage requires high energy density batteries. The highest possible energy density is achieved using reactants with high free energies of reaction and low mass, such as lithium or sodium reacting with elements high up in column 6 and 7 of the periodic table. This also requires that the solid electrolyte remains stable under highly reducing or highly oxidizing conditions. Major advantages of solid electrolytes over liquid electrolytes are the absence of leakage and container problems, better chemical stability, improved safety and the possibility of miniaturization; for example using thin solid films.

There is an increasing demand for microbatteries compatible with microelectronics technology, related to the development of laptop computers or portable cell phones. This led to the development of high energy density and long life-cycle rechargeable lithium batteries, initially based on metallic lithium anodes. However, systems based on metallic lithium suffered from problems due to metal oxidation and poor rechargeability due to the formation of metallic dendrites. The alternative *rocking chair* concept, proposed in 1980 [11.111], based on two lithium insertion compounds Li_xWO_3 and Li_yTiS_2, replaced the unstable lithium electrode, but was unable to provide sufficiently high energy densities. Improved rechargeable lithium ion batteries based instead on nongraphitic *hard* carbons as the lithium insertion anodes have since been developed [11.112]. This was followed by the successful association of hard carbon insertion anodes with the high-voltage $LiCoO_2$ insertion cathode. Due to the relatively high cost of Co, alternative systems based on other cathode materials, such as $LiNiO_2$ or $LiMn_2O_4$, are currently under investigation [11.113]. Polymer (rather than inorganic) electrolytes are used in these applications (see above). An overview of lithium batteries and polymer electrolytes can be found in books by *Julien* and *Nazri* [11.78] and *Gray* [11.114] and several more recent reviews [11.79, 80].

11.4.5 Electrochromic Windows

Electrochromic light transmission modulators – so-called *smart windows* that use solid ionic conductors –

may play a significant role in energy-saving by regulating thermal insulation. In such a system, the window is maintained transparent in the visible and reflecting in the IR during the winter, allowing penetration of sunshine but blocking loss of interior heat. On the other hand, the window is rendered partially opaque during hot summer days, reducing the amount of radiation entering the building. The electrochromic elements [11.115], which color or bleach upon insertion–deinsertion of lithium or hydrogen ions, are sandwiched between two transparent thin-film electrodes and are separated by a solid electrolyte. The transparent electrodes are generally indium tin oxide (ITO). Glass electrolytes appear to be promising choices. Research and development on tungsten trioxide-based electrochromic materials started in the 1970s. Ions fill empty tetrahedral sites in the WO_3 structure [11.116]

$$x\mathrm{Li}^+ + \mathrm{WO}_3 + x\mathrm{e}^- \rightarrow \mathrm{Li}_x\mathrm{WO}_3 \,. \tag{11.30}$$

Key requirements include:

1. Compatible electrochromic and solid electrolyte thin film materials
2. The ability to operate near ambient temperature, in the range −40 to +120 °C
3. Cycle reversibly many thousands of times a year for a lifetime of 20 yr
4. Exhibit significant shifts in reflectivity with the degree of insertion–deinsertion.

11.5 Future Trends

A rapidly converging interest in thin film oxides has been developing in the microelectronics and solid state ionics communities. In the solid state ionics arena, the desire to reduce the operating temperature of solid oxide fuel cells (SOFC) has been stimulating a shift in emphasis from bulk to thin film electrolytes. Likewise, the trend in recent years has shifted away from bulk ceramics towards miniaturized smart sensor systems in which the sensor elements are integrated with electronics and various MEMS-based components, including microheaters, valves and membranes. Considering the continued drive towards ever smaller submicron lateral dimensions in MOSFET technology, it is likely that future efforts will be directed towards the construction of micro- and nanoscale ionic devices.

Specifically, one can envision the embedding of miniaturized thin film or SOFC structures as sensors or power sources together with microelectromechanical (MEM) components and other active electronics in the same silicon wafer. By applying standard Si technology, such as thin film deposition and photolithography, one accesses methods for tailoring electrolyte and electrode geometry (thickness, active electrode area and triple phase boundary length) with exceptionally high dimensional reproducibility, while retaining the ability to scale to larger dimensions. Already mentioned, is the potential for widespread use of memristors as alternative nonvolatile memory devices which benefit from nanoscale dimensions, simple geometries, high switching speeds, low power requirements, long retention time and competitive switching cyles. Attention will need to be focused on the special challenges that the marriage between solid state ionics and electronics implies, including semiconductor-compatible processing, rapid temperature excursions, stress-induced property modifications and interfacial stability.

One will also need to consider how the defect and transport properties of thin films may differ from their bulk counterparts, and how silicon and other materials platform provides opportunities to examine such properties in an in situ manner, and thereby identify novel or distinctive properties associated with low-dimensional structures. Good progress along these lines have been made in recent years. For example, the multibeam optical stress sensor (MOSS) is used to measure stress, with high precision, induced in films, in situ, during growth, or subsequently by changes in temperature or pO_2, by optically determining the curvature of the support wafer. Stoichiometry changes in the films lead to stress given constraints imposed by the underlying substrates. Cyclic expansion–contraction experiments conducted on $TiO_{2-\delta}$ films grown on Si using an in situ MOSS system could be correlated with grain size demonstrating that grain boundaries are associated with significantly higher defect concentrations [11.117]. Chemical capacitance, derived from complex impedance spectroscopy, has recently been demonstrated to enable the nonstoichiometry of thin film $Pr_{0.1}Ce_{0.9}O_{2-\delta}$ to be measured with high precision and over an extensive range of temperature and pO_2. The authors found the thin film to have a lower enthalpy of reduction than the bulk counterpart leading to a greater loss of oxygen per unit volume in the film under the same operating conditions [11.118]. The same authors demonstrated an in situ optical absorption technique also capable of monitoring changes in stoichiometry in $Pr_{0.1}Ce_{0.9}O_{2-\delta}$ thin films and were

able to confirm that the chemical capacitance and optical absorption methods provide self-consistent results [11.119].

Materials, like those used in solid state batteries, fuel cells and permeation membranes often suffer large changes in stoichiometry. This is inevitable, for example, in the lithim battery cathode, such as Li_xCoO_2 in which Li is inserted and extracted during the charging and discharging of the battery. Such changes in stoichiometry carry along with them often large changes in lattice parameter and consequently large dimensional changes. These in turn create stresses that can lead to fracture. Alternatively, the induced strains can lead to changes in defect formation and mobility, stabilization of alternate phases, and modified surface chemistries. These phenomena are now being studied under the category of electro-chemo-mechanical phenomena that have recently been reviewed and categorized [11.120].

Acknowledgments. Support from the Department of Energy, Basic Energy Science (Award No. DE-SC0002633), the National Science Foundation (Award No. DMR-1507047) and the Skoltech-MIT Center for Electrochemical Energy Storage for topics related to this work are highly appreciated. In assembling this work, I drew on earlier journal and proceedings articles published by myself or in in conjunction with colleagues. In particular, I wish to particularly acknowledge joint publications with Prof. P. Knauth of the Aix Marseille Université, France and Dr. S.R. Bishop of MIT. Thanks also go to the International Institute of Carbon Neutral Energy Research (I2CNER), Kyushu University, for its hospitality and support.

References

11.1 H.L. Tuller: Oxygen ion conduction and structural disorder in conductive oxides, J. Phys. Chem. Solids **55**, 1393 (1994)

11.2 P. Knauth, H.L. Tuller: Solid state ionics: roots, status and future prospects, J. Am. Ceram. Soc. **85**, 1654 (2002)

11.3 H.L. Tuller: Highly conducting ceramics. In: *Ceramic Materials for Electronics*, 3rd edn., ed. by R.C. Buchanan (Marcel Dekker, New York 2004) p. 87

11.4 M.K. Paria, H.S. Maiti: Electrical conductivity and defect structure of polycrystalline tin dioxide doped with antimony oxide, J. Mater. Sci. **17**, 3275 (1982)

11.5 J.B. Goodenough: Skeleton structures. In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) p. 393

11.6 F.A. Kröger: *The Chemistry of Imperfect Crystals*, 2nd edn. (North-Holland, Amsterdam 1974)

11.7 J. Hladik (Ed.): *Physics of Electrolytes*, Vol. 1 (Academic, New York 1972)

11.8 J.M. Bassat, P. Odier, A. Villesuzanne, C. Marin, M. Pouchard: Anisotropic ionic transport properties in $La_2NiO_{4+\delta}$ crystals, Solid State Ionics **167**, 341 (2004)

11.9 B.C.H. Steele: Oxygen transport and exchange in oxide ceramics, J. Power Sources **49**, 1 (1994)

11.10 S. Kramer, M. Spears, H.L. Tuller: Conduction in titanate pyrochlores: role of dopants, Solid State Ionics **72**, 59 (1994)

11.11 T. Ishihara, H. Matsuda, Y. Takita: Doped $LaGaO_3$ perovskite type oxide as a new oxide ionic conductor, J. Am. Ceram. Soc. **116**, 3801 (1994)

11.12 S. Wang: *Fundamentals of Semiconductor Theory and Device Physics* (Prentice Hall, Englewood Cliffs 1989)

11.13 S.R. Bishop, W. Higgins, G. Ciampi, A. Churilov, K.S. Shah, H.L. Tuller: The defect and transport properties of donor doped single crystal TlBr, J. Electrochem. Soc. **158**, J47 (2011)

11.14 S.R. Bishop, H.L. Tuller, J. Engel, W. Higgins, G. Ciampi, A. Churilov, K.S. Shah: The defect and transport properties of acceptor doped TlBr: Role of dopant exsolution and association, Phys. Chem. Chem. Phys. **14**, 10160 (2012)

11.15 L.W. Strock: Z. Phys. Chem. **B25**, 441 (1934)

11.16 L.W. Strock: Z. Phys. Chem. **B25**, 132 (1936)

11.17 R. Cava, F. Reidinger, B.J. Wuensch: Single-crystal neutron-diffraction study of AgI between 23 ° and 300 °C, Solid State Commun. **24**, 411 (1977)

11.18 T. Kudo: Survey of types of electrolytes. In: *The CRC Handbook of Solid State Electrochemistry*, ed. by P.J. Gellings, H.J.M. Bouwmeester (CRC, Boca Raton 1997) p. 195

11.19 N. Kamayay, K. Homma, Y. Yamakaway, M. Hirayama, R. Kanno, M. Yonemura, T. Kamiyama, Y. Kato, S. Hama, K. Kawamoto, A. Mitsui: A lithium superionic conductor, Nat. Mater. **10**, 682 (2011)

11.20 C. Cao, Z.-B. Li, X.-L. Wang, X.-B. Zhao, W.-Q. Han: Recent advances in inorganic solid electrolytes for lithium batteries, Front. Energ. Res. **2**, 1 (2014)

11.21 R.M. Dell, A. Hooper: Oxygen ion conductors. In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) p. 291

11.22 H.L. Tuller, A.S. Nowick: Doped ceria as a solid oxide electrolyte, J. Electrochem. Soc. **122**, 255 (1975)

11.23 B.C.H. Steele: Materials science and engineering: The enabling technology for the commercialization of fuel cell systems, J. Mater. Sci. **36**, 1053 (2001)

11.24 V.M. Vishnyakov: Proton exchange membrane fuel cells, Vacuum **80**, 1053 (2006)

11.25 T. Ishihara, T. Shibayama, M. Honda, H. Nishiguchi, Y. Takita: Intermediate temperature solid oxide fuel cells using $LaGaO_3$ electrolyte II. Improvement of oxide ion conductivity and power density by doping Fe for Ga site of $LaGaO_3$, J. Electrochem. Soc. **147**, 1332 (2000)

11.26 S.A. Kramer, H.L. Tuller: A novel titanate-based oxygen ion conductor: $Gd_2Ti_2O_7$, Solid State Ionics **82**, 15 (1995)

11.27 A.A. Yaremchenko, V.V. Kharton, E.N. Naumovich, F.M.B. Marques: Physicochemical and transport properties of BiCuVOx-based ceramics, J. Electroceram. **4**, 233 (2000)

11.28 R.A. De Souza, J.A. Kilner, J.F. Walker: A SIMS study of oxygen tracer diffusion and surface exchange in $La_{0.8}Sr_{0.2}MnO_{3+\delta}$, Mater. Lett. **43**, 43 (2000)

11.29 J. Wu, L.P. Li, W.T.P. Espinosa, S.M. Haile: Defect Chemistry and Transport Properties Of $Ba_xCe_{0.85}M_{0.15}O_{3-\delta}$, J. Mater. Res. **19**, 2366 (2004)

11.30 F. Damay, L.C. Klein: Transport properties of Nafion composite membranes for proton-exchange membranes fuel cells, Solid State Ionics **162–163**, 261 (2003)

11.31 J.S. Kasper: Inorganic silver ion conductors. In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) p. 217

11.32 T. Matsui, J.B. Wagner Jr.: Inorganic copper ion conductors. In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) p. 237

11.33 T. Takahashi: In: *Superionic Solids and Solid Electrolytes: Recent Trends*, ed. by A.L. Laskar, S. Chandra (Academic, San Diego 1989) p. 1

11.34 F.A. Fusco, H.L. Tuller: Fast ion transport in glasses. In: *Superionic Solids and Solid Electrolytes: Recent Trends*, ed. by A.L. Laskar, S. Chandra (Academic, San Diego 1989) p. 43

11.35 Z. Jiang, B. Carroll, K.M. Abraham: Studies of some poly(vinylidene fluoride) electrolytes, Electrochim. Acta **42**, 2667 (1997)

11.36 J.-M. Reau, J. Portier: Fluorine ion conductor. In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) p. 313

11.37 J. Wang, M. Kaffari, D. Choi: On the ionic conduction in β-alumina: potential energy curves and conduction mechanism, J. Chem. Phys. **63**, 772 (1975)

11.38 P.K. Moon, H.L. Tuller: Ionic conduction in the $Gd_2Ti_2O_7$-$Gd_2Zr_2O_7$ system, Solid State Ionics **28–30**, 470 (1988)

11.39 B.J. Wuensch, K.W. Eberman, C. Heremans, E.M. Ku, P. Onnerud, S.M. Haile, J.K. Stalick, J.D. Jorgensen: Solid State Ionics **12**(9), 111 (2000)

11.40 H.A. Harwig, A.G. Gerards: Electrical properties of the α, β, γ, and δ phases of bismuth sesquioxide, J. Solid State Chem. **26**, 265 (1978)

11.41 T. Takahashi, H. Iwahara: Oxide ion conductors based on bismuthsesquioxide, Mater. Res. Bull. **13**, 1447 (1978)

11.42 K.R. Kendall, C. Navas, J.K. Thomas, H.-C. zur Loye: Recent developments in oxide ion conductors: Aurivillius phases, Chem. Mater. **8**, 642 (1996)

11.43 J.C. Boivin, G. Mairesse: Recent material developments in fast oxide ion conductors, Chem. Mater. **1**(0), 2870 (1998)

11.44 F. Abraham, J.C. Boivin, G. Mairesse, G. Nowogrocki: The BiMeVox series: A new family of high performances oxide ion conductors, Solid State Ionics **4**(0–1), 934 (1990)

11.45 E. Wachsman, K.T. Lee: Lowering the temperature of solid oxide fuel cells, Science **334**, 935 (2011)

11.46 S. Sanna, V. Esposito, J.W. Andreasen, J. Hjelm, W. Zhang, T. Kasama, S.B. Simonsen, M. Christensen, S. Linderoth, N. Pryds: Enhancement of the chemical stability in confined δ-Bi_2O_3, Nat. Mater. **14**, 500 (2015)

11.47 H.L. Tuller, D.P. Button, D.R. Uhlmann: Fast ion transport in oxide glasses, J. Non-Cryst. Solids **40**, 93 (1980)

11.48 D. Ravaine, J.L. Souquet: A thermodynamic approach to ionic conductivity in oxide glasses – part 1. Correlation of the ionic conductivity with the chemical potential of alkali oxide in oxide glasses, Phys. Chem. Glasses **18**, 27 (1977)

11.49 D.P. Button, R.P. Tandon, H.L. Tuller, D.R. Uhlmann: Fast Li+ ion conduction in chloro-borate glasses, J. Non-Cryst. Solids **42**, 297 (1980)

11.50 D.P. Button, R.P. Tandon, H.L. Tuller, D.R. Uhlmann: Fast Li+ ion conductance in chloroborate glasses, II: diborates and metaborates, Solid State Ionics **5**, 655 (1981)

11.51 F.A. Fusco, H.L. Tuller, D.P. Button: Lithium, sodium and potasoium transport in fast ion conducting glasses: Trends and models. In: *Proc. Symp. Electro-Ceramics and Solid State Ionics*, ed. by H.L. Tuller, D.M. Smyth (Electrochemical Society, Pennington 1988) p. 167

11.52 D.P. Button, P.K. Moon, H.L. Tuller, D.R. Uhlmann: Structure-transport relationship in oxide glasses: fast ion transport in alkali borate glasses, Glastech. Ber. **56K**, 856 (1983)

11.53 M.A. Ratner, P. Johansson, D.F. Shriver: Polymer electrolytes: Ionic transport mechanisms and relaxation coupling, MRS Bull. **25**, 31 (2000)

11.54 B. Scrosati, C.A. Vincent: Polymer electrolytes: The key to lithium polymer batteries, MRS Bull. **25**, 28 (2000)

11.55 D.Y. Wang, D.S. Park, J. Griffiths, A.S. Nowick: Oxygen-ion conductivity and defect interactions in yttria-doped ceria, Solid State Ionics **2**, 95 (1981)

11.56 H. Iwahara, T. Esaka, H. Uchida, H. Maeda: Proton conduction in sintered oxides and its application to steam electrolysis for hydrogen production, Solid State Ionics **3**(4), 359 (1981)

11.57 A.S. Nowick, Y. Du: High-temperature protonic conductors with perovskite-related structures, Solid State Ionics **7**(7), 137 (1995)

11.58 M. Cherry, M.S. Islam, J.D. Gale, C.R.A. Catlow: Computational studies of protons in perovskite-structured oxides, J. Phys. Chem. **99**(40), 14614

(1995)
11.59 R. Gerhart-Anderson, A.S. Nowick: Ionic conductivity of CeO_2 with trivalent dopants of different ionic radii, Solid State Ionics **5**, 547 (1981)
11.60 H.L. Tuller, A.S. Nowick: Defect structure and electrical properties of nonstoichiometric CeO_2 single crystals, J. Electrochem Soc. **126**, 209 (1979)
11.61 C.C. Liang: Conduction characteristics of the lithium iodide-aluminum oxide solid electrolytes, J. Electrochem Soc. **120**, 1289 (1973)
11.62 J. Maier: Ionic conduction in space charge regions, Prog. Solid State Chem. **23**, 171 (1995)
11.63 D. Pergolesi, E. Fabbri, A. D'Epifanio, E. Di Bartolomeo, A. Tebano, S. Sanna, S. Licoccia, G. Balestrino, E. Traversa: High proton conduction in grain-boundary-free yttrium-doped barium zirconate films grown by pulsed laser deposition, Nat. Mater. **9**, 846 (2010)
11.64 H. Sata, K. Eberman, K. Eberl, J. Maier: Mesoscopic fast ion conduction in nanometre-scale planar heterostructures, Nature **408**, 946 (2000)
11.65 Y.-M. Chiang, E.B. Lavik, I. Kosacki, H.L. Tuller, J.Y. Ying: Nonstoichiometry and electrical conductivity of nanocrystalline CeO_{2-x}, J. Electroceram. **1**, 7 (1997)
11.66 J. Garcia-Barriocanal, A. Rivera-Calzada, M. Varela, Z. Sefrioui, E. Iborra, C. Leon, S.J. Pennycook, J. Santamaria: Colossal ionic conductivity at interfaces of epitaxial ZrO_2:Y_2O_3/$SrTiO_3$ heterostructures, Science **321**, 676 (2008)
11.67 X. Guo: Comment on "Colossal Ionic Conductivity At Interfaces Of Epitaxial Zro_2:Y_2O_3/$SrTiO_3$ Heterostructures", Science **324**, 465 (2009), http://www.sciencemag.org/cgi/content/full/324/5926/465a
11.68 N. Schichtel, C. Korte, D. Hesse, J. Janek: Elastic strain at interfaces and its influence on ionic conductivity in nanoscaled solid electrolyte thin films-theoretical considerations and experimental studies, Phys. Chem. Chem. Phys. **11**, 3043 (2009)
11.69 A. Kushima, B. Yildiz: Oxygen ion diffusivity in strained yttria stabilized zirconia: Where is the fastest strain?, J. Mater. Chem. **20**, 4809 (2010)
11.70 E. Fabbri, D. Pergolesi, E. Traversa: Ionic conductivity in oxide heterostructures: The role of interfaces, Sci. Technol. Adv. Mater. **11**, 054503 (2010)
11.71 H. Seh: Langasite Bulk Acoustic Wave Resonant Sensor for High Temperature Applications, Ph.D. Thesis, Dept. Materials Sc. and Eng. (MIT, Cambridge 2005)
11.72 L. Heyne: Electrochemistry of mixed ionic-electronic conductors. In: *Solid Electrolytes*, ed. by S. Geller (Springer, Berlin, Heidelberg 1977) p. 169
11.73 N.Q. Minh, T. Takahashi: *Science and Technology of Ceramic Fuel Cells* (Elsevier, Amsterdam 1995)
11.74 H.L. Tuller: Materials design and optimization. In: *Oxygen Ion and Mixed Conductors and their Technological Applications*, ed. by H.L. Tuller, J. Schoonman, I. Riess (Kluwer, Dordrecht 2000) p. 245
11.75 B.M. Kulwicki, S.J. Lukasiewicz, S. Subramanyam, A. Amin, H.L. Tuller: Ceromic sensors. In: *Ceramic Materials for Electronics*, 3rd edn., ed. by R.C. Buchanan (Marcel Dekker, New York 2004) p. 377
11.76 J.A. Kilner, M. Burriel: Intermediate-temperature solid-oxide fuel cells, Annu. Rev. Mater. Res. **44**, 365 (2014)
11.77 S.P.S. Badwal, S. Giddey, C. Munnings, A. Kulkarni: Review of progress in high temperature solid oxide fuel cells, J. Aust. Ceram. Soc. **50**(1), 23 (2014)
11.78 C. Julien, G.-A. Nazri: *Solid State Batteries: Materials Design and Optimization* (Kluwer, Boston 1994)
11.79 J. Chen: Recent progress in advanced materials for lithium ion batteries, Materials **6**, 156 (2013)
11.80 J.B. Goodenough, K.-S. Park: The Li-ion rechargeable battery: A perspective, J. Am. Chem. Soc. **135**, 1167 (2013)
11.81 C.G. Granqvist: Electrochromism and electrochiomic derives. In: *The CRC Handbook of Solid State Electrochemistry*, ed. by P.J. Gellings, H.J.M. Bouwmeester (CRC, Boca Raton 1997) p. 587
11.82 P.T. Moseley, B.C. Tofield (Eds.): *Solid State Gas Sensors* (Adam Hilger, Bristol 1987)
11.83 R. Waser, R. Dittmann, G. Staikov, K. Szot: Redox-based resistive switching memories – Nanoionic mechanisms, prospects and challenges, Adv. Mater. **21**, 2632 (2009)
11.84 W. Göpel, T.A. Jones, M. Kleitz, I. Lundström, T. Seiyama (Eds.): *3*, Sensors: A Comprehensive Survey, Chemical and Biochemical Sensors, Vol. 2 (Wiley, New York 1991)
11.85 A.D. Brailsford, M. Yussouff, E.M. Logothetis: A first-principles model of the zirconia oxygen sensor, Sensor. Actuat. B **44**, 321 (1997)
11.86 J. Maier, M. Holzinger, W. Sitte: Fast potentiometric CO_2 sensors with open reference electrodes, Solid State Ionics **74**, 5 (1994)
11.87 N. Yamazoe, N. Miura: Potentiometric gas sensors for oxidic gases, J. Electroceram. **2**, 243 (1998)
11.88 T. Takeuchi: Oxygen sensors, Sensor. Actuat. **14**, 109 (1988)
11.89 W. Göpel, G. Reinhardt, M. Rosch: Trends in the development of solid state amperometric and potentiometric high temperature sensors, Solid State Ionics **136–137**, 519 (2000)
11.90 E. Ivers-Tiffeé, K.H. Härdtl, W. Menesklou, J. Riegel: Oxygen sensors for lean combustion engines. In: *Ceramic Materials and Components for Engines*, ed. by J.G. Heinrich, F. Aldinger (Wiley, Weinheim 2001) p. 39
11.91 A. Rothschild, W. Menesklou, H.L. Tuller, E. Ivers-Tiffée: Electronic structure, defect chemistry, and transport properties of $SrTi_{1-x}Fe_xO_{3-y}$ solid solutions, Chem. Mater. **18**, 3651 (2006)
11.92 A. Rothschild, H.L. Tuller: Resistive oxygen sensors. In: *Science and Technology of Chemiresistor Gas Sensors*, ed. by D.K. Aswal, S.K. Gupta (Nova Science, Hauppauge 2007) p. 215
11.93 H.L. Tuller: Materials for high temperature electrochemical applications: Automotive sensors,

Catal. Traps, Sensors Actuators B **187**, 106 (2013)
11.94 E. Ivers-Tiffée, K.H. Härdtl, W. Menesklou, J. Riegel: Principles of solid state oxygen sensors for lean combustion gas control, Electrochimica Acta **47**, 807 (2001)
11.95 Y. Chen, W. Jung, Z. Cai, J.J. Kim, H.L. Tuller, B. Yildiz: Impact of Sr Segregation on the electronic structure and oxygen reduction activity of $SrTi_{1-x}Fe_xO_3$ surfaces, Energ. Environ. Sci. **5**, 7979 (2012)
11.96 S.C. Singhal: Science and technology of solid-oxide fuel cells, MRS Bull. **25**, 16 (2000)
11.97 H.J.M. Bouwmeester, A.J. Burggraaf: Dense ceramic membrances for oxygen seperation. In: *The CRC Handbook of Solid State Electrochemistry*, ed. by P.J. Gellings, H.J.M. Bouwmeester (CRC, Boca Raton 1997) p. 481
11.98 I. Riess, M. Godickemeier, L.J. Gauckler: Characterization of solid oxide fuel cells based on solid electrolytes or mixed ionic electronic conductors, Solid State Ionics **90**, 91 (1996)
11.99 S.B. Adler, J.A. Lane, B.C.H. Steele: Electrode kinetics of porous mixed-conducting oxygen electrodes, J. Electrochem. Soc. **143**, 3554 (1996)
11.100 S.A. Kramer, M.A. Spears, H.L. Tuller: Novel compatible solid electrolyte-electrode system suitable for solid state electrochemical cells, Patent US55 403 461 (1995)
11.101 N.J. Long, F. Lecarpentier, H.L. Tuller: Structure and electrical properties of Ni-substituted lanthanum gallate perovskites, J. Electroceram. **3:4**, 399 (1999)
11.102 F. Lecarpentier, H.L. Tuller, N. Long: Performance of $La_{0.9}Sr_{0.1}Ga_{0.5}Ni_{0.5}O_3$ as a cathode for a lanthanum gallate fuel cell, J. Electroceram. **5**, 225 (2000)
11.103 D. Chen, S.R. Bishop, H.L. Tuller: Praseodymium-cerium oxide thin film cathodes: Study of oxygen reduction reaction kinetics, J. Electroceramics **28**, 62 (2012)
11.104 W.C. Jung, J.O. Dereux, W.C. Chueh, Y. Hao, S.M. Haile: High electrode activity of nanostructured, columnar ceria films for solid oxide fuel cells, Energ. Environ. Sci. **5**, 8682 (2012)
11.105 B. Ma, U. Balachandran: Transport properties of the mixed-conducting $SrFe_{1.2}Co_{0.3}O_x$, J. Electroceram. **2**, 135 (1998)
11.106 W.K. Hong, G.M. Choi: Oxygen permeation of BSCF membranes with varying membrane thickness and surface coating, J. Membrane Sc. **346**, 353 (2010)
11.107 T.J. Mazanec: Prospects for ceramic electrochemical reactors in industry, Solid State Ionics **70/71**, 11 (1994)
11.108 K. Huang, M. Schroeder, J.B. Goodenough: Oxygen permeation through composite oxide-ion and electronic conductors, J. Electrochem. Solid State Lett. **2**, 375 (1999)
11.109 J.E. ten Elshof, N.Q. Nguyen, M.W. Den Otter, H.J.M. Bouwmeester: Oxygen permeation properties of dense $Bi_{1.5}Er_{0.5}O_3$-Ag cermet membranes, J. Electrochem. Soc. **144**, 4361 (1997)
11.110 H. Takamura, K. Okumura, Y. Koshino, A. Kamegawa, M. Okada: Oxygen permeation properties of ceria-ferrite-based composites, J. Electroceram. **13**, 613 (2004)
11.111 M. Lazzari, B. Scrosati: A cyclable lithium organic electrolyte cell based on two intercalation electrodes, J. Electrochem. Soc. **127**, 773 (1980)
11.112 J.R. Dahn, A.K. Sleigh, H. Shi, B.M. Way, W.J. Weydanz, J.N. Reimers, Q. Zhong, U. von Sacken: Carbons and graphites as substitutes for the lithium anode. In: *Lithium Batteries*, ed. by G. Pistoia (Elsevier, Amsterdam 1994) p. 1
11.113 J.-M. Tarascon, W.R. McKinnon, F. Coowar, T.N. Bowmer, G. Amatucci, D. Guyomard: Synthesis conditions and oxygen stoichiometry effects on Li insertion into the spinel $LiMn_2O_4$, J. Electrochem. Soc. **141**, 1421 (1994)
11.114 F.M. Gray: *Solid Polymer Electrolytes: Fundamentals and Technological Applications* (VCH, New York 1991)
11.115 C.M. Lampert: Electrochromic materials and devices for energy efficient windows, Solar Energy Mat. **11**, 1 (1984)
11.116 M. Green: WO_3-based electrochromic windows – problems and status, Ionics **5**, 161 (2000)
11.117 S. Bhatia, B.W. Sheldon: Compositional stresses in polycrystalline titania films, J. Am. Ceram. Soc. **91**, 3986 (2008)
11.118 D. Chen, S.R. Bishop, H.L. Tuller: Nonstoichiometry in oxide films operating under anodic conditions: A chemical capacitance study of the praseodymium-cerium oxide system, Chem. Mat. **26**, 6622 (2014)
11.119 J.J. Kim, S.R. Bishop, N.J. Thompson, D. Chen, H.L. Tuller: Investigation of nonstoichiometry in oxide thin films by simultaneous in situ optical absorption and chemical capacitance measurements: Pr doped ceria-case study, Chem. Mat. **26**, 1374 (2014)
11.120 S.R. Bishop, D. Marrocchelli, C. Chatzichristodoulou, M. Mogensen, N.H. Perry, H.L. Tuller, E.D. Wachsman: Chemical expansion: Implications for electrochemical energy storage and conversion devices, Annu. Rev. Mater. Res. **44**, 205 (2014)

Part B Growth a

Part B Growth and Characterization

12 **Bulk Crystal Growth: Methods and Materials**
Peter Capper, Southampton, UK

13 **Single-Crystal Silicon: Growth and Properties**
Fumio Shimura, Fukuroi, Japan

14 **Epitaxial Crystal Growth: Methods and Materials**
Peter Capper, Southampton, UK
Stuart Irvine, Swansea, UK
Tim Joyce, Liverpool, UK

15 **Narrow Bandgap II-VI Semiconductors: Growth**
Peter Capper, Southampton, UK

16 **Wide-Bandgap II-VI Semiconductors: Growth and Properties**
Minoru Isshiki, Sendai, Japan
Jifeng Wang, Tokyo, Japan

17 **Structural Characterization**
Paul D. Brown, Nottingham, UK

18 **Surface Chemical Analysis**
David Sykes, Loughborough, UK

19 **Thermal Properties and Thermal Analysis: Fundamentals, Experimental Techniques and Applications**
Safa Kasap, Saskatoon, Canada
Jiří Málek, Pardubice, Czech Republic
Roman Svoboda, Pardubice, Czech Republic

20 **Electrical Characterization of Semiconductor Materials and Devices**
M. Jamal Deen, Hamilton, Canada
Fabien Pascal, Montpellier, France

12. Bulk Crystal Growth: Methods and Materials

Peter Capper

This chapter covers the field of bulk single crystals of materials used in electronics and optoelectronics. These crystals are used in both active and passive modes, that is, to produce devices directly in/on bulk-grown slices of material, or as substrates in epitaxial growth, respectively. Single-crystal material usually provides superior properties to polycrystalline or amorphous equivalents. The various bulk growth techniques are outlined, together with specific critical features, and examples are given of the types of materials, and their current typical sizes, grown by these techniques. Materials covered range from group IVs (Si, Ge, SiGe, diamond, SiC), group III–Vs (e.g., such as GaAs, InP, nitrides, etc.) group II–IVs (e.g., CdTe, ZnSe, HgCdTe (MCT), etc.) through to a wide range of oxide/halide/phosphate/borate/tungstate materials. This chapter is to be treated as a snapshot only; the interested reader is referred to the remainder of the chapters in this handbook for more specific growth and characterization details on the various materials outlined in this chapter. Neither does this chapter cover the more fundamental aspects of the growth of the particular materials covered; again the reader is referred to relevant chapters within the handbook, or to other sources of information in the general literature.

12.1 **Background** 269
12.2 **History** 270
12.3 **Techniques** 271
12.3.1 Verneuil 271
12.3.2 Czochralski 271
12.3.3 Kyropoulos 273
12.3.4 Stepanov 273
12.3.5 Edge-Defined Film Growth 273
12.3.6 Bridgman 273
12.3.7 Vertical Gradient Freeze 275
12.3.8 Float Zone 275
12.3.9 Traveling Heater Method 275
12.3.10 Low-Temperature Solution Growth 276
12.3.11 High-Temperature Solution Growth (Flux) 276
12.3.12 Hydrothermal 277
12.3.13 Growth from the Vapor 278
12.3.14 Multicrystalline Si Growth for Solar Cells 278
12.4 **Materials Grown** 278
12.4.1 Group IV 278
12.4.2 Groups III–V 280
12.4.3 Groups II–VI 283
12.4.4 Oxides/Halides/Phosphates/Borates 286
12.5 **Conclusions** 289
References 289

12.1 Background

Despite the widespread progress in several epitaxial growth techniques for producing electronic and optoelectronic device-quality material, various bulk growth methods are still used to produce tens of thousands of tons of such materials each year. These crystals are used in both active and passive modes; i. e., to produce devices directly in/on bulk-grown slices of material, or used as substrates in epitaxial growth, respectively.

This chapter covers the field of bulk single crystals of materials used in electronics and optoelectronics. Single-crystal material usually provides superior properties to polycrystalline or amorphous equivalents. The various bulk growth techniques are outlined, together with the specific critical features, and examples are given of the types of materials, and their typical current sizes, grown by these techniques. Materials covered range from group IVs (Si, Ge, SiGe, diamond, SiC), group III–Vs (e.g., GaAs, InP, InSb, GaSb, nitrides, etc.) group II–IVs (e.g., CdTe, ZnSe, HgCdTe (MCT), ZnO, etc.) through to a wide range of oxide/halide/phosphate/borate/tungstate materials. This chapter is to be treated as a snapshot only; the interested reader is referred to the remainder of the chapters in this

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_12

handbook for more specific growth and characterization details on the various materials outlined in this chapter. Many of the crystals grown at a small size mainly for R&D purposes, particularly in Universities, are not discussed, and neither, in general, are organic materials. Neither does this chapter cover the more fundamental aspects of the growth of the particular materials covered; again the reader is referred to the relevant chapters within this book, or to other sources of information in the general literature.

12.2 History

Several very useful studies on the history of crystal growth in general can be found in the literature [12.1–9]. There were many significant contributions made to the fundamentals of crystal growth during the 18th and 19th centuries, including the development of thermodynamics, undercooling, and supersaturation [12.6, 7]. In terms of crystal growth techniques, it is accepted that the first to produce usable crystals on a large scale was that of flame fusion by *Verneuil* [12.10].

Before World War II, synthetic crystals (other than ruby) were mainly used in scientific instruments. However, between 1900 and 1940 there were enormous advances in both the theoretical aspects and in producing samples for scientific study. The diffusion boundary layer was applied by *Nernst* [12.11], while ideas on the growth of perfect crystals were proposed by *Volmer* [12.12], *Kossel* [12.13], and *Stranski* [12.14]. Also during this period many of the growth techniques now used were also initially developed. The flame fusion of *Verneuil* [12.10] was followed by hydrothermal growth [12.15, 16], crystal pulling [12.17], *Kyropoulos* [12.18], *Bridgman* growth [12.19–22], and vertical gradient freeze (VGF) [12.23].

During World War II, synthetic crystals were used as piezoelectric transducers (in sonar) as resonant devices (in radar) and as infrared detectors. However, it was the invention of the transistor (announced in 1948) that heralded the modern era of crystal growth for practical purposes rather than for pure scientific interest. *Teal* and *Little* developed crystal pulling of Ge in 1950 [12.24] and *Dash* [12.25] improved this with the *necking* technique to produce dislocation-free material. A key aspect at this stage was that of impurity levels and means to measure them. The semiconductors that were used for such solid-state devices as transistors work because they have desired dopants incorporated selectively in them. Techniques were developed at this time to improve the purity of materials, for example, zone refining [12.26] and float-zone (FZ) refining [12.27, 28]. Theoretical aspects developed in this period include the role of screw dislocations in growth [12.29] and the generalized theory of *Burton* et al. [12.30]. Reducing melt inclusions in crystals was discussed by *Ivantsov* [12.31, 32] in terms of diffusional undercooling, and by *Tiller* et al. [12.33] in terms of constitutional supercooling. Similar work was carried out in solution growth by *Carlson* [12.34], who studied flow effects across crystal faces, and *Scheel* and *Elwell* [12.35], who derived the maximum stable growth rate and optimized supersaturation to produce inclusion-free crystals. Seg-

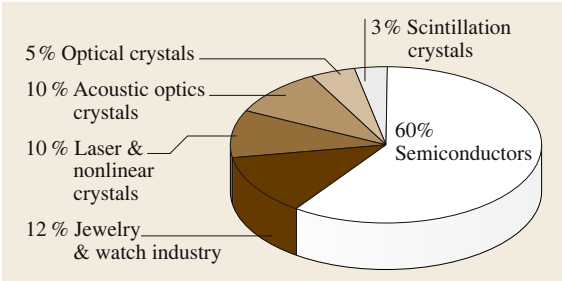


Fig. 12.1 Estimated share of the world production of 20 000 t of bulk crystals. (After [12.1], copyright Elsevier Science, used with permission)

Table 12.1 Estimated worldwide annual production rates of crystals (as at 1986, [12.5]). Some materials are mainly used for non-(electronic/optoelectronic) purposes, for example, cubic zirconia and much of the ruby are used in jewelry, watches, etc. (Items in *brackets* are small in volume but high in value)

Crystal	Rate (t/yr)	Growth methods
Silicon	4000	Czochralski, float-zone, (vapor phase epitaxy VPE)
Metals	4000	Bridgman, Strain anneal
Quartz	800	Hydrothermal
III–V compounds	600	Czochralski, Bridgman (VPE, liquid phase epitaxy LPE)
Alkali halides	500	Bridgman, Kyropoulos
Ruby	500	Verneuil
Germanium	400	Czochralski, Bridgman
Garnets	200	Czochralski
Lithium niobate	100	Czochralski
Phosphates	50	Low-temperature solution
Lithium tantalate	20	Czochralski
Cubic zirconia	15	Skull melting
TGS	10	Low-temperature solution
Diamond	10	High-temperature solution
II–VI compounds	5	Vapor, Bridgman

regation effects, which are related to mass and heat transfer, were studied by *Burton* et al. [12.36] for melt growth and by *van Erk* [12.37] for solution growth, while experimental conditions to produce striation-free material were established by *Rytz* and *Scheel* [12.38]. Forced convection in diffusion-limited growth was recognized as being beneficial for open systems with stirrers [12.39–42], while stirring in sealed containers was accomplished using the accelerated crucible rotation technique [12.43].

In 1986, *Brice* [12.5] estimated the annual production rates of crystals and Table 12.1 reproduces that data. Semiconductor materials clearly dominated at that stage, in particular silicon and III–V compounds. Later, *Scheel* [12.6] gave estimates of 5000 t in 1979, 11 000 t in 1986 (from [12.5]) and approximately 20 000 t in 1999 (Fig. 12.1), with a similar total quoted by *Scheel* in 2008 [12.8]. By 1999, the balance had shifted somewhat from the earlier estimates, but semiconductors continued to dominate, at ≈ 60%. There were roughly equal percentages of scintillator, optical, and acousto-optical crystals, at around 10–12%. The remainder was made up of laser and nonlinear optical crystals and crystals for jewelry and the watch industries.

A recent book edited by *Capper* [12.44] gives a comprehensive update on the bulk growth of many of the materials used in the electronic, optical and optoelectronic fields.

12.3 Techniques

12.3.1 Verneuil

This is the fastest growth method and was the first to be capable of controlling nucleation and thus producing large crystals of high melting point oxide crystals, for example, sapphire and ruby. Currently, a large number of high melting point materials have been grown by this technique, including ZrO_2 (2700 °C), SrO (2420 °C), Y_2O_3 (2420 °C), etc., but the largest use is still Al_2O_3 (often doped to produce ruby, sapphire, etc.). Figure 12.2 shows a schematic of the equipment used. Many different heat sources have been used, for example, solar furnace, glow discharge, plasma torch, arc image and radio frequency heating, but the original gas-flame technique is still the most popular. An oxy-hydrogen flame heats the seed crystal placed on a ceramic pedestal. Powder from a hopper is shaken through the flame and melts, forming a melt surface on the seed. During growth the seed is lowered, controlling the linear growth rate, while the volume growth rate is governed by the powder feed rate. The balance of these two rates controls the crystal diameter and the crystal is normally rotated slowly. Ruby crystals up to 200 mm diameter can be grown. Drawbacks of the technique include high dislocation densities and concentration variations.

12.3.2 Czochralski

Czochralski is a fast growth method and is widely used for both semiconductors and oxide/fluoride materials for optical applications. It also normally produces the most nearly perfect and homogeneous crystals. However, it is only really applicable to those materials that melt congruently or nearly congruently, that is, the solid and the melt compositions are the same at equilibrium. Normally a crucible material is needed that is compatible with the melt, however, the most common crucible material used for silicon, that is, silica, dissolves slowly in the melt and this raises certain process issues in the growth of this material. Capital costs for Czochralski pulling are higher than most other techniques, but it is used when the greatest perfection of grown material is required.

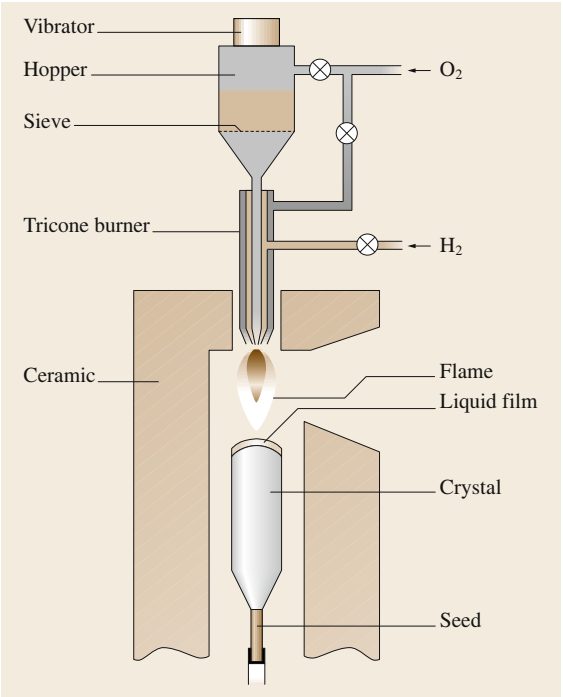


Fig. 12.2 Schematic of Verneuil growth equipment. (After [12.5])

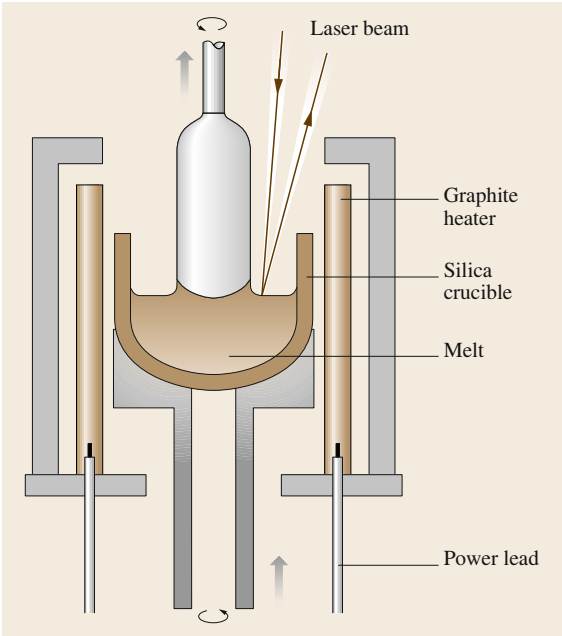


Fig. 12.3 Schematic of Czochralski growth equipment. (After [12.5])

The largest use is clearly for silicon, but probably ≈ 100 materials are commercially grown using this method, while many more are at the research stage. As the diameter of the crystal increases, in silicon for example from 200 to 300 mm, dislocation control becomes more difficult due to larger radial temperature gradients. Figure 12.3 shows a schematic of a Czochralski growth equipment. The basic method is relatively simple: The solid charge is placed in the crucible and heated to a temperature several degrees above the melting point. The seed crystal, rotating slowly, is lowered to contact the melt and the seed then slowly melts. After a short delay (a few minutes) pulling is commenced and new material (with slightly reduced diameter) begins to grow. A long narrow *neck* region is grown to reduce dislocations, then the melt temperature is reduced to increase the diameter. When the crystal attains full diameter growth is maintained until the desired length of crystal has grown, and growth is terminated by sharply increasing the pull rate or increasing the melt temperature so that the diameter reduces to zero.

A constant diameter is maintained by adjusting the power input to the melt. This is done automatically either by directly monitoring the diameter optically, that is, by observing the bright ring around the crystal periphery, or by indirectly measuring the diameter via weighing methods. Crucibles are normally round based for semiconductors but should be relatively flat for oxide/halide growth, that is, ionic materials.

Hurle [12.45] discusses the effects, in detail, of convection, flows in melts, heat transport, mass transport, solute segregation, use of magnetic fields, systems dynamics and automatic diameter control, morphological stability, and defect control. In general, as melt sizes have increased natural convective flows have increased, leading to turbulence, which causes growth rate fluctuations leading to dopant concentration variations (so-called striations).

For growth of III–V compounds, liquid encapsulated Czochralski (LEC) was pioneered by *Mullin* [12.46] for GaAs and GaP. Suppression of the volatility of As and P was crucial to the successful growth of these types of compounds. This technique involves the use of an inert layer of a transparent liquid, usually B_2O_3, which floats on the melt surface and acts as a liquid seal. Most importantly, the encapsulant should wet both the crucible and crystal so that a thin film adheres to the crystal as growth proceeds. The latter prevents dissociation of the hot crystal. Pyrolytic boron nitride (PBN) crucibles are now often used and either resistance or r.f. heating is employed. Semi-insulating GaAs is produced in a high-pressure puller (100–200 atm). The main advantage here is that elemental Ga and As can be used as the starting materials. Diameter is often control by crystal-weight measurements, rather than meniscus observation. Pull rates are typically < 1 cm/h. Axial or transverse magnetic fields can again be used to control melt turbulence and dopant segregation.

Precious metal (Pt, Ir) crucibles are used for oxide crystal growth as high temperatures are used. To prevent reduction of oxide melts, a partial pressure of oxygen is needed. Normally, the crucible is not rotated or translated and the growth rate is greater than the pulling rate as the melt height decreases. Growth rates are a few mm/h, at best, so growth times are normally > 1 week. Large temperature gradients can lead to melt turbulence, which in turn produces a banded structure of solute concentration and varying stoichiometry in the crystals. Control of convection in these systems is therefore of critical importance. *Rudolph* and *Kakimoto* [12.47] have outlined various means of controlling uniformity and stability of crystal growth from the melt by complementing internal parameters (e.g., temperature field, pressure, growth velocity, etc.) by external influences, for example, mechanical, electrical, and mechanical fields. Stirring during crystal growth can often be essential to enhance solute transport through the growth fluid, to homogenize the solute concentration in the fluid and solid, and to reduce the thickness of the solute diffusion boundary layer (in order to prevent constitutional supercooling and hence the morphological instability of the crystal). The question that obviously arises then is why are the properties of the crystal not

markedly affected by the periodic disturbances applied to growth fluids? The answer lies in a comparison of the time scales involved. Movement of solute species in melt growth is normally of the order of hundreds of cm/s. With growth rates in the region of 10^{-2} cm/s, and induced fluctuations in the region of cm/s solute movement dominates. In addition, if the characteristic ratio of D/δ^2 (with D the diffusion coefficient and δ the boundary layer thickness) is less than the frequency of any hydrodynamic changes and their related temperature oscillations then mass diffusivity cannot follow the thermal changes.

12.3.3 Kyropoulos

This simple technique is used where a large diameter is more important than length, for example, for windows, prisms, lenses and other optical components, and for scintillator materials. The set-up and method of growth are similar to those of Czochralski (Fig. 12.4) but after the seed is brought into contact with the melt it is not raised much during growth. As in Czochralski, a short necked region is still grown but then the seed removal is stopped and growth proceeds by reducing the input power to the melt. The resulting crystals normally have diameters of ≈ 80–90% of the crucible diameter. Although the control systems in these equipments are relatively simple, the temperature distribution over the crucible is critical. For alkali halides, the crystal density is greater than the melt density so the melt level decreases with growth and the desired temperature distribution is one of the increasing temperature as the base of the crucible is approached. Average linear growth rates are a few mm/h, corresponding to cooling rates of $< 1\,^{\circ}$C/h. The only other process parameter of concern is the seed rotation rate, which is normally low or zero.

While the method appears to be economically very attractive there are technical deficiencies. The isothermal surfaces are curved, resulting in high dislocation densities and the growth interface is composed of different crystal faces, with consequent inhomogeneities in impurities and vacancy concentrations. High levels of impurities result from the majority of the not well-stirred melt being consumed. Despite these drawbacks many tons of alkali halides for optical applications are grown each year by this technique.

12.3.4 Stepanov

In this technique, a crystal is pulled from a crucible containing a crystal-shaped aperture (Fig. 12.5). Crystals can be pulled vertically upward, downward, or even horizontally. Growth rates are below those of the normal Czochralski technique, but dislocation densities can be reasonably low.

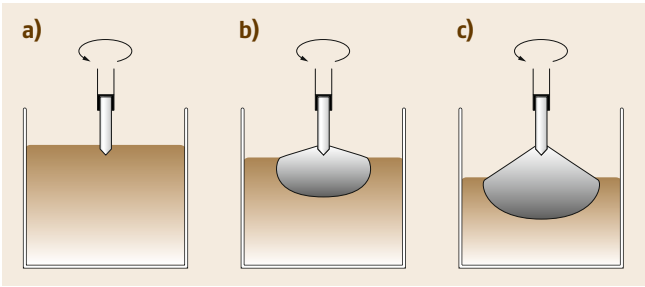


Fig. 12.4a–c Schematic of Kyropoulos growth equipment. **(a)** The seed crystal contacts the melt, a small amount melts, and then cooling is commenced to produce **(b)** and **(c)**. (After [12.5])

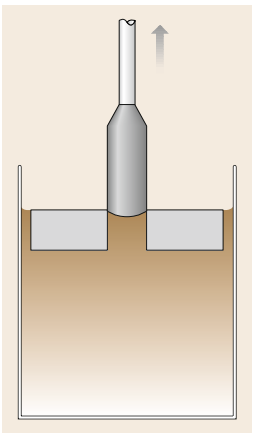


Fig. 12.5 Schematic of Stepanov growth equipment, in which a crystal is pulled through an aperture that defines its shape. (After [12.5])

12.3.5 Edge-Defined Film Growth

A die with a central capillary is placed on the surface of the melt (Fig. 12.6). Surface tension forces cause the melt to wet the die and be drawn up the capillary. A pointed seed crystal is lowered to contact this melt and then pulled upward. The melt is cooled slightly to increase the crystal diameter until it reaches the size of the die. Die can be designed to produce various shapes of crystals, for example, tubes, sheets, etc. Rapid growth is possible, but crystal quality normally suffers. One product is alumina tape (1 mm thick by several cm wide) used as a substrate for the production of high-frequency circuits. In this technique, however, purity can be limited.

12.3.6 Bridgman

In essence, this is a method of producing a crystal from a melt by progressively freezing it from one end to the other (Fig. 12.7). Crystals can be obtained with good

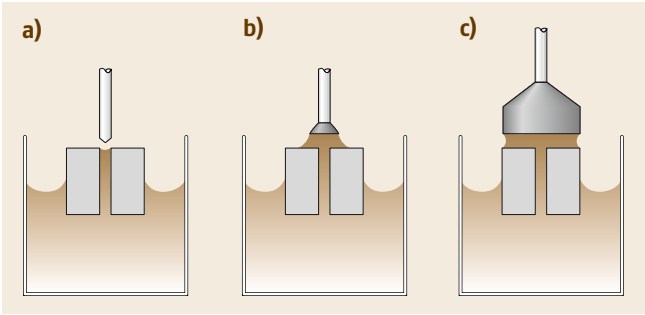


Fig. 12.6a–c Schematic of edge-defined film growth equipment. **(a)** The melt wets the die and is drawn up the capillary; **(b)** the seed is contacted to the melt and pulled up and cooling starts to increase the crystal size; **(c)** the crystal reaches the size of the die. (After [12.5])

dimensional control and the method uses relatively simple technology requiring little supervision. However, as the diameter increases, controlling heat flow becomes progressively more difficult. A wide range of materials have been produced by this technique, including sapphire at a melting point of $\approx 2370\,^{\circ}\mathrm{C}$. One major requirement is that neither the melt nor its vapor must attack the crucible material significantly. Dislocation densities can also be limited to $> 10^4\,\mathrm{cm}^{-2}$, and many materials contain low-angle grain boundaries.

Growth rates can be in the range $0.1{-}30\,\mathrm{mm\,h}^{-1}$. Either the crucible or the furnace can be moved to achieve movement of the freezing isotherm, and both vertical and horizontal orientations are used. Crucible materials include silica (for covalently bonded crystals), graphite, and some metals (e.g., Mo for sapphire growth). Traditionally, crucibles have tapered tips to try to restrict nucleation to one crystal, although seed crystals can also be used. If spurious nucleation occurs it usually forms at the crucible walls, but this can be suppressed by making the growth face concave into the melt. Baffles are often used to separate upper and lower parts of the furnace to ensure thermal isolation. A low radial temperature gradient is needed to reduce dislocation densities. Calculating the temperature distribution is possible [12.48]. The most obvious requirements are a large temperature gradient at the growth interface and low-temperature gradients in the radial direction.

Impurity distributions in crystals are, in general, governed by the type of mixing in the melt. In stirred melts, so-called normal freezing, the relevant equation is

$$C_\mathrm{s} = k_\mathrm{eff} C_0 (1-x) k_\mathrm{eff} - 1\,,$$

where C_s is the concentration in the solid, C_0 the initial concentration, k_eff is the effective distribution coefficient, and x is the fraction solidified. In unstirred melts,

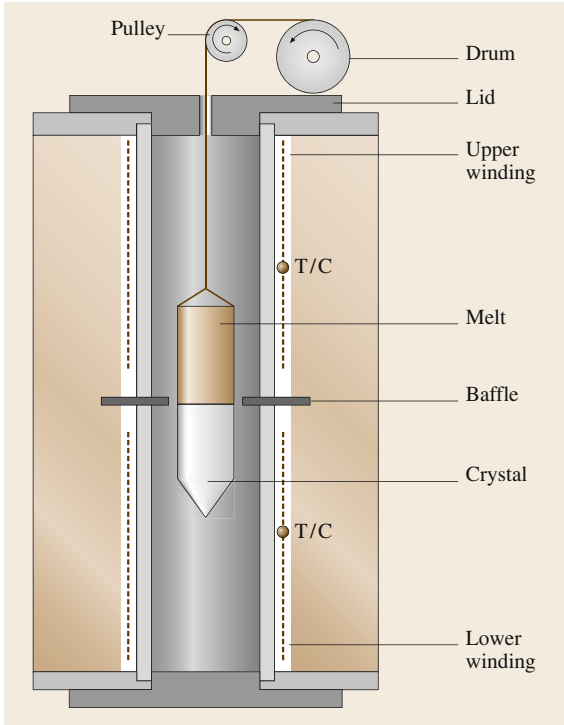


Fig. 12.7 Schematic of Bridgman growth equipment. (After [12.5])

the interface segregation coefficient k^* governs the segregation behavior

$$C_\mathrm{s} = k^* C_0 \left[1 - \frac{(1-k^*)}{k} \times \left(1 - \exp\left[-\frac{(1-k^*)k^* f_z}{D_\mathrm{L}}\right]\right)\right],$$

where f is the growth rate, z the axial distance, and D_L is the diffusion coefficient. For a molten zone of length L, the equation is

$$C_\mathrm{s} = C_0[1 + (k_\mathrm{eff} - 1)\exp(-k_\mathrm{eff}\, x)]\,.$$

For materials with a volatile component, there are several possible refinements of the basic process, for example, sealed ampoules, overpressure, and liquid encapsulation (e.g., using B_2O_3). Normally in sealed-ampoule growth there is a lack of control over stirring. For vertical systems, the accelerated crucible rotation technique (ACRT) can be used, and this author has developed this refinement for the growth of mercury cadmium telluride (MCT) up to 70 mm in diameter [12.49]. Faster stable growth rates are possible with ACRT [12.50], as are larger diameter crystals, improved uniformity and better crystallinity.

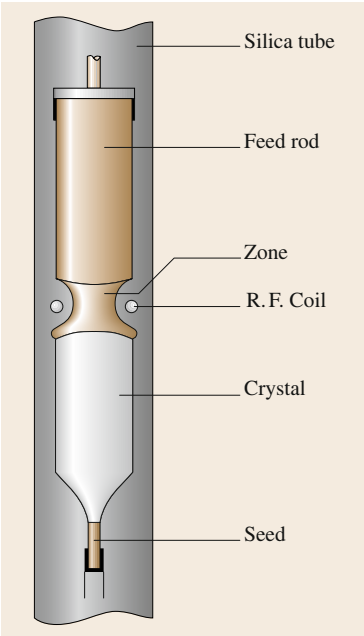


Fig. 12.8 Schematic of float-zone growth equipment. (After [12.5])

Zharikov [12.9, Chap. 3] discussed how axial vibrational control can be used to control the heat and mass transfer and growth kinetics during growth from liquids. Low-frequency, small-amplitude vibrations allow growth of high-quality crystals at high growth rates, and also provide control of stoichiometry and dopant distribution. This technique suppresses temperature oscillations at the interface and improves radial and axial distributions of impurities. It can also be used for non-electroconducting melts and liquids with high Prandtl numbers. Applications range from semiconductors to dielectrics produced from Bridgman, TSSG, Czochralski, and float-zone techniques.

12.3.7 Vertical Gradient Freeze

The vertical gradient freeze (VGF) technique involves the progressive freezing of the lower end of a melt upward. This freezing process can be controlled by moving the furnace past the melt or, preferable, by moving the temperature gradient in a furnace with several independently controlled zones. Low-temperature gradients are normally obtained, leading to reduced dislocation densities, and the crystal is of a defined shape and size. Difficulties include furnace design, the choice of boat material, and the issue of seeding.

The majority of melt-grown crystals are produced by the Bridgman and Czochralski techniques (and their variants). In general, if no suitable crucible can be found then other methods are required, for example, Verneuil.

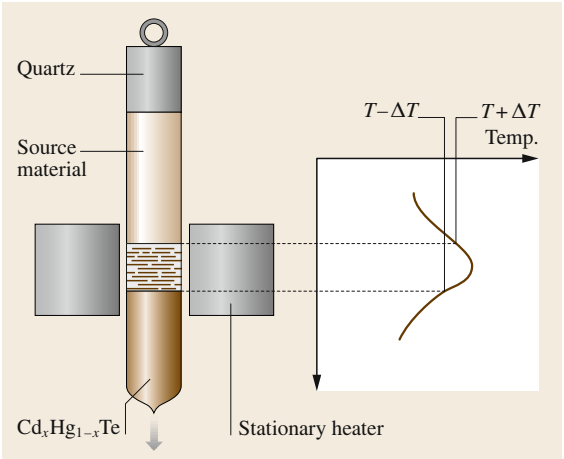


Fig. 12.9 Schematic of THM growth equipment. (After [12.5])

12.3.8 Float Zone

In this technique, a molten zone is maintained between two solid rods (Fig. 12.8). By moving the zone relative to the rods, one of them grows, and a single crystal can be grown if a seed is used. Silicon is the only material grown on a large scale by this technique. The only other use for it is in the small-scale growth of very pure crystals, as no crucible contact is involved. For silicon, r.f. heating is used, with frequencies of 2–3 MHz for diameters > 70 mm. There is a steep temperature gradient that induces flows in the molten zone, and if both the seed and feed rods are rotated then the shape of the solid/liquid interface can be controlled. The molten zone normally moves upward.

12.3.9 Traveling Heater Method

The traveling heater method (THM) technique was developed for II–VI alloy growth (Te-based) by *Triboulet* [12.51]. In this technique, a molten zone is made to migrate through a homogeneous solid source material. This is normally accomplished by slowly moving the ampoule relative to the heater (Fig. 12.9). The key requirement here is to obtain the appropriate temperature profile, also shown in Fig. 12.9. Matter transport is by convection and diffusion across the solvent zone under the influence of the temperature gradient resulting from the movement. For alloy growth, a steady state can be reached where the solvent dissolves a solid of composition C_0 at the upper growth interface and deposits, at near-equilibrium conditions, a material of the same composition at the lower growth interface. Growth occurs at a constant temperature below the solidus temperature and hence shows all the advantages

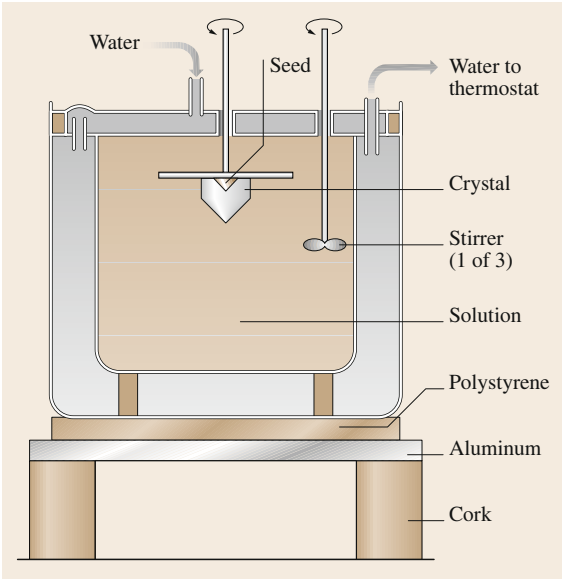


Fig. 12.10 Schematic of low-temperature solution growth equipment. (After [12.5])

of low-temperature growth. When tellurium was used as a molten zone there was a marked purification effect. Seeding was also found to be possible with the technique, as was growth of larger diameter material than with the other common melt techniques used for these compounds.

Problems with THM include the availability of suitable feed material with the required composition and dimensions, although several different routes have been used to overcome this problem. Natural convection is the dominant mechanism of material transport, which led to the addition of ACRT to the basic THM method for some of the ternary alloys.

By producing two bevelled cylinders of binary compounds graded-composition alloys were also produced. These were then used to assess the effects of composition on various electrical and optical properties in the given system. Other modifications to the basic process included the *cold THM* process, in which the relevant metallic elements were used as the source material. This produced a process that accomplished synthesis, growth, and purification at low temperature all in a single run. To avoid the problems resulting from solvent excess in the crystal *sublimation THM* was developed in which an empty space of the same dimension as the molten zone is used. This method was successfully applied to the growth of ZnSe. To improve the purification effect, repeated runs on the same material were used. A drawback of this is one of repeated handling, which in turn was solved by using the *multipass THM* technique, which can be thought of as a variety of the classical zone-melting method.

12.3.10 Low-Temperature Solution Growth

Most of the crystals grown by this technique are water soluble. This limits their use to applications in which moisture can be excluded. Growth rates are low (0.1−10 mm/d), as the growth faces are unstable. This is due to the concentration gradient near the growth face, in addition to which the crystal is normally totally immersed in the solution so that latent heat evolution increases the adverse supersaturation gradient (Fig. 12.10).

Water (both light and heavy) is used in ≈ 95% of the cases, and must be highly pure. All equipment must be carefully cleaned and protected from dust and the solutions must be stirred vigorously while being prepared. Both slow cooling and solvent evaporation techniques are used. Large volumes of solution (> 100 l) are needed for very large crystals of, e.g., KH_2PO_4 (KDP). These large volumes imply very long thermal time constants, which can be increased by adding thermal insulation, up to a limit. Large time constants make it easy to obtain good temperature control. The growth equipment is normally held in temperature-controlled rooms. Although temperature stability is key, other parameters, such as fluctuations in stirring, can lead to large changes in growth rate too.

Solvent evaporation simplifies the temperature control system and makes it more reliable. However, it is difficult to ensure a constant acceptable rate of loss of solvent. This can be approached by using a water-cooled condensation region. One drawback of this technique is that concentrations of impurities increase as the growth proceeds.

12.3.11 High-Temperature Solution Growth (Flux)

This method is used for those materials that melt incongruently; the solvent (flux) reduces the freezing point below the relevant temperature to produce the desired phase of the compound. Both liquid metals (Ga, In, Sn) used in semiconductors and oxides/halides (e.g., PbO, PbF_2) used for ionic materials are employed as solvents. Often, an excess of one of the components will be used (e.g., Ga for GaAs and GaP for instance) or a common ionic material (e.g., K_2CO_3 for $KTa_xNb_{1-x}O_3$). Alternatively, the solvents contain large atoms, such as Pb and Bi, which are too large to enter the lattice of the desired crystal. Mixed solvents, for example, alkali metal ions, break the chains that exist in B_2O_3, SiO_2, and

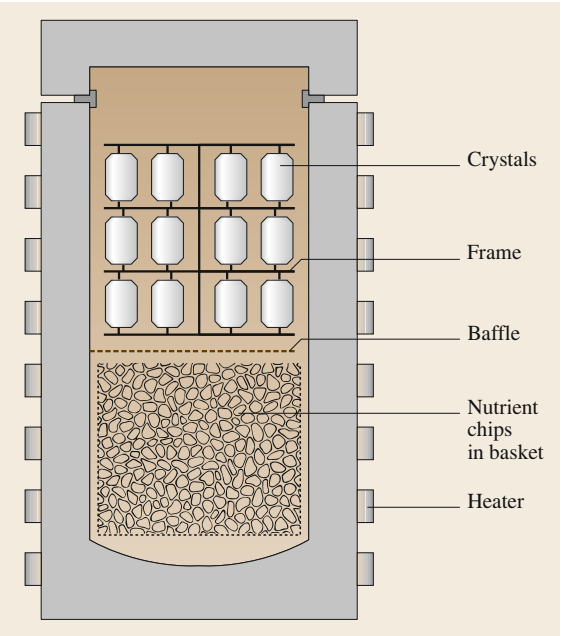


Fig. 12.11 Schematic of hydrothermal growth equipment. (After [12.5])

Bi_2O_3, lowering viscosities and surface tensions. The book by *Elwell* and *Scheel* [12.4] contains much useful information in this field.

Slow cooling of high-temperature solutions was used between 1950 and 1970 to produce hundreds of different materials. Small crystals were normally obtained but these were sufficient to obtain useful measurements of magnetic, optical, and dielectric properties. Later, seeded growth and stirring using ACRT led to much larger crystals. A Czochralski-type pulling technique can also be used in high-temperature solution growth to produce larger crystals.

12.3.12 Hydrothermal

This is growth from aqueous solution at high temperature and pressure. Most materials grown have low solubilities in pure water, so other materials (called mineralizers, for example, NaOH and Na_2CO_3) are added to increase the solubility. Quartz is the only material grown on a significant scale by this method, although ZnO is also produced commercially by this technique.

Natural quartz crystals are used as seeds, as they have lower dislocation densities. Growth proceeds in the temperature-gradient transfer mode. Nutrient is held in the lower part of the liquid in an autoclave. This region is held at between 5 and 50 °C above the upper portion of the autoclave. Convection carries the hot supersaturated liquid to the cooler upper regions, where deposition occurs on the seed crystals (Fig. 12.11). These autoclaves can be operated at up to 400 °C and 2000 atm pressure for ≈ 10 yr without degradation. Autoclaves can be up to 1 m in diameter and 2.5 m high for low-pressure growth from Na_2CO_3 solution, while 40 cm diameter and 8 m high autoclaves are typical for growth from NaOH.

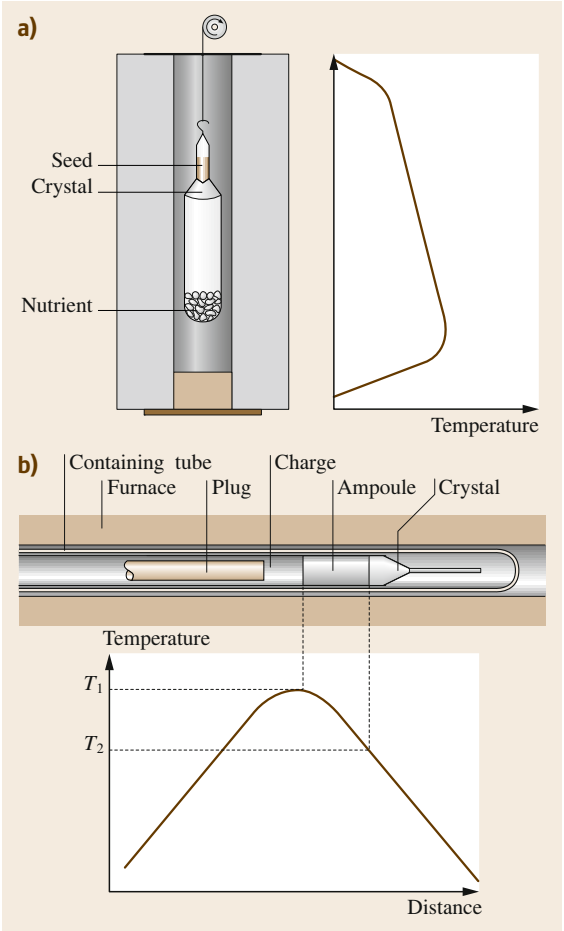


Fig. 12.12a,b Schematic of vapor growth equipment. **(a)** Seeded growth; **(b)** Piper–Polich set-up. (After [12.5])

Resistive heating elements are strapped to the outside of the autoclave and connected to two control systems, allowing independent control of the temperatures in the solution and growth zones. *Brice* [12.5] presents a table that shows some typical hydrothermal growth conditions for a wide range of materials.

The essential problem in the growth of quartz is to produce useful devices with reasonable yields with OH concentration (measured by IR absorption) of < 100 ppma. Several parameters restrict the maximum growth rate to ≈ 0.5 mm/d.

12.3.13 Growth from the Vapor

Vapor growth has some applications, particularly in high melting point materials and in II–VI compounds such as ZnSe and ZnS (Fig. 12.12). High-temperature vapor transport, or sublimation, of SiC has progressed recently to producing 1–2 in crystals. Growth proceeds by sublimation of a SiC source at 2000–2500 °C in near-vacuum conditions onto a seed held at ≈ 1000 °C lower than the source. 6H-polytype material oriented at ⟨0001⟩ is produced. Growth rates are ≈ 2 mm/h. Problems remaining include impurities (≈ 10^{16} cm^{-3}), voids, high dislocation densities, and mixed polymorphism.

12.3.14 Multicrystalline Si Growth for Solar Cells

Green [12.52] discusses the production of multicrystalline Si slices for photovoltaic applications and notes that in 2011 approximately 48% of the world production was based on this material. Advantages over the Czochralski process are noted as lower capital costs, higher throughput, and a higher tolerance to poor feedstock quality. Basically, the process involves the controlled solidification of molten silicon in a container to give silicon ingots with large columnar grains, normally propagating from the container base (Fig. 12.13a). These rectilinear ingots can be very large, up to 400–600 kg, and they are sawn into smaller more manageable sections, ≈ 150 mm in size (Fig. 12.13b), before being sliced into wafers by continuous wire sawing. This material is capable of producing solar cells with ≈ 90% of the performance of single-crystal Si, but due to the higher packing density possible due to the square geometry, this difference is masked at the module level where the overall performance is equivalent.

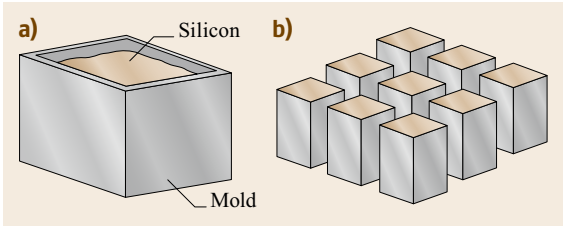


Fig. 12.13 (a) Direction solidification of silicon within a mold: (b) smaller sections are sawn from the main ingot. (After [12.52])

12.4 Materials Grown

This section summarizes the position regarding the wide range of materials used in electronics and optoelectronics. Again the reader is referred to the following relevant chapters for more in-depth coverage of a particular material. The section is subdivided roughly into group IV, group III–V, group II–VI, oxides, halides, and finally phosphates/borates.

12.4.1 Group IV

Silicon (Germanium)

The growth of silicon still dominates the electronics industry in terms of size of activity. *Hurle* [12.45] describes in detail the growth of silicon by the Czochralski method. In essence, a polysilicon charge is placed in a high-purity quartz crucible mounted in pre-baked graphite supports. Chains or cables are used in the pulling system, rather than rods, to limit the height of modern pullers. Both the crystal and the crucible can be rotated, although the details of this aspect vary from manufacturer to manufacturer and are proprietary. As the growth proceeds, the crucible is raised to maintain the position of the melt surface in the heater.

Silicon has two major advantages from a crystal growth perspective: a high thermal conductivity (which permits removal of latent heat of solidification) and a high critical resolved shear stress (which allows high thermal gradients without dislocation generation). These two factors enable high growth rates, leading to economic benefits. Heat is provided by *picket-fence*-type graphite heaters. Most commercial pullers are highly automated with full computer control. In most cases, diameter control is achieved via monitoring of the *bright ring* around the meniscus of the growing crystal. Oxygen is a particular problem and is introduced via erosion of the silica crucible by the melt. Convection moves the oxygen within the melt toward the crystal, where it is incorporated. A similar situation arises with carbon monoxide loss from the melt. Problems with convection have increased as the diameters of silicon crystals have increased over the past several decades. This convection controls concentrations and distributions of oxygen, carbon, and other unwanted impurities.

Although current production of large-diameter material is at the 300 mm level, there have been reports of 400 mm diameter material [12.53] and even 450 mm (≈ 450 kg) [12.8, Chap. 6]. In this case, it was necessary to grow a secondary neck below the Dash dislocation-reducing one as the thin Dash neck could not hold the

weight of the crystal. One report [12.54] shows how co-doping Si with B and Ge can obviate the need for this narrow-neck region, while maintaining dislocation-free growth. *Rudolph* and *Kakimoto* [12.47] report how the use of a 0.35 T magnetic field can reduce turbulence in such large melts.

The highest grade of silicon i. e., the highest purity, is probably still produced by the float-zone (FZ) method, rather than the Czochralski technique [12.55, 56]. The higher purity (particularly regarding carbon and oxygen) and better microdefect control result in higher solar cell efficiencies in FZ material. Faster growth rates and heating/cooling times together with the absence of crucibles and hot-zone consumables also give FZ material an economic advantage over Czochralski material.

Table 12.2 summarizes the present position on sizes of various crystals, including silicon and germanium and silicon/germanium. The author does not claim that this list is complete nor is it authoritative, it merely serves to show the range of crystals and sizes currently being grown by bulk growth techniques. The information is taken from a number of published conference proceedings [12.57–66] and other sources [12.8, 9, 44, 49, 67–69].

Silicon Carbide

The growth of SiC for semiconductor devices, as opposed to abrasive applications, became successful when substrates could be reproducibly grown up to large sizes [12.70–72]. Currently, substrates up to 50 mm diameter are commercially available and 100 mm diameter substrates have been reported by *Talalaev* et al. [12.9, Chap. 2]. The current method is that of seeded sublimation growth. A water-cooled quartz reactor enclosure surrounds a graphite crucible, which is heated by r.f. means. Crucible sizes are slightly larger than the required crystal size and SiC powder (particle size 20–200 μm) and/or sintered polysilicon is used as the source. Typically a distance of 1–20 mm between the seed and the source is used. At the high process temperatures of 1800–2400 °C, volatile species of Si_2C, SiC_2, and Si evaporate from the source and deposit on the seed. The source temperature determines the rate of evaporation, and the temperature difference between the source and the seed determines the diffusion transport rate; together these two temperatures govern the growth rate. The Si : C ratio is a key control parameter for proper growth of the desired material. Sublimation is normally performed in vacuum or inert gas, such as argon. The pressure of this gas also controls the growth rate to an extent. Growth rates can reach up to 0.5–1.0 mm/h, and crystal lengths of up to 20–40 mm are achieved. Doping can be accomplished via nitrogen and aluminum addition for n- and p-type material, respectively.

Polytype control is obtained by carefully choosing the seed orientation, with growth on the (0001) Si face giving 6H material, while 4H material grows on the $(000\bar{1})$ face. Both growth temperature and pressure and intentional/unintentional impurities can also affect the polytype obtained. One remaining problem is that of micropipes within the material. These are hollow core defects of 1–10 μm diameter, although their density has been reduced from 1000 to 1 cm^{-2} in the best samples. Maintaining good control over the nucleation conditions can reduce micropipe density.

Diamond

High-pressure diamond is produced under conditions where it is thermodynamically more stable than graphite [12.73]. However, very high temperatures and pressures are required for growth, unless catalysts are also employed. Clearly a detailed knowledge of the diamond–graphite pressure–temperature phase diagram is a prerequisite for successful growth. For example, in static compression, 8–20 GPa of pressure and temperatures of 1000–3000 °C are needed. Carbon sources include graphite, amorphous carbon, glassy carbon, and C_{60}. Dynamic compression techniques employ pressures of 7–150 GPa. A wide variety of catalysts have been used ranging from conventional ones, for example, transition metals, to carbide-forming elements (Ti, Zr, etc.), to Mg, to oxygen-containing materials (carbonates, hydroxides, etc.), to inert elements (P, Cu, Zn, etc.) to hydrides.

Crystals above 1 mm size are grown by the temperature gradient method. Growth temperature, temperature gradient, growth time, type of catalyst, and impurities all affect the growth. Crystals up to 10–12 mm in size are commercially available from high-temperature–high-pressure growth, see *Sumiya* [12.74]. Growth rates can reach 10 mg/h and are governed by the growth temperature and the temperature gradient. If a high-quality diamond seed is used crystals free of major defects, such as inclusions, stacking faults and dislocations, can be obtained. Impurities, mainly form the catalysts used, are still an issue, as is their nonuniform distribution.

More recent work on CVD techniques has led to diamond crystals up to 100 mm in diameter and 1–2 mm thickness [12.75]. This book also gives details of the use of diamond in a very wide range of device applications, including radiation/neutron detectors, electronic devices, for example, field effect transistors (FETs), electrochemical, and biological sensors and microelectromechanical devices. It even describes the area of superconductivity in diamond films.

Table 12.2 Typical sizes (i. e., diameter) of a selection of bulk-grown crystals (in mm and/or weight in kg) currently in production and in R&D

Crystal	Commercial	R&D
Silicon (Cz)	300 (250 kg)	450 (450 kg)
Silicon (FZ)	100	
Silicon (multicrystalline)	680 × 680 mm (400–600 kg)	
Germanium	75	200–300
Silicon/Germanium		35 (120 mm long)
SiC	50	100
Diamond	100 (1–2 mm thick)	100
GaAs	200 (50 kg)	200
InP	150	150
GaSb	50	75
GaInSb	50	
InSb	75	100
InAs	50	
InGaAs		10
GaInAsSb		30
GaN (sublimation)	50	
AlN (physical vapor transport, PVT)	25 (5 mm thick)	40 (core in 50 mm slices)
MCT	50	70
CdZnTe	140 (10 kg)	140 (12 kg)
Bi_2Te_3	12	28
ZnSe (vapor, LEC)	60	
ZnTe	80	
ZnO	50	75 (10 mm thick)
HgMnTe	30	
CdMnTe		40
HgZnTe	30	
$LiNbO_3$		40
$YCa_4O(BO_3)_3$	80	
$PbZnNbTiO_3$		75
YAG	20	75
GGG	110 (25 kg)	
$LiTaO_3$	100	125
Ruby	200	
Sapphire	350 × 500 × 40	500
BGO	140	
$La_3Ga_5SiO_{14}$	100	
Quartz	15–30 × 150–250, (0.5–18 kg)	
β-BaB_2O_4		60
CaF_2	385 (100 kg)	
Alkali halides	520 (550 kg)	1000 (poly)
BaBrI	400	
Ce:$Li_6Lu(BO_3)_3$	30	
KDP	660 × 500 (380 kg)	
$Ca_5(PO_4)_3F$	70	
TGS	40	

12.4.2 Groups III–V

Gallium Arsenide

Mullin [12.46] provided a comprehensive discussion on all aspects of the growth of III–V and II–VI compounds. He starts his treatment with the early work done on purification in Ge and Si, which highlighted the need for both high purity and for single-crystal growth. He notes that *Pfann* [12.76] initiated zone melting, in which, in the simplest case, a horizontal molten charge is progressively frozen from one end. Both zone leveling, in which a liquid zone within a solid bar is moved through

the bar in one direction and then in the reverse direction, thus producing a uniform distribution of a dopant, and zone melting, in which liquid zones are repeatedly passed through a solid bar to force impurities to segregate to the ends of the bar and then these are subsequently removed before the bar, are used in growth. Horizontal techniques were also used to produce single-crystal material by slight back-melting of a seed crystal, followed by progressive freezing. These various horizontal techniques were subsequently applied to the purification of both elements and compounds in the materials from groups III–V and II–VI.

Table 12.3, taken from [12.46], summarizes the growth techniques that have been used for the various compounds in the groups III–V and II–VI. There has been much debate about the relative advantages and disadvantages of these techniques as applied to compound crystal growth.

GaAs is the most important group III–V compound. Horizontal growth can be used to produce low-temperature gradients at the solid/liquid interface, which reduces stress-induced slip and hence dislocations. This is particularly beneficial for the use of GaAs in laser diodes, where dislocation densities are required to be very low. However, there is a potential problem in using these low-temperature gradients and that is constitutional supercooling, which may result in unstable growth and second-phase inclusions. In addition, there is the issue of impurity uptake from the crucible and nucleation of twins, grain boundaries, and even polycrystalline growth from the crucible walls. These problems can be minimized by taking care in experimental detail but they still make horizontal growth unattractive where large-area uniformity is required, particularly as scaling up to ever larger diameters is difficult.

Early efforts to produce this material by the vertical Czochralski method failed due to dissociation of the melt. This problem was solved by the LEC technique [12.46]. A layer of boric oxide floats on the melt and the seed crystal is dipped through this to contact the melt (Fig. 12.14). PBN crucibles are normally used and heating can be by resistance or r.f. means but commercial systems use graphite heaters. Most semi-insulating GaAs, which is required for integrated circuits, is produced at pressures of 100–200 atm. The advantage of this modification is that elemental As and Ga can be used as the source. Diameter control is normally via crystal weighing. Currently, LEC produces material of 150 mm diameter [12.77] from 28–40 kg melts using 400 mm diameter PBN crucibles. Also reported were modifications to the basic Czochralski technique to reduce stresses, and hence dislocations, in crystals up to 100–150 mm diameter. These are vapor-pressure-controlled Czochralski (VCZ), fully encapsulated Czochralski and hot-wall Czochralski. The reader is referred to the reference for more details of these techniques. Friedrich et al. [12.8, Chap. 5] note that 200 mm diameter crystals have also been grown.

An alternative to Czochralski is that of vertical gradient freeze. For reproducible growth of large-diameter crystals by this technique it was found necessary to use B_2O_3 in a PBN crucible. Freezing is accomplished by moving the temperature gradient via furnace-controller changes, rather than movement of the furnace itself. This naturally produces low-temperature gradients, which give low dislocation densities, and it pro-

Table 12.3 General applicability of growth techniques to compound semiconductors. (The more *asterisks* there are the more appropriate the technique; P: potentially applicable; C: conventional VGF; L: LEC VGF.) (After [12.46])

Technique/ compound	Zone melting, horizontal Bridgman	VGF, vertical Bridgman	Conventional vertical pulling	Liquid encapsulation pulling	Vapor growth
InSb	***	P	***		P
GaSb	***	P	***		P
InAs	***	P		***	P
GaAs	***	C***: L***		***	P
InP	*	C*: L**		***	P
GaP	*	C*: L**		***	P
HgSe		**			P
HgTe		***			P
CdSe	*	**			***
CdTe	***	***			***
ZnSe		**			***
ZnTe		**			***
HgS					***
ZnS					***
CdS					***

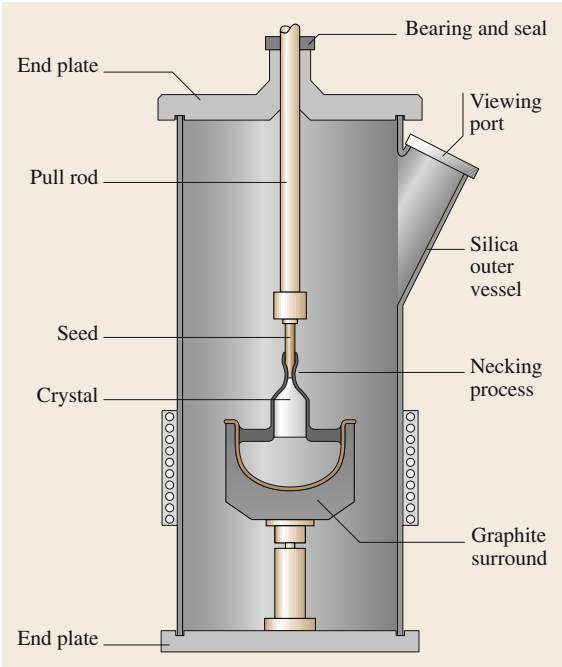


Fig. 12.14 Schematic of liquid-encapsulated Czochralski growth equipment. A silica outer vessel with viewing port is held between end-plates. The induction heating coils couple into the graphite surround. The seed is fixed in the chuck on the pull rod that rotates and moves through the bearing and seal. The crystal grows from the seed through a necking process, and on withdrawal pulls out a layer of B_2O_3 over its surface. (After [12.44], copyright Wiley-VCH (1998), used with permission)

duces a crystal of the right size and shape for subsequent slicing and processing. *Rudolph* [12.77] reports that in 1999 LEC growth of GaAs accounted for ≈ 90% of all SI GaAs, with the remaining 10% via vertical Bridgman or VGF growth. By 2000, he puts the figures at ≈ 50% LEC and ≈ 50% VGF. Rudolph also notes that VGF has been reported to grow up to 150 mm diameter material. *Rudolph* and *Kakimoto* [12.47] have shown how application of a magnetic field in VGF growth can produce nearly striation-free material.

Indium and Gallium Phosphides

As for GaAs, high-pressure LEC growth of both InP and GaP is normally used [12.46]. The problems were analogous to those of GaAs, with the major addition of an increased tendency to twinning. Loss of phosphorus (P_4) is still a problem and leads to deterioration in crystal quality. Dislocation densities are still seen to be high. *Asahi* et al. [12.78] report that some of these problems have been alleviated by modifications to the basic LEC process, for example, thermal-baffled LEC, phosphorus-vapor-pressure controlled LEC, and VCZ. Dislocations tend to increase with crystal diameter, although crystals up to 100 mm with low-disclocation densities have been grown by VCZ. Lower dislocation densities are produced by the VGF method, at a given crystal diameter. High-pressure VGF growth of InP is difficult and temperature fluctuations cause twinning. By improving the temperature control (to within ±0.03 °C). Asahi et al. show that twin-free ⟨100⟩ InP single crystals of 100 mm diameter and 80 mm length can be obtained. Temperature gradients of < 10 °C/cm reduced the etch pit densities below those of comparable LEC material. *Friedrich* et al. [12.8, Chap. 8] report that crystals of InP up to 150 mm diameter have also been produced.

Indium and Gallium Antimonides

Mullin [12.46] has detailed the early work on InSb; much less work has been done on GaSb. The low melting point combined with the negligible vapor pressure of Sb over the melt make zone-refining methods attractive. However, certain impurities render this route troublesome, although means to overcome this issue were found. Crystal-pulling methods were preferred for their versatility, although twinning was seen as a major issue. Growth on the (111) Sb face produces this twinning and it is recommended that growth takes place instead on [211] Sb or [311] Sb faces.

Micklethwaite and *Johnson* [12.79] summarized the position for InSb production for IR detector applications. InSb expands by ≈ 13% on freezing, leading to a net flow away from the growth interface. This *density effusion* must be accounted for in modeling this system. Hydrogen is used as the atmosphere, as it reduces floating InO_x. The basic Czochralski method is used for commercial production and 75 mm diameters are routine. Defect densities can be as low as 10 cm^{-2}. Many of the details of the growth process are proprietary. Growth proceeds without a liquid encapsulant, and as such resembles Si growth more closely than its arsenide and phosphide cousins. However, there are significant differences from the silicon case in terms of reduced radiative heat transfer and increased convective flow in the hydrogen atmosphere. Fluid mechanics in the melt are complicated both by the *density effusion* and the facet effect [12.46].

Dutta [12.8, Chap. 12] reports on the VB and VGF (+ACRT) growth of ternary III–V crystals, for example, GaInSb, up to 50 mm diameter and discusses the reasons why ternaries are so much more difficult to grow than binaries. Later, he described the growth of sev-

eral other III–V ternaries, for example, GaInAs, from quaternary melts by VB/VGF, including the addition of melt replenishment during growth [12.9, Chap. 9].

Group III Nitrides

There was an increase in interest in these compounds following successful development of blue laser diodes. Ideally, epitaxial thin films of these nitrides would be grown on bulk substrates of similar compounds, to minimize lattice mismatch. However, growth of these compounds by the more normal Czochralski and Bridgman techniques is not possible due to the high melting temperatures (2200–3500 K) and high-decomposition pressures (0.2–60 kbar) [12.80, 81]. One growth method that can be used is high-pressure solution growth. In this technique, liquid Ga (plus 0.2–0.5 at % Mg, Ca, Be or Zn) is held in a high-pressure chamber with a N_2 atmosphere. The maximum pressure is $\approx$ 20 kbar and the maximum temperature is 2000 K. Both pressure and temperature are controlled within tight limits and in situ annealing in vacuum is also allowed for. A temperature gradient of 2–20 °C/cm is maintained along the crucible axis and the N_2 dissolves in the hotter end and GaN crystallizes in the cooler end. No seeding was used and after 120–150 h of growth hexagonal needles or platelets are produced of size 10 × 10 mm. Growth by metal-organic vapor phase epitaxy (MOVPE) and molecular beam epitaxy (MBE) of GaN and InGaN multi-quantum wells (MQWs) on these bulk GaN samples was also reported by *Grzegory* et al. [12.81]. Light emitting diodes (LEDs) and laser diodes based on these epitaxial layers were also described.

This technique only produces very small crystals of both AlN and InN.

Nishino and *Sakai* [12.82] describe the sublimation growth of both GaN and AlN. Initially, the source powder is synthesized by heating Ga metal in NH_3. This powder is then annealed in NH_3 before loading into the growth chamber. Temperatures are 1000–1100 °C but atmospheric pressure can now be employed. Both NH_3 and N_2 are introduced into the growth chamber. Either small, 3 mm × several hundred μm, *bulk* crystals or free-standing films (10–30 μm thick) on sapphire substrates are produced. Thicker samples (up to 500 μm thick) can be produced by the sublimation sandwich method in which the distance from source to substrate is reduced to 2–5 mm. A hot-pressed polycrystalline source of AlN and a SiC-coated graphite crucible was used to grow 0.3–1.0 mm thick films on the 6H-SiC seed (10 × 10 mm in size) at temperatures of 1950–2250 °C. *Talalaev* et al. [12.9, Chap. 2] note the growth of 50 mm diameter AlN substrates with a central 40 mm diameter single-crystal core, plus totally single-crystal high-purity 25 mm AlN substrates.

12.4.3 Groups II–VI

Several reviews of the growth of a wide range of binary and ternary II–VI compounds have been published [12.46, 67, 83, 84]. One of the key concepts in the growth of these compounds is the ease with which phase transitions occur. This imposes limits on certain growth techniques for particular compounds.

There are two main types of bulk growth techniques, namely growth from the liquid and growth from the vapor. Most narrow-gap II–VI compounds are characterized by high melting points and/or high component partial pressures and early work concentrated on various forms of vapor growth, particularly for S- and Se-containing compounds. Problems of low growth rate and/or small crystal size and purity led to a switch to melt growth techniques, although there have been some significant developments in the growth of larger crystals from the vapor.

Many techniques have been used to grow narrow-gap II–VI compounds. Solid-state recrystallization (SSR) was used for ternary systems where there is a wide separation between solidus and liquidus, for example, MCT. Other names that have been used for this process are quench anneal (QA) and cast recrystallize anneal (CRA). The term anneal is used in the first case to define a high-temperature grain growth process while in CRA it is a low-temperature process to adjust stoichiometry. Strictly speaking SSR is crystal growth from the solid phase at temperatures close to the melting point but is included here for convenience.

Bridgman growth is the most widely used of the bulk growth techniques applied to narrow-gap II–VI compounds. It is the only technique that has produced crystals of all the binary and ternary compounds studied. Numerous modifications have been applied to the basic process but the three principle means of achieving growth are to move an ampoule through a temperature gradient, to move the furnace past a stationary ampoule and to move a temperature profile from high to low temperatures with both the furnace and ampoule stationary, the so-called vertical gradient freeze (VGF) method.

Usually, silica ampoules are used but various coatings have been applied and the use of other materials such as PBN has been reported [12.85]. Charges are prepared either using the appropriate elements or, for ternaries and quaternaries, pre-formed compounds of the binaries. Both vertical and horizontal configurations have been used and large crystal sizes coupled

with a relatively high growth rate (1–2 mm/h) make the basic technique relatively cheap and versatile. Marked segregation of both matrix elements, in ternaries and quaternaries, and impurities can occur. Difficulties in controlling component vapor pressures can be overcome by employing a reservoir of one of the components at a temperature lower than that of the growing crystal. High partial pressures in Hg-containing compounds can cause ampoule failure but means have been developed to cope with this situation [12.86].

The traveling heater method (THM) has mainly been used for Te-based binary and ternary compounds. Growth by THM combines the advantages of low-temperature solution growth with steady-state conditions, as in zone melting. A homogeneous alloy can be used as the starting ingot or segments of the binaries can be employed. If the Te zone height is made equal to the ring heater size then a flat interface is obtained, for particular geometries of crystal and furnace and growth parameters [12.87]. Very low growth rates (0.1 mm/h) are typical but diameters up to 40 mm have been accomplished. As in Bridgman, many variations on the basic THM method have been attempted. These include multipass-THM [12.88] and *cold* THM [12.89].

Hydrothermal growth has mainly been applied to S- and Se-based binary compounds and was reviewed by *Kuznetsov* [12.90]. The main advantages are in reducing growth temperature (to 200–250 °C) and dealing with the high partial pressures. Growth rates are very low (0.05 mm/d), and crystals are small (few mm) and contain subgrains and inclusions, but can still provide useful fundamental data.

Growth from the vapor is usually carried out at much lower temperatures than growth from the liquid, hence below unwanted phase transition temperatures, and can lead to less defective crystals. However, growth rates are generally much lower and uncontrolled nucleation and twinning are more prevalent. Techniques can be split essentially into unseeded and seeded methods. The simplest technique is that of chemical vapor transport (CVT), usually using iodine as the transport agent. The compound to be grown is reacted at a high temperature with the iodine, transported to a cooler region and deposited. The reverse reaction produces iodine, which diffuses back to the high-temperature region and the process repeats itself. The method uses a 2-zone furnace but has the disadvantages of iodine incorporation and small crystal sizes. *Mimila* and *Triboulet* [12.91] experimented with sublimation and CVT using water as a transporting agent for ZnSe and Triboulet has succeeded in the growth of CdTe by this means.

Most other vapor techniques are in some sense a derivative of the Piper–Polich method (Fig. 12.12b). Although this was developed originally for the growth of CdS [12.92], the technique may be applied to the vapor growth of any compound that sublimes readily below the melting point temperature. The crystal is grown in a closed crucible that is moved through a steep temperature gradient, such that the source material is always hotter than the growing surface and mass transport occurs from source to crystal.

Vapor growth under the controlled partial pressure of one of the constituents has been obtained by including small quantities of one of the elements in an extended part of the capsule located in a cooler part of the furnace [12.93]. The technique, termed *Durham*, is a vertical unseeded growth procedure, where the source is sublimed from bottom to top in an evacuated silica capsule over several days. Constituent partial pressures are independently controlled by placing one of the elements in a separately heated reservoir connected to the main capsule via a small orifice.

The vertical unseeded vapor growth (VUVG) method differs from the Durham technique in that the partial pressures are controlled by the initial deviation from stoichiometry of the charge material. It was initially developed to grow CdTe. The crystal is grown in an evacuated cylindrical silica capsule in a vertical furnace. A long quartz tube, which extends outside the furnace, is attached to the top of the capsule providing a heat pipe to create a cooler area for nucleation, which is closely monitored to ensure that the nucleus is single grained.

The sublimation traveling heater method (STHM) was developed from THM by *Triboulet* and *Marfaing* [12.88]. The Te-rich molten zone in conventional THM is replaced by an empty space or vapor zone. A problem arises in that the constituent partial pressures are not only functions of the temperature. However, the vapor can be maintained near the stoichiometric P_{min} condition if a small capillary, one end of which is at room temperature, is connected to the vapor chamber [12.94]. Excess species are sublimed down the capillary, preserving the relative constituent partial pressures and hence the growth rate.

In one version of the seeded technique, an oriented seed is placed on top of a long sapphire rod inside a sealed quartz capsule. The source material is placed in a basket above the seed and sublimation is top-down. The sapphire rod is carefully centered within the growth capsule to ensure a narrow gap between it and the walls of the capsule. The crystals grow clear of the container walls and generally consist of two or three large grains, some of which maintain the seed orientation. This *Markov* technique has been adapted for commercial production by the ELMA Research and Development Association of Moscow (cited in *Durose*

et al. [12.95]). Large crystals of CdTe, 50 mm in diameter and up to 10 mm in length have been grown.

HgS is grown by vapor means or by hydrothermal techniques, both of which are at low temperatures and hence low S partial pressures. HgSe has been produced by vapor growth and also by the Bridgman technique. Thick-walled silica ampoules are required in the growth of the Hg-containing compounds to contain the high Hg partial pressures. More details of the growth of the Hg-based binaries can be found in [12.89] for vapor and Bridgman growth, while Kuznetsov has reviewed the hydrothermal growth of HgS [12.90].

Early work on the growth of the Cd-based binaries was summarized by *Lorenz* [12.96]. Vapor growth was favored for CdS and CdSe, although hydrothermal growth was also discussed. The latter technique, as applied to CdS, was discussed in [12.90]. CdTe was grown by various melt growth techniques, including seeded zone melting and Bridgman, and by vapor growth methods. *Zanio* [12.97] reviewed the growth, properties and applications of CdTe, including radiation detectors. By that time, vapor growth had been essentially supplanted by both Bridgman and THM growth. A more recent review of CdTe can be found in *Capper* and *Brinkman* [12.98]. The main uses of CdTe (and CdZnTe/CdTeSe) are as a substrate material for the epitaxial growth of MCT and, more recently, as radiation detectors [12.99]. *Triboulet* et al. [12.100, 101] applied SSR to the growth of both CdSe and CdTe.

A wide variety of techniques for growth from the liquid have been used to produce crystals of CdTe. These include solvent evaporation [12.102], liquid encapsulated Czochralski [12.103], zone refining [12.104], VGF [12.105, 106], solution growth [12.107], heat exchanger method [12.108], float-zoning [12.109], Bridgman [12.98] and THM [12.51], and *Triboulet* et al. [12.100] have reviewed these materials as substrates for MCT epitaxial growth and compared them to the alternatives based on GaAs, sapphire, and Si. Problems with the growth of CdTe include low thermal conductivity, difficulty in seeding due to the need for superheated melts, ease of twin formation, tilts and rotations in the lattice, stoichiometry control, and impurities. However, *Triboulet* et al. [12.101] concluded that despite the progress made in the alternative substrate materials, lattice-matched substrates based on CdTe produced MCT epitaxial layers that gave the best device performance at that time.

Within the Bridgman growth technique, many alternatives have been attempted. These include holding excess Cd at a lower temperature in a separate reservoir [12.110], vibration-free growth [12.111], vibrational stirring [12.112], horizontal growth [12.113], low-temperature gradients [12.114], high-pressure (100 atm) growth [12.115], and the addition of the accelerated crucible rotation technique (ACRT) [12.116]. Work in the United States [12.113] aimed at improving both the horizontal and vertical techniques and scaling-up to large melts, up to 8 kg (horizontal). *Szeles* et al. [12.117] showed that crystals up to 140 mm diameter and 10 kg in weight can be grown by careful control of the temperature fields.

In the Bridgman process, elemental Cd and Te are loaded into a clean carbon-coated silica ampoule, homogenized by melting/rocking, heated to temperatures in excess of 1100 °C, and then frozen at rates of a few mm/h in a vertical or horizontal system. Single crystals or, more normally, large-grained ingots of size ≈ 75 mm in diameter and 10–15 cm in length are produced.

Growth by variations on the basic THM process have included the addition of ACRT [12.118, 119], focused radiant heating [12.120], sublimation-THM and multipass-THM [12.88], and *cold* THM at 780 °C [12.89]. Crystals are grown either from the precompounded binaries or the elements (as in cold THM). Growth rates tend to be low, few mm/d, but crystals are very pure due to segregation effects and high-resistivity material results (10^7 Ωcm in [12.88]). This makes THM material ideally suited for gamma- and x-ray detectors.

The vast majority of work on bulk growth of MCT has been from the melt. Although rapid progress has taken place in the epitaxial growth techniques for MCT, material grown by several bulk methods is still in use for infrared detection, particularly for photoconductive detectors (this is the case in this author's lab and others worldwide, for example, SPITFIRE in New Zealand). Several historical reviews of the development of bulk MCT have been published [12.121–123]. Many techniques were tried in the early years but three prime techniques survived: SSR, Bridgman and THM.

Two fundamentally different approaches have been followed to improve the basic Bridgman process. These are based on control of melt mixing and of heat flows, respectively. The former has been studied by this author and coworkers [12.124] while the latter includes the work of *Szofran* and *Lehoczky* [12.125] among others. Marked segregation of CdTe with respect to HgTe occurs in the axial direction but this leads to an advantage of the Bridgman process over other techniques, that is, material in all ranges of interest (1–2, 3–5 and 8–12 μm, for $x = 0.7$, 0.3 and 0.2 regions, respectively) is produced in a single run.

A means of stirring melts contained in sealed, pressurized ampoules was needed and the ACRT of *Scheel* and *Shulz-Dubois* [12.43], in which the melt is subjected to periodic acceleration/deceleration at rotation rates of up to 60 rpm, was chosen. The first report

of ACRT in MCT Bridgman growth was given in a patent [12.126]. These effects were developed and discussed in more detail in later papers, which are reviewed in [12.124]. Crystals were produced up to 20 mm in diameter and with x values up to 0.6 in the tip regions of some crystals (this has recently been increased to ≈ 0.93 at 20 mm diameter in this author's lab, *Capper* et al. [12.44, 49, 127], up to 0.4 in 40 mm diameter crystals, while the size has also been increased up to 70 mm).

Triboulet [12.51] developed the THM technique for MCT where diameters up to 40 mm were accomplished and x values up to 0.7 [12.128] for optical communication devices. *Durand* et al. [12.129] employed seeds to produce large, oriented crystals, an advantage of this technique over other bulk methods. *Gille* et al. [12.130] adopted a slightly different approach. A pre-THM step is carried out to quench a Te-rich (53–60%) MCT melt. The entire growth procedure takes several months but gives uniform material. This group also used rotation in the horizontal growth by THM with some success [12.131]. Two groups have applied ACRT to THM and obtained enhanced material properties. *Royer* et al. [12.132] used a saw-tooth ACRT sequence and obtained improved radial and axial compositional uniformity. *Bloedner* and *Gille* [12.133] used ACRT although no significant dependence of crystallinity on rotation sequence was seen. They did, however, achieve an increase in growth rate from 1.5 to 8.5 mm/d.

Bridgman is the main melt growth technique for the Hg- and Cd-based ternary and quaternary compounds while work on the vapor growth of these compounds has been limited. The various Hg-based ternaries (e.g., HgZnTe and HgMnTe) have been studied as potential alternatives to MCT for infrared detectors (*Rogalski* [12.134] gave more details including phase equilibria plots). This work was initiated by a theoretical prediction that the Hg–Te bond is stabilized by the addition of Zn, in particular. Rogalski concluded that THM produced the best quality HgZnTe while Bridgman growth of HgMnTe was described as being similar to MCT growth but with reduced segregation. Crystals of HgMnTe up to 40 mm diameter were produced. *Triboulet* [12.135] discussed alternatives to MCT. He concluded that HgMgTe, HgCdSe, and HgZnSe were not suitable, for a variety of reasons, but that HgMnTe and HgZnTe were potentially suitable.

The main Cd-based ternaries are CdZnTe and CdTeSe, which are used as substrates for epitaxial MCT growth by LPE and MBE. A vast literature exists on these compounds and it is still a very active area of research. Vertical and horizontal Bridgman are the main techniques used [12.98] although VGF has been used [12.98] as has THM [12.51]. High-quality crystals up to 100 mm (and even 125 mm) diameter can be produced by VGF [12.78, 106]. Even larger crystals are being grown by horizontal Bridgman [12.117, 136]. Two main problems remain in these compounds. These are uniformity of Zn, or Se, and impurities.

Pelliciari et al. [12.137] showed how solvent evaporation from a Te-rich solution of CdTe in an open tube system can produce large-grained and even single crystals of CdTe up to 300 mm diameter. This large-area material is aimed at the fabrication of x-ray and γ-ray detectors.

Dilute magnetic semiconductors (DMSs) are a class of semiconductor where the semiconducting properties are changed by the addition of a magnetic ion (e.g., Mn^{2+} or Fe^{2+}). These materials therefore display normal electrical and magnetic properties, and novel ones such as large Faraday rotations and giant magnetoresistance. *Pajaczkowska* [12.138] gave an early review of the phase diagrams, lattice parameters, and growth of many of the Mn- and Fe-based DMS compounds. Vapor growth and hydrothermal growth were mainly used, although the Te-based compounds were grown by the Bridgman technique. The most comprehensive review of DMSs is given in [12.139]. In that book, Giriat and Furdyna review the crystal structure of, and growth methods used for, the Mn-based compounds. The majority of the compounds, particularly the Te-based ones, were grown using Bridgman, or unspecified modifications of it. CdMnS was grown in graphite crucibles under high inert gas pressure, or by CVT (using iodine as transport agent). HgMnS was grown hydrothermally, yielding 2 mm diameter crystals. For Fe-based compounds similar preparation methods to those used for the Mn-based compounds were applied but less Fe can be incorporated in the lattice.

Bismuth telluride (strictly speaking a V–VI compound) is used in thermoelectric applications close to room temperature [12.140]. This material is also grown by the Bridgman technique. Both n-type and p-type materials are required to make the thermoelectric devices, and these can be achieved either by stoichiometric deviations or deliberate doping. In this author's lab bismuth telluride is grown at similar rates of 0.5–2.9 mm/h for 12 mm diameter ingots up to 100 mm length using antimony and iodine as acceptor and donor dopants, respectively.

12.4.4 Oxides/Halides/Phosphates/Borates

Oxides

Hurle [12.45] summarized the issues of the growth of oxides by the Czochralski technique. The high melting points of most oxides of interest necessitate the use

of precious metal crucibles, for example, iridium, platinum, although molybdenum and tungsten ones can be used for lower melting point materials. Extensive and efficient thermal insulation is required, using alumina, magnesia, zirconia, and thoria. Some materials require a partial pressure of oxygen in the growth chamber to prevent reduction of the melt. However, this can lead to oxidation of the crucible, which in turn enters the melt. Careful management of the oxygen pressure and gas flow is needed, together with physical baffles. Heating is normally by r.f. means. Mounting of the seeds is problematic and long seeds are necessary to avoid thermal degradation of the seed holder and pull rod. Rotation or translation of the crucible is normally not used, so the rate of growth has to be above the pulling rate to account for the fall in melt height. Crystals are normally limited to less than half the crucible diameter for stable growth. Growth rates are low, a few mm/h or less, to avoid unstable growth. Afterheaters are also employed to reduce the built-in stress in the crystal during the later stages of growth and subsequent cool down. Growth runs are several days so automatic diameter control, in this instance using weighing techniques, is mandatory. The high temperatures lead to high-temperature gradients and thus strong bouyancy-driven convection that is turbulent by nature. This turbulence gives rise to nonuniformities in the crystals. In addition, catastrophic failure can result from coupling between this turbulent flow and the flow due to the crystal rotation. Thus, understanding the convective flows in oxide growth is critical if large-diameter highly perfect crystals are to be produced.

Fukuda et al. [12.141, 142] have presented detailed reviews of the growth of a wide range of oxide materials of interest for optical and optoelectronic applications. They note that only $LiNbO_3$ (LN), $LiTaO_3$ (LT), and $Y_3Al_5O_{12}$ (YAG) have as yet reached production status. Many of the other materials exhibit superior properties but are not available in large quantities at sufficient perfection to warrant them being used in devices. In general, the growth methods are based on liquid–solid transitions, either from the melt or from high-temperature solutions (flux), for congruently and incongruently melting materials, respectively. Czochralski growth is still the mainstay of oxide growth, although problems with shape (diameter) control of some oxides (rare-earth vanadates, rutile, etc.) necessitate the use of edge-defined film-fed growth (EFG). Growth from fluxes is used for a range of oxide materials, although some incongruently melting materials can be produced from melts of the same composition as the crystal by a micro-pulling down technique.

In the micro-pulling down technique the melt is held in a Pt crucible and a micro-nozzle (die) is arranged at the bottom of the melt. Below this arrangement is placed an afterheater. This combination produces a steep temperature gradient (300 °C/mm), which ensures stable growth at high rates. Single crystals of 0.05–1.0 mm diameter are produced onto seeds. LN crystals, for example, are grown at 12–90 mm/h from stoichiometric melts up to 100 mm in length, assumed only to be limited by the melt volume used. Dislocations are introduced if the diameter of the crystals exceeds 0.8 mm. Growth of $K_3Li_2Nb_5O_{12}$ (KLN), which has outstanding electro-optic and nonlinear optical properties, is problematic by either Czochralski or Kyropoulos methods due to segregation effects at the solid/liquid interface and cracking. This material was grown by the micro-pulling down technique at rates of 20–80 mm/h with diameters of 0.15–0.5 mm. The crystal composition was found to be near to the melt composition and crystals were single domain.

Crystals in the $KTiOPO_4$ (KTPO) family were grown by the flux method using slow cooling from 1100 to 800 °C. Crystals of a few mm in size were obtained for second-harmonic generation measurements.

Crystals from the langasite family ($La_3Ga_3SiO_{12}$) are used for lasers. These were grown by the Czochralski method at 50 mm diameter and 130 mm length. High-purity oxide powders were used as starting materials and an Ar/O_2 mixture was used to suppress evaporation of the gallium suboxide. A pulling rate of 1.0 mm/h was used together with a crystal rotation rate of 10 rpm. The resulting crystals were single phase and crack free. These crystals show superior properties, as filters, to quartz. A range of langasite-type crystals up to 100 mm diameter have also been grown by both Czochralski and VB techniques ([12.8, Chap. 15], [12.143]).

Another means of overcoming problems in Czochralski growth of these materials is to use the EFG method. Rutile (TiO_2) is used as a polarizer in optical isolators and prisms for optical communication systems. Crystals of this material can be grown by Verneuil and float-zone methods but growth by Czochralski is difficult. Iridium is used for the die material and crystals up to 8×1 cm and 0.5 mm thick were grown. A modified die was used to grow a rod of 15 mm diameter and ≈ 40 mm length. A double-die system was shown that was used to produce a core-doped crystal of LN, using Nd^{3+} and Cr^{3+} as the dopants. A 5 mm diameter crystal some 60 mm in length was produced. Such crystals have potential in optical applications where pumping energy is only absorbed in the core region, reducing heat-induced energy losses.

Quartz is used in both piezoelectric and in various optical applications [12.144]. *Balitsky* [12.145] noted

that some 2500 t were produced per year at that time. Hydrothermal growth is the main growth technique used in order to keep the growth temperature below the β- (nonpiezoelectric) to α-quartz (piezoelectric) phase transition. Growth at temperatures of 250–500 °C and pressures of 50–200 MPa are used in both alkaline and fluoride solutions. Growth rates of 0.5–1.0 mm/d are achieved on (0001) or (1120) seeds and crystals up to $200 \times 70 \times 60$ mm in size weighing from 1–18 kg.

Various oxide materials are used widely in scintillator applications. The most common are $Bi_4Ge_2O_{12}$ (BGO), up to 140 mm diameter and 400 mm long [12.8, Chap. 8] and $PbWO_4$ (PWO) [12.146, 147]. These materials are grown by normal Czochralski and Bridgman techniques and the requirement is for thousands of each for large high-energy/astrophysics projects in several centers throughout the world. For example, some 85 000 PWO crystals were required for one application at the Large Hadron Collider at CERN. High-purity starting materials are important as are certain growth parameters and post-growth annealing, for example, in oxygen, to reduce stress in the material.

Sapphire is grown by several techniques for a wide variety of optical, microwave and microelectronics applications [12.148]. Verneuil [12.149] and Czochralski and EFG [12.45] have been used extensively to produce large single crystals, particularly for laser rods, and shaped pieces using various forms of dies. In addition, the horizontal directional solidification crystallization (HDC) technique [12.148] has been used. A molybdenum boat filled with the charge is simply moved through the temperature gradient. Low-temperature gradients lead to crystals with low stress and low-dislocation densities (compared with the Verneuil and pulling techniques. The growth rate is high at 8 mm/h and crystals up to 10 kg and up to $30 \times 30 \times 2.5\,cm^3$ and even $50 \times 35 \times 4\,cm^3$ [12.150] can be produced. Even larger crystals, up to 20 kg with a diameter of 200 mm) can be produced by a modified Kyropolous method [12.151] and 500 mm diameter crystals were noted by *Friedrich* [12.8, Chap. 5].

Another technique used is the heat exchanger method, which is a solidification technique from the melt [12.152]. By independently controlling the heat input and the heat output, large crystals can be grown onto seeds. The seed is prevented from melting by a helium gas flow at the bottom of the melt. By progressively increasing the helium flow, following partial melting of the seed, and/or decreasing the furnace temperature growth is progressed. After complete solidification, an annealing step is performed prior to cooling to room temperature. The submerged solid/liquid interface damps out any thermal and mechanical variations and also produces a low-temperature gradient at the interface. There is impurity segregation into the last-to-freeze parts of the crystal that form near the crucible walls. These areas are removed before machining takes place. Crystals up to 340 mm in diameter (65 kg) have been produced in this way. Slabs cut from these crystals are used as optical windows, between 5–25 mm thick. *Demina* et al. [12.9, p. 213] note that in 2007 the total volume of 2 in equivalent wafers for all applications was $\approx 5\,000\,000$.

Bruni [12.9, p. 267] reports on the growth of a 110 mm GGG crystal weighing 25 kg, which was used in laser applications.

There has been renewed interest recently in the growth and applications of ZnO [12.153, 154]. It is shown that 50 mm and even 75 mm diameter crystals (up to 10 mm thick) can be grown by both vapor and hydrothermal techniques.

Halides

Another class of scintillator materials are halides. These are used in a range of applications from high-energy physics, nuclear medicine, environmental monitoring and security systems [12.155]. Both large diameter, comparable with the human body, and large numbers, for the high-energy physics applications, are required. The simple Bridgman growth process is still used for many of these materials, although problems do arise due to the contact between crystal and crucible. Insufficient melt mixing also leads to inclusions and striations in crystals. The alternative technique for growth is that of continuously fed Czochralski/Kyropolous growth. A melt-level sensor is normally used to control the diameter over the tens/hundreds of hours of growth of large halide crystals. Replenishment of the melt by molten material from a closed feeder enables purification of the starting materials, and a conical crucible that is not too deep is used. Crystals of CsI up to 500 mm diameter and 750 mm height, weighing ≈ 550 kg have been produced [12.156] and CdF_2 up to 400 mm diameter [12.8, p. 140]. The ultimate aim is to produce crystals > 1 m in diameter so that entire crystals can be used to monitor the human body.

Phosphates, Borates and Tungstates

Nonlinear optical materials are very important for laser-frequency-conversion applications. One of the most important of the phosphates is potassium dihydrogen phosphate (KDP) that is used for higher harmonic generation in large laser systems for fusion experiments [12.105, 157]. Growth takes place at room temperature to 60 °C and growth rates can be as high as 10–20 mm/d with sizes of $\approx 40 \times 40 \times 85$ cm [12.157] or $45 \times 45 \times 70$ cm [12.105] or $50 \times 50 \times 50$ cm [12.158], the latter weighing some 300 kg, some taking over

a year to grow! *Schaffers* et al. [12.9, p. 229] noted that the National Ignition Facility required some 75 such large-scale crystals weighing over 100 t. They also reported on the production of 70 mm diameter crystals of $Ca_5(PO_4)_3F$ and 80 mm diameter crystals of $YCa_4O(BO_3)_2$ for laser applications, for example, frequency-conversion materials.

Another important phosphate is potassium titanyl phosphate (KTP) used to get green light by frequency doubling a Nd:YAG laser. Growth in this case is from high-temperature solution at about 950 °C [12.105]. Sizes up to $32 \times 42 \times 87$ mm (weight 173 g) can be grown in 40 days.

Borates, including barium borate, lithium borate, cesium borate and coborates, for example, cesium lithium borate are used in UV-generation applications. Crystals are again grown by the high-temperature solution method up to $14 \times 11 \times 11$ cm in size, weighing 1.8 kg, in 3 weeks [12.105]. Crystals of β-BaB_2O_4 have recently been produced up to 60 mm diameter [12.159], while $YCa_4O(BO_3)_2$ (YCOB) material up to 100 mm diameter for SHG have also been grown [12.160].

Crystals of $CaWO_4$ up to 40 mm diameter weighing 1.8 kg are currently being used in dark-matter research [12.161].

12.5 Conclusions

This chapter has summarized the current position on bulk growth of crystals for optoelectronic and electronic applications. It is not intended to be a completely comprehensive view of the field, merely serving to introduce the reader to the wide range of materials produced and the numerous crystal growth techniques that have been developed to grow single crystals. An historical perspective has been attempted to give the reader a feel for the scale of some of the activities. The sections on specific materials try to summarize the particular growth techniques employed, and those that cannot in some cases, and outline the typical sizes currently produced commercially and in the R&D areas. For more details on current developments see the books [12.8, 9, 44, 49, 69].

Acknowledgments. The author wishes to thank the management of Leonardo MW Ltd for permission to publish this chapter.

References

12.1 H.J. Scheel: J. Cryst. Growth **211**, 1 (2000)
12.2 H.E. Buckley: *Crystal Growth* (Wiley, New York 1951)
12.3 J.G. Burke: *Origins of the Science of Crystals* (Univ. California Press, Berkeley 1966)
12.4 D. Elwell, H.J. Scheel: *Crystal Growth from High-Temperature Solutions* (Academic, New York 1975)
12.5 J.C. Brice: *Crystal Growth Processes* (Blackie, London 1986)
12.6 H.J. Scheel: In: *The Technology of Crystal Growth and Epitaxy*, ed. by H.J. Scheel, T. Fukuda (Wiley, Chichester 2003)
12.7 A.A. Chernov: J. Mater. Sci. Mater. Electron. **12**, 437 (2001)
12.8 H.J. Scheel, P. Capper (Eds.): *Crystal Growth Technology. From Fundamentals and Simulation to Large-Scale Production* (Wiley, Weinheim 2008)
12.9 P. Capper, P. Rudolph (Eds.): *Crystal Growth Technology. Semiconductors and Dielectrics* (Wiley, Weinheim 2010)
12.10 A.V.L. Verneuil: Compt. Rend. **135**, 791 (1902)
12.11 W. Nernst: Z. Phys. Chem. **47**, 52 (1904)
12.12 M. Volmer: Z. Phys. Chem. **102**, 267 (1927)
12.13 W. Kossel: Nachr. Ges. Wiss. Göttingen Math.-Phys. **KI**, 135 (1927)
12.14 I.N. Stranski: Z. Phys. Chem. **136**, 259 (1928)
12.15 G. Spezia: Acad. Sci. Torino Atti **30**, 254 (1905)
12.16 G. Spezia: Acad. Sci. Torino Atti **44**, 95 (1908)
12.17 J. Czochralski: Z. Phys. Chem. **92**, 219 (1918)
12.18 S. Kyropoulos: Z. Anorg. Chem. **154**, 308 (1926)
12.19 P.W. Bridgman: Proc. Am. Acad. Arts Sci. **58**, 165 (1923)
12.20 P.W. Bridgman: Proc. Am. Acad. Arts Sci. **60**, 303 (1925)
12.21 F. Stöber: Z. Kristallogr. **61**, 299 (1925)
12.22 D.C. Stockbarger: Rev. Sci. Instrum. **7**, 133 (1936)
12.23 H.C. Ramsberger, E.H. Malvin: J. Opt. Soc. Am. **15**, 359 (1927)
12.24 G.K. Teal, J.B. Little: Phys. Rev. **78**, 647 (1950)
12.25 W.C. Dash: J. Appl. Phys. **30**, 459 (1959)
12.26 W.G. Pfann: Trans. AIME **194**, 747 (1952)
12.27 H.C. Theurer: Method of Processing Semiconductor Materials, US Patent (Application) 3 060 123 (1952)
12.28 P.H. Keck, M.J.E. Golay: Phys. Rev. **89**, 1297 (1953)
12.29 F.C. Frank: Discuss. Faraday Soc. **5**, 48 (1949)
12.30 W.K. Burton, N. Cabrera, F.C. Frank: Philos. Trans. A **243**, 299 (1951)
12.31 G.P. Ivantsov: Dokl. Akad. Nauk SSSR **81**, 179 (1952)
12.32 G.P. Ivantsov: Dokl. Akad. Nauk SSSR **83**, 573 (1953)
12.33 W.A. Tiller, K.A. Jackson, J.W. Rutter, B. Chalmers: Acta Metall. Mater. **1**, 428 (1953)
12.34 A.E. Carlson: PhD Thesis (Univ. Utah, Utah 1958)

12.35 H.J. Scheel, D. Elwell: J. Cryst. Growth **12**, 153 (1972)
12.36 J.A. Burton, R.C. Prim, W.P. Slichter: J. Chem. Phys. **21**, 1987 (1953)
12.37 W. van Erk: J. Cryst. Growth **57**, 71 (1982)
12.38 D. Rytz, H.J. Scheel: J. Cryst. Growth **59**, 468 (1982)
12.39 L. Wulff: Z. Krystallogr. **11**, 120 (1886)
12.40 L. Wulff: Z. Krystallogr. **100**, 51 (1886)
12.41 F. Krüger, W. Finke: Kristallwachstumsvorrichtung, Deutsches Reichs Patent DRP 228 246 (1910), (5.11.1910)
12.42 A. Johnsen: *Wachstum und Auflösung der Kristalle* (Wilhelm Engelmann, Leipzig 1910)
12.43 H.J. Scheel, E.O. Schulz-Dubois: J. Cryst. Growth **8**, 304 (1971)
12.44 P. Capper (Ed.): *Bulk Crystal Growth of Electronic, Optical and Optoelectronic Materials* (Wiley, Chichester 2005)
12.45 D.T.J. Hurle: *Crystal Pulling from the Melt* (Springer, Berlin, Heidelberg 1993)
12.46 J.B. Mullin: Compound semiconductor processing. In: *Compound Semiconductor Devices: Structures and Processing*, ed. by K.A. Jackson (Wiley, Weinheim 1998) p. 1
12.47 P. Rudolph, K. Kakimoto: Mater. Res. Soc. Bull. **34**, 1 (2009)
12.48 C.J. Jones, P. Capper, J.J. Gosney, I. Kenworthy: J. Cryst. Growth **69**, 281 (1984)
12.49 P. Capper: J. Mater. Sci. Mater. Electron. **26**, 4380 (2015)
12.50 P. Capper: Prog. Cryst. Growth Charact. Mater. **28**, 1 (1994)
12.51 R. Triboulet: Prog. Cryst. Growth Charact. Mater. **28**, 85 (1994)
12.52 M.A. Green: Developments in crystalline silicon cells. In: *Solar Cell Materials: Developing Technologies*, ed. by G.J. Conibeer, A.F.W. Willoughby (Wiley, Chichester 2014) p. 67
12.53 M. Shiraishi, K. Takano, J. Matsubara, N. Iida, N. Machida, M. Kuramoto, H. Yamagishi: J. Cryst. Growth **229**, 17 (2001)
12.54 K. Hoshikawa, Huang Xinming, T. Taishi: J. Cryst. Growth **275**(1/2), 276 (2004)
12.55 L. Jensen: *Paper Given at 1st International School on Crystal Growth and Technology* (Beatenberg, Switzerland 1998)
12.56 T. Ciszek: Silicon crystal growth for photovoltaics. In: *Crystal Growth Technology*, ed. by H.J. Scheel, T. Fukuda (Wiley, Chichester 2003) p. 267
12.57 Proc. 12th Int. Conf. Cryst. Growth, J. Cryst. Growth **189** (1999)
12.58 Proc. 10th Int. Conf. Vapor Growth Epitaxy, J. Cryst. Growth **199** (1999)
12.59 Proc. 13th Int. Conf. Crsyt. Growth, J. Cryst. Growth **237** (2002)
12.60 Proc. 11th Int. Conf. Vapor Growth Epitaxy, J. Cryst. Growth **239** (2002)
12.61 Proc. 1st Asian Conf. Cryst. Growth Technol., J. Cryst. Growth **229** (2001)
12.62 Proc. 14th Int. Conf. Cryst. Growth, J. Cryst. Growth **275** (2005)
12.63 Proc. 12th Int. Conf. Vapor Growth Epitaxy, J. Cryst. Growth **275** (2005)
12.64 Proc. 15th Int. Conf. Cryst. Growth, J. Cryst. Growth, **310** (2007)
12.65 Proc. 16th Int. Conf. Cryst. Growth, J. Cryst. Growth, **318** (2011)
12.66 Proc. 17th Int. Conf. Cryst. Growth Epitaxy J. Cryst. Growth, **401** (2014)
12.67 P. Capper (Ed.): *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications* (Chapman Hall, London 1997)
12.68 P. Capper, C.T. Elliott (Eds.): *Infrared Detectors and Emitters: Materials and Devices* (Kluwer, Boston 2001)
12.69 H.J. Scheel, T. Fukuda (Eds.): *Crystal Growth Technology* (Wiley, Chichester 2003)
12.70 S. Nishino: Chemical vapour deposition of SiC. In: *Properties of Silicon Carbide*, EMIS Datarev., Vol. 13, ed. by G.L. Harris (IEE, London 1995) p. 204
12.71 A.O. Konstantinov: Sublimation growth of SiC. In: *Properties of Silicon Carbide*, EMIS Datarev., Vol. 13, ed. by G.L. Harris (IEE, London 1995) p. 170
12.72 N. Nordell: Bulk and epitaxel growth of SiC. In: *Process Technology for Silicon Carbide Devices*, ed. by C.M. Zetterling (IEE, London 2002) p. 13
12.73 H. Kanda, T. Sekine: High temperature high pressure sysnthesis of single crystal diamond. In: *Properties, Growth and Applications of Diamond*, EMIS Datarev., Vol. 26, ed. by M.H. Nazare, A.J. Neves (IEE, London 2001) p. 247
12.74 H. Sumiya: High pressure synthesis of large high-quality single crystal diamonds, Proc. Int. Conf. Cryst. Growth Epitaxy, Warsaw (2013) p. 70
12.75 R.S. Sussmann (Ed.): *CVD Diamond for Electronic Devices and Sensors* (Wiley, Chichester 2009)
12.76 W.G. Pfann: *Zone Melting*, 2nd edn. (Wiley, New York 1966)
12.77 P. Rudolph: Thermodynamic fundamentals of phase transitions applied to crystal growth processes. In: *Crystal Growth Technology*, ed. by H.J. Scheel, T. Fukuda (Wiley, Chichester 2003) p. 15
12.78 T. Asahi, K. Kainosho, K. Kohiro, A. Noda, K. Sato, O. Oda: Growth of III–V and II–VI single crystals by the vertical-gradient freeze method. In: *Crystal Growth Technology*, ed. by H.J. Scheel, T. Fukuda (Wiley, Chichester 2003) p. 323
12.79 W.F.J. Micklethwaite, A.J. Johnson: InSb: Materials and devices. In: *Infrared Detectors and Emitters: Materials and Devices*, ed. by P. Capper, C.T. Elliott (Kluwer, Boston 2001) p. 177
12.80 I. Grzegory, S. Porowski: High pressure solution growth of GaN and related compounds. In: *Properties, Processing and Applications of Gallium Nitride and Related Semiconductors*, EMIS Datarev., Vol. 23, ed. by J.H. Edgar, S. Strite, I. Akasaki, H. Amano, C. Wetzel (IEE, London 1999) p. 359
12.81 I. Grzegory, S. Krukowski, M. Leszczynski, P. Perlin, T. Suski, S. Porowski: High pressure crystalization of GaN. In: *Nitride Semiconductors: Handbook on Materials and Devices*, ed. by P. Ruterana, M. Albrecht, J. Neugebauer (Wiley, Weinheim 2003) p. 3
12.82 K. Nishino, S. Sakai: Sublimination growth of GaN and AlN. In: *Properties, Processing and Applica-*

tions of Gallium Nitride and Related Semiconductors, EMIS Datarev., Vol. 23, ed. by J.H. Edgar, S. Strite, I. Akasaki, H. Amano, C. Wetzel (IEE, London 1999) p. B12
12.83 P. Rudolph: Melt growth of II–VI compound single crystals. In: *Recent Developments of Bulk Crystal Growth 1998*, ed. by M. Isshiki (Research Signpost, Trivandrum 1998) p. 127
12.84 H. Hartmann, K. Bottcher, D. Siche: In: *Recent Developments of Bulk Crystal Growth 1998*, ed. by M. Isshiki (Research Signpost, Trivandrum 1998) p. 165
12.85 B.J. Fitzpatrick, P.M. Harnack, S. Cherin: Philips J. Res. **41**, 452 (1986)
12.86 P. Capper, J.E. Harris, D. Nicholson, D. Cole: J. Cryst. Growth **46**, 575 (1979)
12.87 R. Triboulet, T. Nguyen Duy, A. Durand: J. Vac. Sci. Technol. A **3**, 95 (1985)
12.88 R. Triboulet, Y. Marfaing: J. Cryst. Growth **51**, 89 (1981)
12.89 R. Triboulet, K. Van Pham, G. Didier: J. Cryst. Growt **101**, 216 (1990)
12.90 V.A. Kuznetsov: Prog. Cryst. Growth Charact. **21**, 163 (1990)
12.91 J. Mimila, R. Triboulet: Mater. Lett. **24**, 221 (1995)
12.92 W.W. Piper, S.J. Polich: J. Appl. Phys. **32**, 1278 (1961)
12.93 G.J. Russell, J. Woods: J. Cryst. Growth **46**, 323 (1979)
12.94 P. Blanconnier, P. Henoc: J. Cryst. Growth **17**, 218 (1972)
12.95 K. Durose, A. Turnbull, P.D. Brown: Mater. Sci. Eng. B **16**, 96 (1993)
12.96 M.R. Lorenz: In: *Physics and Chemistry of II–VI Compounds*, ed. by M. Aven, J.S. Prener (North-Holland, Amsterdam 1967)
12.97 K. Zanio: Semicond. Semimet. **13** (1978)
12.98 P. Capper, A. Brinkman: Growth of CdTe, CdZnTe and CdTeSe by bulk methods. In: *Properties of Narrow Gap Cadmium-Based Compounds*, EMIS Datarev, Vol. 10, ed. by P. Capper (IEE, London 1994) p. 369
12.99 A.W. Brinkman: In: *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications*, ed. by P. Capper (Chapman Hall, London 1997)
12.100 R. Triboulet, J.O. Ndap, A. El Mokri, A. Tromson-Carli, A. Zozime: J. Phys. IV **5**(C3), C3–141 (1995)
12.101 R. Triboulet, A. Tromson-Carli, D. Lorans, T. Nguyen Duy: J. Electron. Mater. **22**, 827 (1993)
12.102 J.B. Mullin, C.A. Jones, B.W. Straughan, A. Royle: J. Cryst. Growth **59**, 135 (1982)
12.103 H.M. Hobgood, B.W. Swanson, R.N. Thomas: J. Cryst. Growth **85**, 510 (1987)
12.104 R. Triboulet, Y. Marfaing: J. Electrochem. Soc. **120**, 1260 (1973)
12.105 T. Sasaki, Y. Mori, M. Yoshimura: Growth of nonlinear optical crystals for laser frequency conversion. In: *The Technology of Crystal Growth and Epitaxy*, ed. by H.J. Scheel, T. Fukuda (Wiley, Chichester 2003), Chap. 20
12.106 R. Hirano, H. Kurita: Bulk growth of CdZnTe/CdTe crystals. In: *Bulk Crystal Growth of Electronic, Optical and Optoelectronic Materials*, ed. by P. Capper (Wiley, Chichester 2005)
12.107 K. Zanio: J. Electron. Mater. **3**, 327 (1974)
12.108 C.P. Khattak, F. Schmid: Proc. SPIE **1106**, 47 (1989)
12.109 W.M. Chang, W.R. Wilcox, L. Regel: Mater. Sci. Eng. B **16**, 23 (1993)
12.110 N.R. Kyle: J. Electrochem. Soc. **118**, 1790 (1971)
12.111 J.C. Tranchart, B. Latorre, C. Foucher, Y. LeGouce: J. Cryst. Growth **72**, 468 (1985)
12.112 Y.-C. Lu, J.-J. Shiau, R.S. Fiegelson, R.K. Route: J. Cryst. Growth **102**, 807 (1990)
12.113 J.P. Tower, S.B. Tobin, M. Kestigian, P.W. Norton, A.B. Bollong, H.F. Schaake, C.K. Ard: J. Electron. Mater. **24**, 497 (1995)
12.114 S. Sen, S.M. Johnson, J.A. Kiele, W.H. Konkel, J.E. Stannard: Mater. Res. Soc. Symp. Proc. **161**, 3 (1990)
12.115 J.F. Butler, F.P. Doty, B. Apotovsky, J. Lajzerowicz, L. Veiger: Mater. Sci. Eng. B **16**, 291 (1993)
12.116 P. Capper, J.E. Harris, E. O'Keefe, C.L. Jones, C.K. Ard, P. Mackett, D.T. Dutton: Mater. Sci. Eng. B **16**, 29 (1993)
12.117 C. Szeles, S.E. Cameron, S.A. Soldner, J.-O. Ndap, M.D. Reed: J. Electron. Mater. **33**(6), 742 (2004)
12.118 A. El Mokri, R. Triboulet, A. Lusson, A. Tromson-Carli, G. Didier: J. Cryst. Growth **138**, 168 (1995)
12.119 R.U. Bloedner, M. Presia, P. Gille: Adv. Mater. Opt. Electron. **3**, 233 (1994)
12.120 R. Schoenholz, R. Dian, R. Nitsche: J. Cryst. Growth **72**, 72 (1985)
12.121 W.F.H. Micklethwaite: Semicond. Semimet. **18**, 3 (1981)
12.122 P.W. Kruse: Semicond. Semimet. **18**, 1 (1981)
12.123 H. Maier: Proc. N.A.T.O. Adv. Res. Workshop Future Small-Gap II–VI Semicond., Liege (1988)
12.124 P. Capper: Prog. Cryst. Growth Charact. **19**, 259 (1989)
12.125 F.R. Szofran, S.L. Lehoczky: J. Cryst. Growth **70**, 349 (1984)
12.126 P. Capper, J.J.G. Gosney: UK Patent (Application) 8115911 (1981)
12.127 P. Capper, C. Maxey, C. Butler, M. Grist, J. Price: J. Mater. Sci. Mater. Electron. **15**, 721 (2004)
12.128 Y. Nguyen Duy, A. Durand, J.L. Lyot: Mater. Res. Soc. Symp. Proc. **90**, 81 (1987)
12.129 A. Durand, J.L. Dessus, T. Nguyen Duy, J.F. Barbot: Proc. SPIE **659**, 131 (1986)
12.130 P. Gille, F.M. Kiessling, M. Burkert: J. Cryst. Growth **114**, 77 (1991)
12.131 P. Gille, M. Pesia, R.U. Bloedner, N. Puhlman: J. Cryst. Growth **130**, 188 (1993)
12.132 M. Royer, B.R. Jean, A.R. Durand, R. Triboulet: French Patent (Application) 8804370 (1988)
12.133 R.U. Bloedner, P. Gille: J. Cryst. Growth **130**, 181 (1993)
12.134 A. Rogalski: *New Ternary Alloy Systems for Infrared Detectors* (SPIE, Bellingham 1994)
12.135 R. Triboulet: Semicond. Sci. Technol. **5**, 1073 (1990)
12.136 R. Korenstein, R.J. Olson Jr., D. Lee: J. Electron. Mater. **24**, 511 (1995)

12.137 B. Pelliciari, F. Dierre, D. Brellier, B. Schaub: J. Cryst. Growth **275**, 99 (2005)

12.138 A. Pajaczkowska: Prog. Cryst. Growth Charact. **1**, 289 (1978)

12.139 W. Giriat, J.K. Furdyna: Semicond. Semimet. **25**, 1 (1988)

12.140 M.C.C. Custodio, A.C. Hernandes: J. Cryst. Growth **205**, 523 (1999)

12.141 T. Fukuda, V.I. Chani, K. Shimamura: Growth of oxide single crystals for electrical and electro-optical applications. In: *Recent Developments of Bulk Crystal Growth 1998*, ed. by M. Isshiki (Research Signpost, Trivandrum 1998) p. 191

12.142 T. Fukuda, V.I. Chani, K. Shimamura: Melt growth of oxide crystals for SAW, piezoelectric and nonlinear optical applications. In: *The Technology of Crystal Growth and Epitaxy*, ed. by H.J. Scheel, T. Fukuda (Wiley, Chichester 2003)

12.143 S. Uda, S.Q. Wang, N. Konishi, H. Inaba, J. Harada: J. Cryst. Growth **237/239**, 707 (2002)

12.144 F. Iwasaki, H. Iwasaki: J. Cryst. Growth **237/239**, 820 (2002)

12.145 V.S. Balitsky: In: *1st Int. School Cryst. Growth Technol., Beatenberg* (1998)

12.146 M. Korzhik: In: *1st Int. School Cryst. Growth Technol., Beatenberg* (1998)

12.147 P.J. Li, Z.W. Yin, D.S. Yan: In: *1st Int. School Cryst. Growth Technol., Beatenberg* (1998)

12.148 K.S. Bagdasarov, E.V. Zharikov: In: *1st Int. School Cryst. Growth Technol., Beatenberg* (1998)

12.149 L. Lytvynov: In: *2nd Int. School on Cryst. Growth Technol., Zao* (2000)

12.150 V.M. Puzikov, G. Adonkin, V. Baranov, A. Budnikov, L. Gryu, S.I. Nizhankovskyi, N. Sidelnikova, A. Tauko: Growth of large-size Sapphire and Ti:Sapphire crystals with high optical and structural characteristics by HDC method, Proc. Int. Conf. Cryst. Growth Epitaxy, Warsaw (2013)

12.151 M.I. Moussatov, E.V. Zharikov: Proc. 1st Int. School Cryst. Growth Technol., Beatenberg (1998)

12.152 F. Schmid, C.P. Khattak: Proc. 1st Int. School Cryst. Growth Technol., Beatenberg (1998)

12.153 M.J. Callahan, D. Ehrentraut, M.N. Alexander, B. Wang: Growth mechanisms and properties of hydrothermal ZnO. In: *Zinc Oxide Materials for Electronic and Optoelectronic Device Applications*, ed. by C.W. Litton, D.C. Reynolds, T.C. Collins (Wiley, Chichester 2011) p. 189

12.154 R. Shimada, H. Morkoç: Growth and characterization of GaN/ZnO hetroepitaxy and ZnO-based hybrid devices. In: *Zinc Oxide Materials for Electronic and Optoelectronic Device Applications*, ed. by C.W. Litton, D.C. Reynolds, T.C. Collins (Wiley, Chichester 2011) p. 221

12.155 A.V. Gektin, B.G. Zaslavsky: In: *1st Int. School Cryst. Growth Technol., Beatenberg* (1998)

12.156 A.V. Gektin: In: *2nd Int. School Cryst. Growth Technol., Zao* (2000)

12.157 N. Zaitseva, L. Carman, I. Smolsky: J. Cryst. Growth **241**, 363 (2002)

12.158 X. Zhuang, L. Ye, G. Zheng, G. Su, Y. He, X. Lin, Z. Xu: J. Crystal Growth **318**(1), 700 (2011)

12.159 E.A. Simonova, N.G. Kononova, P. Federov, A.E. Kokh, V.S. Shevchenko: Investigation on the Li,Ba/Bo_2, F ternary reciprocal system and growth of bulk β-BaB_2O_4 crystals, Int. Conf. Cryst. Growth Epitaxy, Warsaw (2013) p. 105

12.160 X. Tu, Y. Zheng, K. Xiong, Y. Tu, E. Shi: Crystal growth and characterization of 4 inches $YCa_4O(BO_3)_3$ crystals, Int. Conf. Cryst. Growth Epitaxy, Warsaw (2013) p. 432

12.161 E. Cartlidge: *Physics World* (Institute of Physics, Bristol 2013) p. 12

13. Single-Crystal Silicon: Growth and Properties

Fumio Shimura

It is clear that silicon, which has been the dominant material in the semiconductor industry for some time, will carry us into the coming ultra-large-scale integration (ULSI) and system-on-a-chip (SOC) eras, even though silicon is not the optimum choice for every electronic device. Semiconductor devices and circuits are fabricated through many mechanical, chemical, physical, and thermal processes. The preparation of silicon single-crystal substrates with mechanically and chemically polished surfaces is the first step in the long and complex device fabrication process. In this chapter, the approaches currently used to prepare silicon materials (from raw materials to single-crystalline silicon) are discussed.

13.1 **Overview** 294
13.2 **Starting Materials** 295
13.2.1 Metallurgical-Grade Silicon 295
13.2.2 Polycrystalline Silicon 295
13.3 **Single-Crystal Growth** 296
13.3.1 Floating-Zone Method 297
13.3.2 Czochralski Method 299
13.3.3 Impurities in Czochralski Silicon 301
13.4 **New Crystal Growth Methods** 304
13.4.1 Czochralski Growth with an Applied Magnetic Field (MCZ) 304
13.4.2 Continuous Czochralski Method (CCZ) 305
13.4.3 Neckingless Growth Method 306
References 306

Silicon, which has been and will continue to be the dominant material in the semiconductor industry for some time to come [13.1], will carry us into the ultra-large-scale integration (ULSI) era and the system-on-a-chip (SOC) era.

As electronic devices have become more advanced, device performance has become more sensitive to the quality and the properties of the materials used to construct them.

Germanium (Ge) was originally utilized as a semiconductor material for solid state electronic devices. However, the narrow bandgap (0.66 eV) of Ge limits the operation of germanium-based devices to temperatures of approximately 90 °C because of the considerable leakage currents observed at higher temperatures. The wider bandgap of silicon (1.12 eV), on the other hand, results in electronic devices that are capable of operating at up to ≈ 200 °C. However, there is a more serious problem than the narrow bandgap: germanium does not readily provide a stable passivation layer on the surface. For example, germanium dioxide (GeO_2) is water-soluble and dissociates at approximately 800 °C. Silicon, in contrast to germanium, readily accommodates surface passivation by forming silicon dioxide (SiO_2), which provides a high degree of protection to the underlying device. This stable SiO_2 layer results in a decisive advantage for silicon over germanium as the basic semiconductor material used for electronic device fabrication. This advantage has lead to a number of new technologies, including processes for diffusion doping and defining intricate patterns. Other advantages of silicon are that it is completely nontoxic, and that silica (SiO_2), the raw material from which silicon is obtained, comprises approximately 60% of the mineral content of the Earth's crust. This implies that the raw material from which silicon is obtained is available in plentiful supply to the integrated circuit (IC) industry. Moreover, electronic-grade silicon can be obtained at less than one-tenth the cost of germanium. All of these advantages have caused silicon to almost completely replace germanium in the semiconductor industry.

Although silicon is not the optimum choice for every electronic device, its advantages mean that it will almost certainly dominate the semiconductor industry for some time yet.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_13

13.1 Overview

Very fruitful interactions have occurred between the users and manufacturers of semiconductor material since the invention of the point-contact transistor in 1947, when the necessity for *perfect and pure* crystals was recognized. The competition was often such that the crystal quality demanded by new devices could only be met by controlling crystal growth using electronic equipment built with these new devices. Since dislocation-free silicon crystals were grown as early as the 1960s using the *Dash technique* [13.2], semiconductor material research and developmental efforts have concentrated on material purity, production yields, and problems related to device manufacture.

Semiconductor devices and circuits are fabricated using a wide variety of mechanical, chemical, physical and thermal processes. A flow diagram for typical semiconductor silicon preparation processes is shown in Fig. 13.1. The preparation of silicon single-crystal substrates with mechanically and chemically polished surfaces is the first step in the long and complex process of device fabrication.

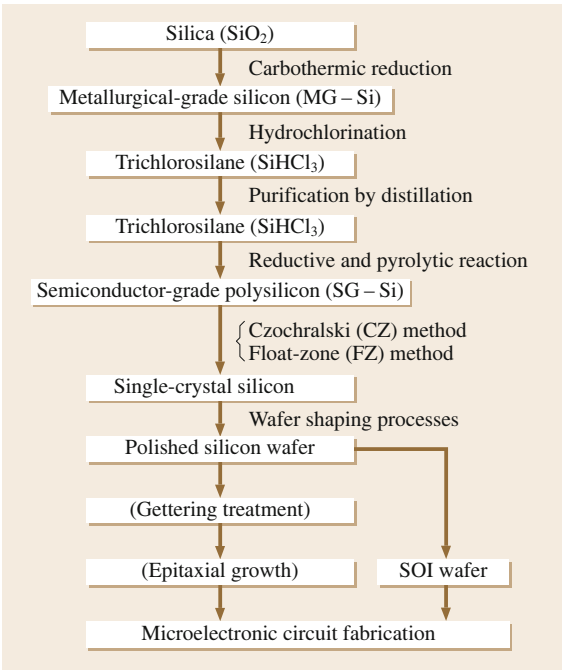


Fig. 13.1 Flow diagram for typical semiconductor silicon preparation processes. (After [13.1])

As noted above, silicon is the second most abundant element on Earth; more than 90% of the Earth's crust is composed of silica and silicates. Given this boundless supply of raw material, the problem is then to transform silicon into the usable state required by the semiconductor technology. The first and main requirement is that the silicon used for electronic devices must be extremely pure, since very small amounts of some impurities have a strong influence on the electronic characteristics of silicon, and therefore the performance of the electronic device. The second requirement is for large-diameter crystals, since the chip yield per wafer increases substantially with larger diameters, as shown in Fig. 13.2 for the case of DRAM [13.3], one of the most common electronic devices. Besides the purity and the diameter, the cost of production and the specifications of the material, including the grown-in defect density and the resistive homogeneity, must meet current industrial demands.

In this chapter, current approaches to the preparation of silicon – converting the raw material into single-crystalline silicon (see Fig. 13.1) – are discussed.

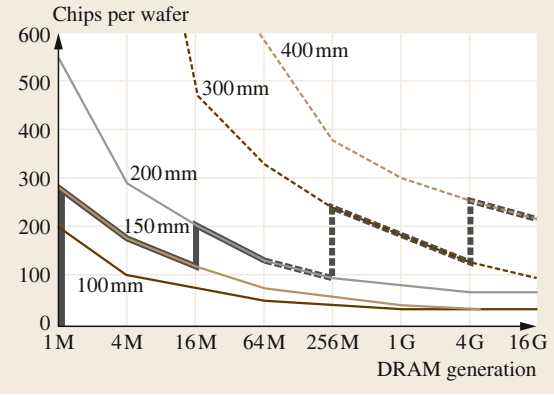


Fig. 13.2 Chips per wafer as a function of DRAM generation. (After [13.3])

13.2 Starting Materials

13.2.1 Metallurgical-Grade Silicon

The starting material for high-purity silicon single crystals is silica (SiO_2). The first step in silicon manufacture is the melting and reduction of silica. This is accomplished by mixing silica and carbon in the form of coal, coke or wood chips and heating the mixture to high temperatures in a submerged electrode arc furnace. This carbothermic reduction of silica produces fused silicon

$$SiO_2 + 2C \rightarrow Si + 2CO\,. \tag{13.1}$$

A complex series of reactions actually occur in the furnace at temperatures ranging from 1500 to 2000 °C. The lumps of silicon obtained from this process are called metallurgical-grade silicon (MG-Si), and its purity is about 98–99%.

13.2.2 Polycrystalline Silicon

Intermediate Chemical Compounds

The next step is to purify MG-Si to the level of semiconductor-grade silicon (SG-Si), which is used as the starting material for single-crystalline silicon. The basic concept is that powdered MG-Si is reacted with anhydrous HCl to form various chlorosilane compounds in a fluidized-bed reactor. Then the silanes are purified by distillation and chemical vapor deposition (CVD) to form SG-polysilicon.

A number of intermediate chemical compounds have been considered, such as monosilane (SiH_4), silicon tetrachloride ($SiCl_4$), trichlorosilane ($SiHCl_3$) and dichlorosilane (SiH_2Cl_2). Among these, trichlorosilane is most commonly used for subsequent polysilicon deposition for the following reasons:

1. It can be easily formed by the reaction of anhydrous hydrogen chloride with MG-Si at reasonably low temperatures (200–400 °C).
2. It is liquid at room temperature, so purification can be accomplished using standard distillation techniques.
3. It is easy to handle and can be stored in carbon steel tanks when dry.
4. Liquid trichlorosilane is easily vaporized and, when mixed with hydrogen, it can be transported in steel lines.
5. It can be reduced at atmospheric pressure in the presence of hydrogen.
6. Its deposition can take place on heated silicon, eliminating the need for contact with any foreign surfaces that may contaminate the resulting silicon.
7. It reacts at lower temperatures (1000–1200 °C) and at faster rates than silicon tetrachloride.

Hydrochlorination of Silicon

Trichlorosilane is synthesized by heating powdered MG-Si at around 300 °C in a fluidized-bed reactor. That is, MG-Si is converted into $SiHCl_3$ according to the following reaction

$$Si + 3HCl \rightarrow SiHCl_3 + H_2\,. \tag{13.2}$$

The reaction is highly exothermic and so heat must be removed to maximize the yield of trichlorosilane. While converting MG-Si into $SiHCl_3$, various impurities such as Fe, Al, and B are removed by converting them into their halides ($FeCl_3$, $AlCl_3$, and BCl_3, respectively), and byproducts such as $SiCl_4$ and H_2 are also produced.

Distillation and Decomposition of Trichlorosilane

Distillation has been widely used to purify trichlorosilane. The trichlorosilane, which has a low boiling point (31.8 °C), is fractionally distilled from the impure halides, resulting in greatly increased purity, with an electrically active impurity concentration of less than 1 ppba. The high-purity trichlorosilane is then vaporized, diluted with high-purity hydrogen, and introduced into the deposition reactor. In the reactor, thin silicon rods called slim rods supported by graphite electrodes are available for surface deposition of silicon according to the reaction

$$SiHCl_3 + H_2 \rightarrow Si + 3HCl\,. \tag{13.3}$$

In addition this reaction, the following reaction also occurs during polysilicon deposition, resulting in the formation of silicon tetrachloride (the major byproduct of the process)

$$HCl + SiHCl_3 \rightarrow SiCl_4 + H_2\,. \tag{13.4}$$

This silicon tetrachloride is used to produce high-purity quartz, for example.

Needless to say, the purity of the slim rods must be comparable to that of the deposited silicon. The slim rods are preheated to approximately 400 °C at the start of the silicon CVD process. This preheating is required in order to increase the conductivity of high-purity (high-resistance) slim rods sufficiently to allow for resistive heating. Depositing for 200–300 h at around 1100 °C results in high-purity polysilicon rods of 150–200 mm in diameter. The polysilicon rods are

shaped into various forms for subsequent crystal growth processes, such as chunks for Czochralski melt growth and long cylindrical rods for float-zone growth. The process for reducing trichlorosilane on a heated silicon rod using hydrogen was described in the late 1950s and early 1960s in a number of process patents assigned to Siemens; therefore, this process is often called the *Siemens method* [13.4].

The major disadvantages of the Siemens method are its poor silicon and chlorine conversion efficiencies, relatively small batch size, and high power consumption. The poor conversion efficiencies of silicon and chlorine are associated with the large volume of silicon tetrachloride produced as the byproduct in the CVD process. Only about 30% of the silicon provided in the CVD reaction is converted into high-purity polysilicon. Also, the cost of producing high-purity polysilicon may depend on the usefulness of the byproduct, $SiCl_4$.

Monosilane Process

A polysilicon production technology based on the production and pyrolysis of monosilane was established in the late 1960s. Monosilane potentially saves energy because it deposits polysilicon at a lower temperature and produces purer polysilicon than the trichlorosilane process; however, it has hardly been used due to the lack of an economical route to monosilane and due to processing problems in the deposition step [13.5]. However, with the recent development of economical routes to high-purity silane and the successful operation of a large-scale plant, this technology has attracted the attention of the semiconductor industry, which requires higher purity silicon.

In current industrial monosilane processes, magnesium and MG-Si powder are heated to 500 °C under a hydrogen atmosphere in order to synthesize magenesium silicide (Mg_2Si), which is then made to react with ammonium chloride (NH_4Cl) in liquid ammonia (NH_3) below 0 °C to form monosilane (SiH_4). High-purity polysilicon is then produced via the pyrolysis of the monosilane on resistively heated polysilicon filaments at 700–800 °C. In the monosilane generation process, most of the boron impurities are removed from silane via chemical reaction with NH_3. A boron content of 0.01–0.02 ppba in polysilicon has been achieved using a monosilane process. This concentration is very low compared to that observed in polysilicon prepared from trichlorosilane. Moreover, the resulting polysilicon is less contaminated with metals picked up through chemical transport processes because monosilane decomposition does not cause any corrosion problems.

Granular Polysilicon Deposition

A significantly different process, which uses the decomposition of monosilane in a fluidized-bed deposition reactor to produce free-flowing granular polysilicon, has been developed [13.5]. Tiny silicon seed particles are fluidized in a monosilane/hydrogen mix, and polysilicon is deposited to form free-flowing spherical particles that are an average of 700 μm in diameter with a size distribution of 100–1500 μm. The fluidized-bed seeds were originally made by grinding SG-Si in a ball or hammer mill and leaching the product with acid, hydrogen peroxide and water. This process was time-consuming and costly, and tended to introduce undesirable impurities into the system through the metal grinders. However, in a new method, large SG-Si particles are fired at each other by a high-speed stream of gas causing them to break into particles of a suitable size for the fluidized bed. This process introduces no foreign materials and requires no leaching.

Because of the greater surface area of granular polysilicon, fluidized-bed reactors are much more efficient than traditional Siemens-type rod reactors. The quality of fluidized-bed polysilicon has been shown to be equivalent to polysilicon produced by the more conventional Siemens method. Moreover, granular polysilicon of a free-flowing form and high bulk density enables crystal growers to obtain the most from of each production run. That is, in the Czochralski crystal growth process (see the following section), crucibles can be quickly and easily filled to uniform loadings which typically exceed those of randomly stacked polysilicon chunks produced by the Siemens method. If we also consider the potential of the technique to move from batch operation to continuous pulling (discussed later), we can see that free-flowing polysilicon granules could provide the advantageous route of a uniform feed into a steady-state melt. This product appears to be a revolutionary starting material of great promise for silicon crystal growth.

13.3 Single-Crystal Growth

Although various techniques have been utilized to convert polysilicon into single crystals of silicon, two techniques have dominated the production of them for electronics because they meet the requirements of the microelectronics device industry. One is a zone-melting method commonly called the *floating-zone*

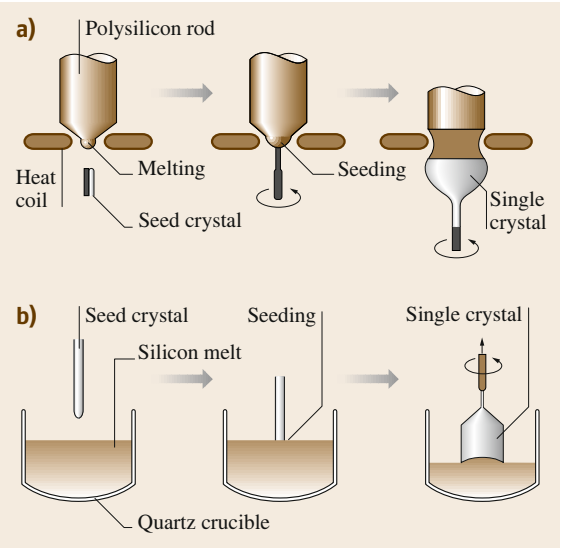


Fig. 13.3a,b Principles of single-crystal growth by **(a)** floating-zone method and **(b)** Czochralski method. (After [13.1])

(FZ) method, and the other is a pulling method traditionally called the *Czochralski (CZ) method*, although it should actually be called the *Teal–Little method*. The principles behind these two crystal growth methods are depicted in Fig. 13.3. In the FZ method, a molten zone is passed through a polysilicon rod to convert it into a single-crystal ingot; in the CZ method, a single crystal is grown by pulling from a melt contained in a quartz crucible. In both cases, the *seed crystal* plays a very important role in obtaining a single crystal with a desired crystallographic orientation.

It is estimated that about 95% of all single-crystal silicon is produced by the CZ method and the rest mainly by the FZ method. The silicon semiconductor industry requires high purity and minimum defect concentrations in their silicon crystals to optimize device manufacturing yield and operational performance. These requirements are becoming increasingly stringent as the technology changes from LSI to VLSI/ULSI and then SOC. Besides the quality or perfection of silicon crystals, crystal diameter has also been steadily increasing in order to meet the demands of device manufacturers. Since microelectronic chips are produced via a *batch system*, the diameters of the silicon wafers used for device fabrication significantly affect the productivity (as shown in Fig. 13.2), and in turn the production cost.

In the following sections, we first discuss the FZ method and then move on to the CZ method. The latter will be discussed in more detail due to its extreme importance to the microelectronics industry.

13.3.1 Floating-Zone Method

General Remarks

The FZ method originated from zone melting, which was used to refine binary alloys [13.6] and was invented by *Theuerer* [13.7]. The reactivity of liquid silicon with the material used for the crucible led to the development of the FZ method [13.8], which permits the crystallization of silicon without the need for any contact with the crucible material, which is needed to be able to grow crystals of the required semiconductor purity.

Outline of the Process

In the FZ process, a polysilicon rod is converted into a single-crystal ingot by passing a molten zone heated by a needle-eye coil from one end of the rod to the other, as shown in Fig. 13.3a. First, the tip of the polysilicon rod is contacted and fused with a seed crystal with the desired crystal orientation. This process is called *seeding*. The seeded molten zone is passed through the polysilicon rod by simultaneously moving the single crystal seed down the rod. When the molten zone of silicon solidifies, polysilicon is converted into single-crystalline silicon with the help of the seed crystal. As the zone travels along the polysilicon rod, single-crystal silicon freezes at its end and grows as an extension of the seed crystal.

After seeding, a thin neck about 2 or 3 mm in diameter and 10–20 mm long is formed. This process is called *necking*. Growing a neck eliminates dislocations that can be introduced into newly grown single-crystal silicon during the seeding operation due to thermal shock. This necking process, called the *Dash technique* [13.2], is therefore fundamental to growing dislocation-free crystals and is used universally in both the FZ and the CZ methods. An x-ray topograph of the seed, neck and conical part of a silicon single-crystal grown by the FZ method is shown in Fig. 13.4. It is apparent that dislocations generated at the melt contact are completely eliminated by necking. After the conical part is formed, the main body with the full target diameter is grown. During the entire FZ growth process, the shape of the molten zone and the ingot diameter are determined by adjusting the power to the coil and the travel rate, both of which are under computer control. The technique most commonly used to automatically control the diameter in both the FZ and CZ methods employs an infrared sensor focused on the meniscus. The shape of the meniscus on a growing crystal depends on its angle of contact at the three-phase boundary, the crystal diameter, and the magnitude of the surface tension. A change in meniscus angle (and therefore crystal diameter) is sensed, and the information is fed back in order to automatically adjust the growth conditions.

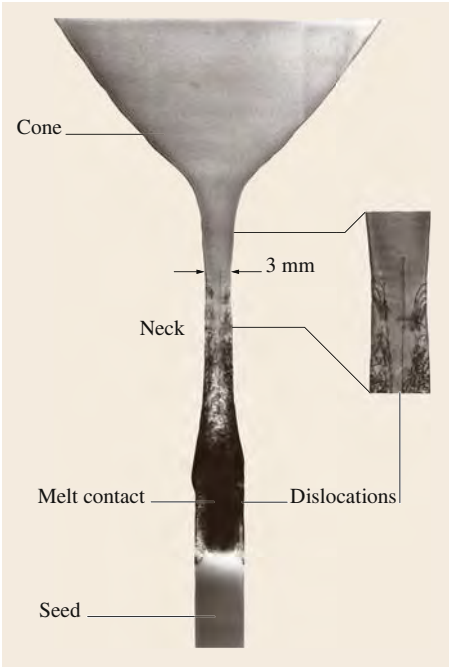


Fig. 13.4 X-ray topography of seed, neck and conical part of floating-zone silicon. (Courtesy of Dr. T. Abe)

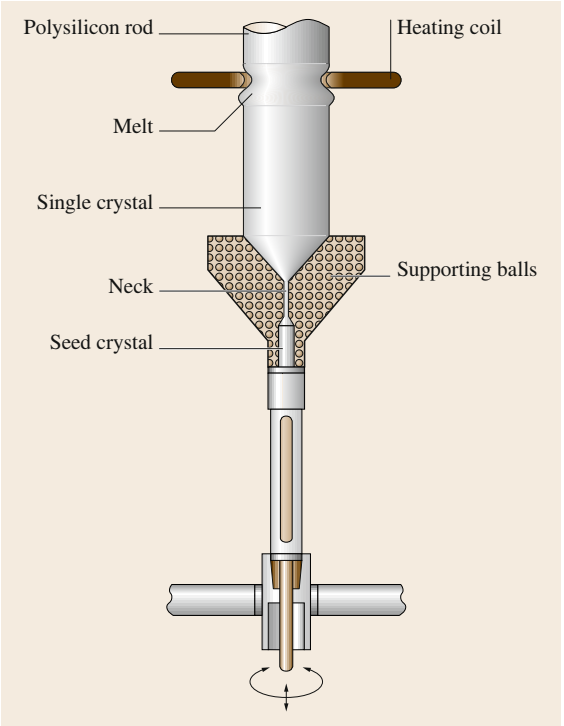


Fig. 13.5 Supporting system for floating-zone silicon crystal. (After [13.9])

In contrast with CZ crystal growth, in which the seed crystal is dipped into the silicon melt and the growing crystal is pulled upward, in the FZ method the thin seed crystal sustains the growing crystal, as does the polysilicon rod from the bottom (Fig. 13.3). As a result, the rod is balanced precariously on the thin seed and neck during the entire growth process. The seed and neck can support a crystal of up to a 20 kg so long as the center of gravity of the growing crystal remains at the center of the growth system. If the center of gravity moves away from the center line, the seed will easily fracture. Hence, it was necessary to invent a crystal stabilizing and supporting technique before long and heavy FZ silicon crystals could be grown. For large crystals, it is necessary to support the growing crystal in the way shown in Fig. 13.5 [13.9], particularly in the case of recent FZ crystals with large diameters (150–200 mm), since their weights easily exceed 20 kg.

Doping

In order to obtain n- or p-type silicon single-crystals of the required resistivity, either the polysilicon or the growing crystal must be doped with the appropriate donor or acceptor impurities, respectively. For FZ silicon growth, although several doping techniques have been tried, the crystals are typically doped by blowing a dopant gas such as phosphine (PH_3) for n-type silicon or diborane (B_2H_6) for p-type silicon onto the molten zone. The dopant gas is usually diluted with a carrier gas, such as argon. The great advantage of this method is that the silicon crystal manufacturer does not need to store polysilicon sources with different resistivities.

Since the segregation (discussed in the next subsection) of elemental dopants for n-type silicon is much less than unity, FZ crystals doped by the traditional method have radial dopant gradients. Moreover, since the crystallization rate varies in the radial direction on the microscopic scale, the dopant concentrations distribute cyclically and give rise to so-called *dopant striations*, resulting in radial resistivity inhomogeneities. In order to obtain more homogeneously doped n-type silicon, neutron transmutation doping (NTD) has been applied to FZ silicon crystals [13.10]. This procedure involves the nuclear transmutation of silicon to phosphorus by bombarding the crystal with thermal neutrons according to the reaction

$$^{30}\mathrm{Si}(\mathrm{n},\gamma) \rightarrow {}^{31}\mathrm{Si} \xrightarrow{2.6\,\mathrm{h}} {}^{31}\mathrm{P} + \beta\,. \qquad (13.5)$$

The radioactive isotope ^{31}Si is formed when ^{30}Si captures a neutron and then decays into the stable isotope ^{31}P (donor atoms), whose distribution is not dependent on crystal growth parameters. Immediately after irradiation the crystals exhibit high resistivity, which

is attributed to the large number of lattice defects arising from radiation damage. The irradiated crystal, therefore, must be annealed in an inert ambient at temperatures of around 700 °C in order to annihilate the defects and to restore the resistivity to that derived from the phosphorus doping. Under the NTD scheme, crystals are grown without doping and are then irradiated in a nuclear reactor with a large ratio of thermal to fast neutrons in order to enhance neutron capture and to minimize damage to the crystal lattice.

The application of NTD has been almost exclusively limited to FZ crystals because of their higher purity compared to CZ crystals. When the NTD technique was applied to CZ silicon crystals, it was found that oxygen donor formation during the annealing process after irradiation changed the resistivity from that expected, even though phosphorus donor homogeneity was achieved [13.11]. NTD has the additional shortcoming that no process is available for p-type dopants and that an excessively long period of irradiation is required for low resistivities (in the range of 1−10 Ω cm).

Properties of FZ-Silicon Crystal

During FZ crystal growth, the molten silicon does not come into contact with any substance other than the ambient gas in the growth chamber. Therefore, an FZ silicon crystal is inherently distinguished by its higher purity compared to a CZ crystal which is grown from the melt – involving contact with a quartz crucible. This contact gives rise to high oxygen impurity concentrations of around 10^{18} atoms/cm^3 in CZ crystals, while FZ silicon contains less than 10^{16} atoms/cm^3. This higher purity allows FZ silicon to achieve high resistivities not obtainable using CZ silicon. Most of the FZ silicon consumed has a resistivity of between 10 and 200 Ω cm, while CZ silicon is usually prepared to resistivities of 50 Ω cm or less due to the contamination from the quartz crucible. FZ silicon is therefore mainly used to fabricate semiconductor power devices that support reverse voltages in excess of 750−1000 V. The high-purity crystal growth and the precision doping characteristics of NTD FZ-Si have also led to its use in infrared detectors [13.12], for example.

However, if we consider mechanical strength, it has been recognized for many years that FZ silicon, which contains fewer oxygen impurities than CZ silicon, is mechanically weaker and more vulnerable to thermal stress during device fabrication [13.13, 14]. High-temperature processing of silicon wafers during electronic device manufacturing often produces enough thermal stress to generate slip dislocations and warpage. These effects bring about yield loss due to leaky junctions, dielectric defects, and reduced lifetime, as well as reduced photolithographic yields due to the degradation of wafer flatness. Loss of geometrical planarity due to warpage can be so severe that the wafers are not processed any further. Because of this, CZ silicon wafers have been used much more widely in IC device fabrication than FZ wafers have. This difference in mechanical stability against thermal stresses is the dominant reason why CZ silicon crystals are exclusively used for the fabrication of ICs that require a large number of thermal process steps.

In order to overcome these shortcomings of FZ silicon, the growth of FZ silicon crystals with doping impurities such as oxygen [13.15] and nitrogen [13.16] has been attempted. It was found that doping FZ silicon crystals with oxygen or nitrogen at concentrations of $1-1.5\times10^{17}$ atoms/cm^3 or 1.5×10^{15} atoms/cm^3, respectively, results in a remarkable increase in mechanical strength.

13.3.2 Czochralski Method

General Remarks

This method was named after J. Czochralski, who established a technique for determining the crystallization velocities of metals [13.17]. However, the actual pulling method that has been widely applied to single-crystal growth was developed by *Teal* and *Little* [13.18], who modified Czochralski's basic principle. They were the first to successfully grow single-crystals of germanium, 8 inches in length and 0.75 inches in diameter, in 1950. They subsequently designed another apparatus for the growth of silicon at higher temperatures. Although the basic production process for single-crystal silicon has changed little since it was pioneered by Teal and coworkers, large-diameter (up to 400 mm) silicon single-crystals with a high degree of perfection that meet state-of-the-art device demands have been grown by incorporating the Dash technique and successive technological innovations into the apparatus.

Today's research and development efforts concerning silicon crystals are directed toward achieving microscopic uniformity of crystal properties such as the resistivity and the concentrations of impurities and microdefects, as well as microscopic control of them, which will be discussed elsewhere in this Handbook.

Outline of the Process

The three most important steps in CZ crystal growth are shown schematically in Fig. 13.3b. In principle, the process of CZ growth is similar to that of FZ growth: (1) melting polysilicon, (2) seeding and (3) growing. The CZ pulling procedure, however, is more complicated than that of FZ growth and is distinguished from it by the use of a quartz crucible to contain the molten

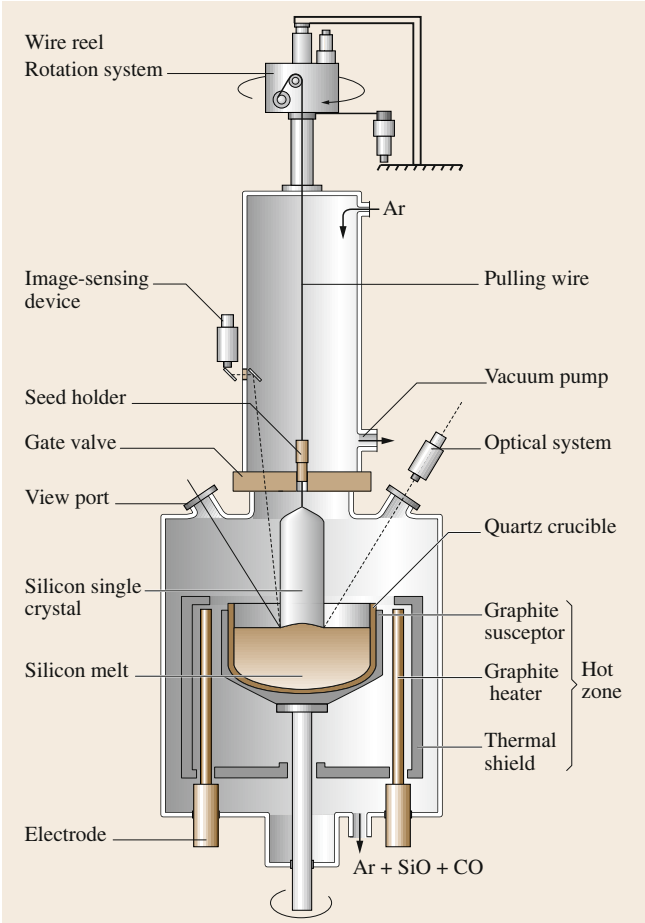


Fig. 13.6 Schematic view of typical Czochralski silicon crystal growing system. (After [13.1])

silicon. Figure 13.6 shows a schematic view of typical modern CZ crystal growth equipment. Important steps in the actual or standard CZ silicon crystal growth sequence are as follows:

1. Polysilicon chunks or grains are placed in a quartz crucible and melted at temperatures higher than the melting point of silicon (1420 °C) in an inert ambient gas.
2. The melt is kept at a high temperature for a while in order to ensure complete melting and ejection of tiny bubbles, which may cause voids or negative crystal defects, from the melt.
3. A seed crystal with the desired crystal orientation is dipped into the melt until it begins to melt itself. The seed is then withdrawn from the melt so that the neck is formed by gradually reducing the diameter; this is the most delicate step. During the entire crystal growth process, inert gas (usually argon) flows downward through the pulling chamber in order to carry off reaction products such as SiO and CO.
4. By gradually increasing the crystal diameter, the conical part and shoulder are grown. The diameter is increased up to the target diameter by decreasing the pulling rate and/or the melt temperature.
5. Finally, the cylindrical part of the body with a constant diameter is grown by controlling the pulling rate and the melt temperature while compensating for the drop in the melt level as the crystal grows. The pulling rate is generally reduced toward the tail end of a growing crystal, mainly due to increasing heat radiation from the crucible wall as the melt level drops and exposes more crucible wall to the growing crystal. Near the end of the growth process, but before the crucible is completely drained of molten silicon, the crystal diameter must be gradually reduced to form an end-cone in order to minimize thermal shock, which can cause slip dislocations at the tail end. When the diameter becomes small enough, the crystal can be separated from the melt without the generation of dislocations.

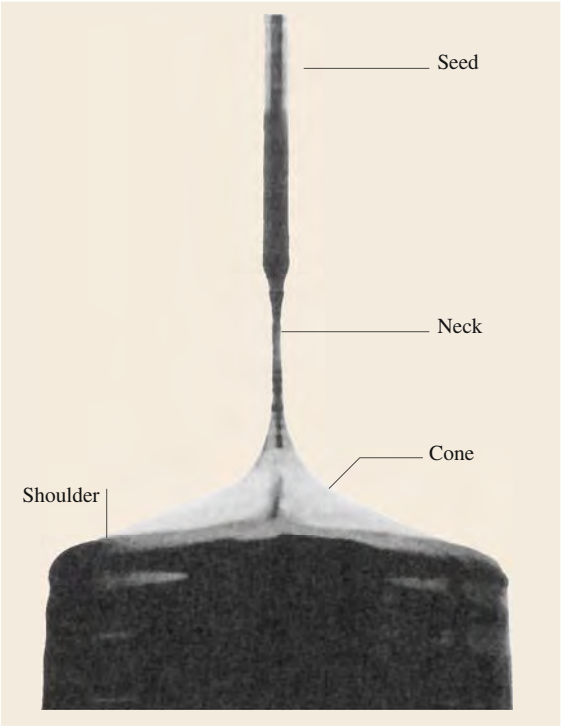


Fig. 13.7 Seed-end part of as-grown Czochralski silicon crystal

Figure 13.7 shows the seed-end part of an as-grown CZ silicon crystal. Although a seed-corn, which is the

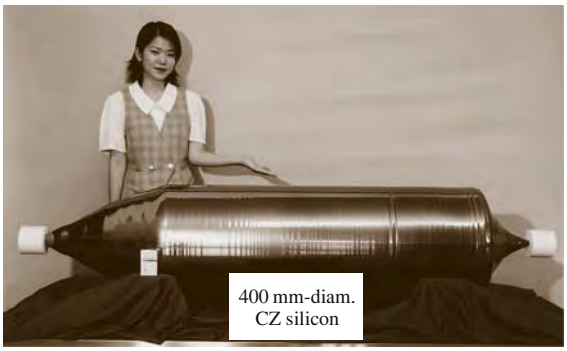


Fig. 13.8 Extra-large as-grown Czochralski silicon ingot 400 mm in diameter and 1800 mm in length. (Courtesy of Super Silicon Crystal Research Institute Corporation, Japan)

transition region from the seed to the cylindrical part, is usually formed to be rather flat for economic reasons, a more tapered shape might be desirable from a crystal quality point of view. The shoulder part and its vicinity should not be used for device fabrication because this part is considered a transition region in many senses and it exhibits inhomogeneous crystal characteristics due to the abrupt change in growth conditions.

Figure 13.8 shows an extra-large as-grown CZ silicon crystal ingot 400 mm in diameter and 1800 mm in length grown by the Super Silicon Crystal Research Institute Corporation in Japan [13.3].

Influence of Spatial Location in a Grown Crystal

As Fig. 13.9 clearly depicts, each portion of a CZ crystal is grown at a different time with different growth conditions [13.19]. Thus, it is important to understand that each portion has a different set of crystal characteristics and a different thermal history due to its different position along the crystal length. For example, the seed-end portion has a longer thermal history, ranging from the melting point of 1420 to around 400 °C in a puller, while the tail-end portion has a shorter history and is cooled down rather rapidly from the melting point. Ultimately, each silicon wafer prepared from a different portion of a grown crystal could exhibit different physico-chemical characteristics depending on its location in the ingot. In fact, it has been reported that the oxygen precipitation behavior exhibits the greatest location dependence, which, in turn, affects the generation of bulk defects [13.20].

Also, a nonuniform distribution of both crystal defects and impurities occurs across the transverse section of a flat wafer prepared from a CZ crystal silicon melt crystallized or solidified successively at the crystal–melt interface, which is generally curved in the CZ

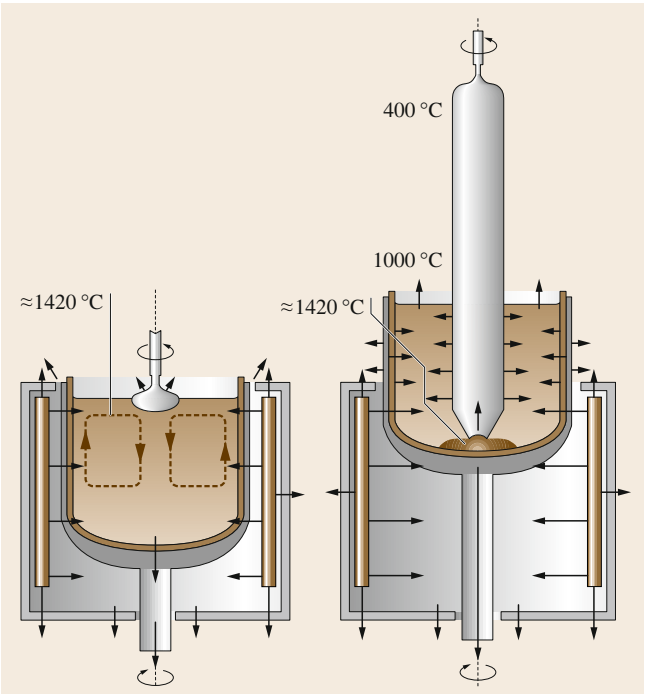


Fig. 13.9 Thermal environment during Czochralski crystal growth at initial and final stages. *Arrows* indicate approximate directions of heat flow. (After [13.19])

crystal growth process. Such inhomogeneities can be observed as *striations*, which are discussed later.

13.3.3 Impurities in Czochralski Silicon

The properties of the silicon semiconductors used in electronic devices are very sensitive to impurities. Because of this sensitivity, the electrical/electronic properties of silicon can be precisely controlled by adding a small amount of dopant. In addition to this dopant sensitivity, contamination by impurities (particularly transition metals) negatively affects the properties of silicon and results in the serious degradation of device performance. Moreover, oxygen is incorporated at levels of tens of atoms per million into CZ silicon crystals due to the reaction between the silicon melt and the quartz crucible. Regardless of how much oxygen is in the crystal, the characteristics of silicon crystals are greatly affected by the concentration and the behavior of oxygen [13.21]. In addition, carbon is also incorporated into CZ silicon crystals either from polysilicon raw materials or during the growth process, due to the graphite parts used in the CZ pulling equipment. Although the concentration of carbon in commercial CZ silicon crystals is normally less than 0.1 ppma, carbon is an impurity that greatly affects the behavior

Fig. 13.10 Growth striations, revealed by chemical etching, in a shoulder of Czochralski silicon

of oxygen [13.22, 23]. Also, nitrogen-doped CZ silicon crystals [13.24, 25] have recently attracted much attention due to their high microscopic crystal quality, which may meet the requirements for state-of-the-art electronic devices [13.26, 27].

Impurity Inhomogeneity

During crystallization from a melt, various impurities (including dopants) contained in the melt are incorporated into the growing crystal. The impurity concentration of the solid phase generally differs from that of the liquid phase due to a phenomenon known as *segregation*.

Segregation. The equilibrium segregation behavior associated with the solidification of multicomponent systems can be determined from the corresponding phase diagram of a binary system with a *solute* (the impurity) and a *solvent* (the host material) as components.

The ratio of the solubility of impurity A in solid silicon $[C_A]_s$ to that in liquid silicon $[C_A]_L$

$$k_0 = \frac{[C_A]_s}{[C_A]_L} \qquad (13.6)$$

is referred to as the *equilibrium segregation coefficient*. The impurity solubility in liquid silicon is always higher than that in solid silicon; that is, $k_0 < 1$.

The equilibrium segregation coefficient k_0 is only applicable to solidification at negligibly slow growth rates. For finite or higher solidification rates, impurity atoms with $k_0 < 1$ are rejected by the advancing solid at a greater rate than they can diffuse into the melt. In the CZ crystal growth process, segregation takes place at the start of solidification at a given seed–melt interface, and the rejected impurity atoms begin to accumulate in the melt layer near the growth interface and diffuse in the direction of the bulk of the melt. In this situation, an *effective segregation coefficient* k_{eff} can be defined at any moment during CZ crystal growth, and the impurity concentration $[C]_s$ in a CZ crystal can be derived by

$$[C]_s = k_{eff}[C_0](1-g)^{k_{eff}-1}\,, \qquad (13.7)$$

where $[C_0]$ is the initial impurity concentration in the melt and g is the fraction solidified.

Consequently, it is clear that a macroscopic longitudinal variation in the impurity level, which causes a variation in resistivity due to the variation in the dopant concentration, is inherent to the CZ batch growth process; this is due to the segregation phenomenon. Moreover, the longitudinal distribution of impurities is influenced by changes in the magnitude and the nature of melt convection that occur as the melt aspect ratio is decreased during crystal growth.

Striations. In most crystal growth processes, there are transients in the parameters such as instantaneous microscopic growth rate and the diffusion boundary layer thickness which result in variations in the effective segregation coefficient k_{eff}. These variations give rise to microscopic compositional inhomogeneities in the form of *striations* parallel to the crystal–melt interface. Striations can be easily delineated with several techniques, such as preferential chemical etching and x-ray topography. Figure 13.10 shows the striations revealed by chemical etching in the shoulder part of a longitudinal cross-section of a CZ silicon crystal. The gradual change in the shape of the growth interface is also clearly observed.

Striations are physically caused by the segregation of impurities and also point defects; however, the striations are practically caused by temperature fluctuations near the crystal–melt interface, induced by unstable thermal convection in the melt and crystal rotation in an asymmetric thermal environment. In addition, mechanical vibrations due to poor pulling control mechanisms in the growth equipment can also cause temperature fluctuations.

Figure 13.11 schematically illustrates a CZ-grown crystal cross-section containing a curved crystal–melt interface, which results in inhomogeneities on the surface of a slice. As each planar wafer is sliced, it contains different portions of several curved striations. Different *phonograph rings*, referred to as *swirl*, can then occur in each wafer, which can be observed across the wafer using the techniques mentioned above.

Doping

In order to obtain the desired resistivity, a certain amount of dopant (either donor or acceptor atoms) is added to a silicon melt according to the resistivity–concentration relation. It is common practice to add dopants in the form of highly doped silicon particles or chunks of about 0.01 Ω cm resistivity, which are called the dopant fixture, since the amount of pure dopant needed is unmanageably small, except for heavily doped silicon materials (n^+ or p^+ silicon).

The criteria for selecting a dopant for a semiconductor material are that it has the following properties:

1. Suitable energy levels
2. High solubility
3. Suitable or low diffusivity
4. Low vapor pressure.

A high diffusivity or high vapor pressure leads to undesirable diffusion or vaporization of dopants, which results in unstable device operation and difficulties in achieving precise resistivity control. A solubility that is too small limits the resistivity that can be obtained. In addition to those criteria, the chemical properties (the toxicity for example) must be considered. A further consideration from the viewpoint of crystal growth is that the dopant has a segregation coefficient that is close to unity in order to make the resistivity as uniform as possible from the seed-end to the tail-end of the CZ crystal ingot. Consequently, phosphorus (P) and boron (B) are the most commonly used donor and acceptor dopants for silicon, respectively. For n^+ silicon, in which donor atoms are heavily doped, antimony (Sb) is usually used instead of phosphorus because of its smaller diffusivity, in spite of its small segregation coefficient and high vapor pressure, which lead to large variations in concentration in both the axial and the radial directions.

Oxygen and Carbon

As shown schematically in Figs. 13.3b and 13.6, a quartz (SiO_2) crucible and graphite heating elements are used in the CZ-Si crystal growth method. The surface of the crucible that contacts the silicon melt is gradually dissolved due to the reaction

$$SiO_2 + Si \rightarrow 2SiO\,. \qquad (13.8)$$

This reaction enriches the silicon melt with oxygen. Most of the oxygen atoms evaporate from the melt surface as volatile silicon mono-oxide (SiO), but some of them are incorporated into a silicon crystal through the crystal–melt interface.

However, the carbon in CZ silicon crystals originates mainly from the polycrystalline starting material. Levels of carbon ranging from 0.1 to 1 ppma,

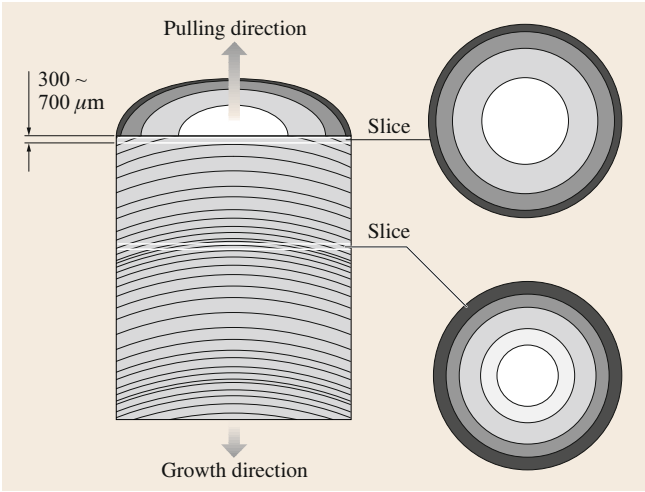


Fig. 13.11 Schematic illustration of Czochralski crystal cross-section containing a curved crystal–melt interface and planar wafers sliced into different portions. (After [13.1])

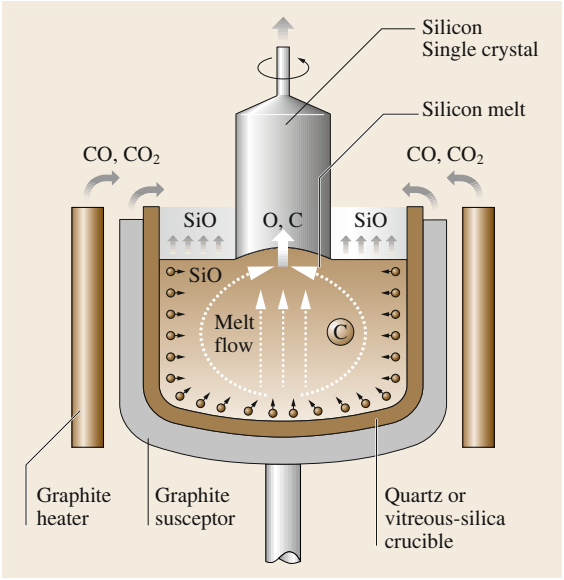


Fig. 13.12 Incorporation of oxygen and carbon into Czochralski silicon crystal. (After [13.1])

depending on the manufacturer, are found in the polysilicon. Sources of carbon in polysilicon are assumed to be mainly carbon-containing impurities found in the trichlorosilane used in the production of polysilicon. Graphite parts in CZ pulling equipment can also contribute to carbon contamination by reacting with oxygen, which is always present during the ambient growth. The resulting products of CO and CO_2 are dissolved into the silicon melt and account for the carbon impurities in silicon crystals. Thus, oxygen and

carbon are the two major nondoping impurities that are incorporated into CZ silicon crystals in the way shown schematically in Fig. 13.12. The behavior of these impurities in silicon, which affect a number of the properties of CZ silicon crystals, has been the subject of intensive study since the late 1950s [13.21].

13.4 New Crystal Growth Methods

Silicon crystals used for microelectronic device fabrication must meet a variety of requirements set by device manufacturers. In addition to the requirements for silicon *wafers*, the following crystallographic demands have become more common due to high-yield and high-performance microelectronic device manufacturing:

1. Large diameter
2. Low or controlled defect density
3. Uniform and low radial resistivity gradient
4. Optimum initial oxygen concentration and its precipitation.

It is clear that silicon crystal manufacturers must not only meet the above requirements but also produce those crystals economically and with high manufacturing yields. The main concerns of silicon crystal growers are the crystallographic perfection and the axial distribution of dopants in CZ silicon. In order to overcome some problems with the conventional CZ crystal growth method, several new crystal growth methods have been developed.

13.4.1 Czochralski Growth with an Applied Magnetic Field (MCZ)

The melt convection flow in the crucible strongly affects the crystal quality of CZ silicon. In particular, unfavorable growth striations are induced by unsteady melt convection resulting in temperature fluctuations at the growth interface. The ability of a magnetic field to inhibit thermal convection in electrically conducting fluid was first applied to the crystal growth of indium antimonide via the horizontal boat technique [13.28] and the horizontal zone-melting technique [13.29]. Through these investigations, it was confirmed that a magnetic field of sufficient strength can suppress the temperature fluctuations that accompany melt convection, and can dramatically reduce growth striations.

The effect of the magnetic field on growth striations is explained by its ability to decrease the turbulent thermal convection of a melt and in turn decrease the temperature fluctuations at the crystal–melt interface. The fluid flow damping caused by the magnetic field is due to the induced magnetomotive force when the flow is orthogonal to the magnetic flux lines, which results in an increase in the effective kinematic viscosity of the conducting melt.

Silicon crystal growth by the magnetic field applied CZ (MCZ) method was reported for the first time in 1980 [13.30]. Originally MCZ was intended for the growth of CZ silicon crystals that contain low oxygen concentrations and therefore have high resistivities with low radial variations. In other words, MCZ silicon was expected to replace the FZ silicon almost exclusively used for power device fabrication. Since then, various magnetic field configurations, in terms of the magnetic field direction (horizontal or vertical) and the type of magnets used (normal conductive or superconductive), have been developed [13.31]. MCZ silicon produced with a wide range of desired oxygen concentrations (from low to high) has been of great interest for different device applications. The value of MCZ silicon lies in its high quality and its ability to control the oxygen concentration over a wide range, which cannot be achieved using the conventional CZ method [13.32], as well as its enhanced growth rate [13.33].

As far as the crystal quality is concerned, there is no doubt that the MCZ method provides the silicon crystals most favorable to the semiconductor device industry. The production cost of MCZ silicon may be higher than that of conventional CZ silicon because the MCZ method consumes more electrical power and re-

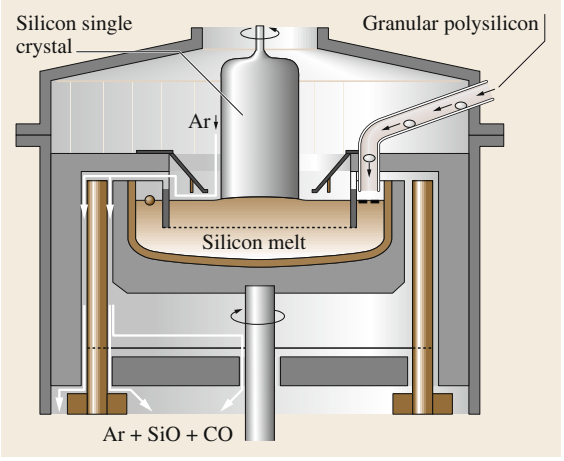


Fig. 13.13 Schematic illustration of the continuous-charging Czochralski method. (After [13.34])

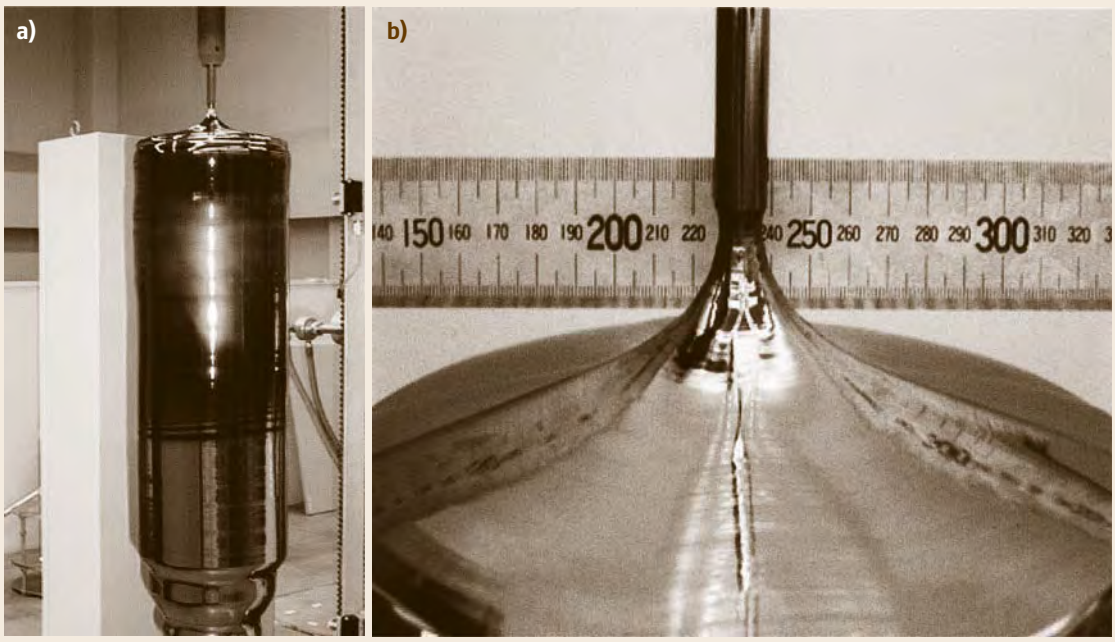


Fig. 13.14a,b 200 mm-diameter dislocation-free Czochralski silicon crystal grown without the Dash necking process. **(a)** Whole body, **(b)** seed and cone. (Courtesy of Prof. K. Hoshikawa)

quires additional equipment and operating space for the electromagnets; however, taking into account the higher growth rate of MCZ, and when superconductive magnets that need smaller space and consume less electrical power compared with conductive magnets are used, the production cost of MCZ silicon crystals may become comparable to that of conventional CZ silicon crystals. In addition, the improved crystal quality of MCZ silicon may increase production yields and lower the production cost.

13.4.2 Continuous Czochralski Method (CCZ)

Crystal production costs depend to a large extent on the cost of materials, in particular the cost of those used for quartz crucibles. In the conventional CZ process, called a *batch process*, a crystal is pulled from a single crucible charge, and the quartz crucible is used only once and is then discarded. This is because the small amount of remaining silicon cracks the crucible as it cools from a high temperature during each growth run.

One strategy for replenishing a quartz crucible with melt economically is to continuously add feed as the crystal is grown and thereby maintain the melt at a constant volume. In addition to saving crucible costs, the continuous-charging Czochralski (CCZ) method provides an ideal environment for silicon crystal growth. As already mentioned, many of the inhomogeneities in crystals grown by the conventional CZ batch process are a direct result of the unsteady kinetics arising from the change in melt volume during crystal growth. The CCZ method aims not only to reduce production costs but also to grow crystals under steady conditions. By maintaining the melt volume at a constant level, steady thermal and melt flow conditions can be achieved (see Fig. 13.9, which shows the change in thermal environments during conventional CZ growth).

Continuous charging is commonly performed by polysilicon feeding, as shown in Fig. 13.13 [13.34]. This system consists of a hopper for storing the polysilicon raw material and a vibratory feeder that transfers the polysilicon to the crucible. In the crucible that contains the silicon melt, a quartz baffle is required to prevent the melt turbulance caused by feeding in the solid material around the growth interface. Free-flowing polysilicon granules such as those mentioned previously are obviously advantageous for the CCZ method.

The CCZ method certainly solves most of the problems related to inhomogeneities in crystal grown by the conventional CZ method. Moreover, the combination of MCZ and CCZ (the magnetic-field-applied continuous CZ (MCCZ) method) is expected to provide the ultimate crystal growth method, giving ideal silicon crystals for a wide variety of microelectronic applications [13.1]. Indeed, it has been used to grow

high-quality silicon crystals intended for microelectronic devices [13.35].

However, it should be emphasized that the different thermal histories of different parts of the crystal (from the seed to the tail ends, as shown in Fig. 13.9) must be considered even when the crystal is grown by the ideal growth method. In order to homogenize the grown crystal or to obtain axial uniformity in the thermal history, some form of post-treatment, such as high-temperature annealing [13.36], is required for the crystal.

13.4.3 Neckingless Growth Method

As mentioned previously, Dash's necking process (which grows a thin neck 3–5 mm in diameter, Fig. 13.7) is a critical step during CZ crystal growth because it eliminates grown-in dislocations. This technique has been the industry standard for more than 40 years. However, recent demands for large crystal diameters (> 300 mm, weighing over 300 kg) have resulted in the need for larger diameter necks that do not introduce dislocations into the growing crystal, since a thin neck 3–5 mm in diameter can not support such large crystals.

Large diameter seeds that are typically 170 mm long, with a minimum diameter of > 10 mm and an average of 12 mm grown from silicon melt heavily doped with boron (> 10^{19} atoms/cm^3) have been used to grow dislocation-free 200 mm-diameter CZ silicon crystals [13.37, 38]. It is estimated that large diameter necks 12 mm in diameter can CZ support crystals as heavy as 2000 kg [13.39]. Figure 13.14a shows a 200 mm-diameter dislocation-free CZ silicon crystal grown without the Dash necking process, and Fig. 13.14b shows its enlarged seed (compare with Fig. 13.7). The mechanism by which dislocations are not incorporated into the growing crystal has been primarily attributed to hardening effect of the heavy doping of boron in the silicon.

References

13.1 F. Shimura: *Semiconductor Silicon Crystal Technology* (Academic, New York 1988)
13.2 W.C. Dash: J. Appl. Phys. **29**, 736 (1958)
13.3 K. Takada, H. Yamagishi, H. Minami, M. Imai: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1998) p. 376
13.4 J.R. McCormic: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1986) p. 43
13.5 P.A. Taylor: Solid State Technol. **July**, 53 (1987)
13.6 W.G. Pfann: Trans. Am. Inst. Min. Metall. Eng. **194**, 747 (1952)
13.7 C.H. Theuerer: U.S. Patent 3060123 (1962)
13.8 P.H. Keck, M.J.E. Golay: Phys. Rev. **89**, 1297 (1953)
13.9 W. Keller, A. Mühlbauer: *FLoating-Zone Silicon* (Marcel Dekker, New York 1981)
13.10 J.M. Meese: *Neutron Transmutation Doping in Semiconductors* (Plenum, New York 1979)
13.11 H.M. Liaw, C.J. Varker: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1977) p. 116
13.12 E.L. Kern, L.S. Yaggy, J.A. Barker: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1977) p. 52
13.13 S.M. Hu: Appl. Phys. Lett. **31**, 53 (1977)
13.14 K. Sumino, H. Harada, I. Yonenaga: Jpn. J. Appl. Phys. **19**, L49 (1980)
13.15 K. Sumino, I. Yonenaga, A. Yusa: Jpn. J. Appl. Phys. **19**, L763 (1980)
13.16 T. Abe, K. Kikuchi, S. Shirai: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1981) p. 54
13.17 J. Czochralski: Z. Phys. Chem. **92**, 219 (1918)
13.18 G.K. Teal, J.B. Little: Phys. Rev. **78**, 647 (1950)
13.19 W. Zulehner, D. Huber: In: *Crystals 8: Silicon, Chemical Etching* (Springer, Berlin, Heidelberg 1982) p. 1
13.20 H. Tsuya, F. Shimura, K. Ogawa, T. Kawamura: J. Electrochem. Soc. **129**, 374 (1982)
13.21 F. Shimura (Ed.): *Oxygen in Silicon* (Academic, New York 1994)
13.22 S. Kishino, Y. Matsushita, M. Kanamori: Appl. Phys. Lett. **35**, 213 (1979)
13.23 F. Shimura: J. Appl. Phys. **59**, 3251 (1986)
13.24 H.D. Chiou, J. Moody, R. Sandfort, F. Shimura: VLSI science technology, Proc. 2nd Int. Symp. Very Large Scale Integr. (The Electrochemical Society, Pennington 1984) p. 208
13.25 F. Shimura, R.S. Hocket: Appl. Phys. Lett. **48**, 224 (1986)
13.26 A. Huber, M. Kapser, J. Grabmeier, U. Lambert, W. v. Ammon, R. Pech: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 2002) p. 280
13.27 G.A. Rozgonyi: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 2002) p. 149
13.28 H.P. Utech, M.C. Flemings: J. Appl. Phys. **37**, 2021 (1966)
13.29 H.A. Chedzey, D.T. Hurtle: Nature **210**, 933 (1966)
13.30 K. Hoshi, T. Suzuki, Y. Okubo, N. Isawa: Ext. Abstr. Electrochem. Soc. 157th Meet. (The Electrochemical Society, Pennington 1980) p. 811
13.31 M. Ohwa, T. Higuchi, E. Toji, M. Watanabe, K. Homma, S. Takasu: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1986) p. 117
13.32 M. Futagami, K. Hoshi, N. Isawa, T. Suzuki, Y. Okubo, Y. Kato, Y. Okamoto: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington

1986) p. 939
13.33 T. Suzuki, N. Isawa, K. Hoshi, Y. Kato, Y. Okubo: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1986) p. 142
13.34 W. Zulehner: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1990) p. 30
13.35 Y. Arai, M. Kida, N. Ono, K. Abe, N. Machida, H. Futuya, K. Sahira: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1994) p. 180
13.36 F. Shimura: In: *VLSI Science and Technology* (The Electrochemical Society, Pennington 1982) p. 17
13.37 S. Chandrasekhar, K.M. Kim: In: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1998) p. 411
13.38 K. Hoshikawa, X. Huang, T. Taishi, T. Kajigaya, T. Iino: Jpn. J. Appl. Phys. **38**, L1369 (1999)
13.39 K.M. Kim, P. Smetana: J. Cryst. Growth **100**, 527 (1989)

14. Epitaxial Crystal Growth: Methods and Materials

Peter Capper, Stuart Irvine, Tim Joyce

Epitaxial growth of thin films of material for a wide range of applications in electronics, optoelectronics, and magneto-optics is a critical activity in many industries. The original technique, in most instances, was liquid-phase epitaxy (LPE) as this was the simplest and often the cheapest route to producing device-quality layers. While some production processes are still based on LPE, most of the research activities and, increasingly, much of the production of electronic and optoelectronic devices is now centered on metal organic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE). These latter techniques are more versatile, although the equipment is more expensive, and can readily produce multilayer structures with atomic-layer control, which is becoming more and more fundamental to the nanoscale engineering being called upon now to produce device structures in as-grown multilayers. This chapter covers these three basic techniques, including some of their more common variants, and outlines the relative advantages and disadvantages of each of them. Some examples of growth in various materials systems of importance are also outlined for each of the three techniques.

14.1 **Liquid-Phase Epitaxy (LPE)** 309
14.1.1 Introduction and Background 309
14.1.2 History and Status 310
14.1.3 Characteristics 310
14.1.4 Apparatus and Techniques 311
14.1.5 Group IVs 313
14.1.6 Group III–Vs 314
14.1.7 Group II–VIs 316
14.1.8 Garnets 318
14.1.9 Oxides/Fluorides 318
14.1.10 Atomically Flat Surfaces 318
14.1.11 Conclusions 319
14.2 **Metal Organic Chemical Vapor Deposition** 319
14.2.1 Introduction and Background 319
14.2.2 Basic Reaction Kinetics 319
14.2.3 Precursors 321
14.2.4 Reactor Cells 322
14.2.5 III–V MOCVD 324
14.2.6 II–VI MOCVD 327
14.2.7 Conclusions 329
14.3 **Molecular Beam Epitaxy (MBE)** 329
14.3.1 Introduction and Background 329
14.3.2 Reaction Mechanisms 330
14.3.3 MBE Growth Systems 331
14.3.4 Gas Sources in MBE 333
14.3.5 Growth of III–V Materials by MBE 334
14.3.6 Conclusions 337
References 337

This chapter outlines the three major epitaxial growth processes used to produce layers of material for electronic, optical, optoelectronic and magneto-optical applications. These are liquid-phase epitaxy (LPE), metal organic chemical vapor deposition (MOCVD), and molecular beam epitaxy (MBE) and their main variants. All three techniques have their relevant advantages and disadvantages for particular material systems and these will be highlighted where appropriate in the following sections.

14.1 Liquid-Phase Epitaxy (LPE)

14.1.1 Introduction and Background

Liquid-phase epitaxy (LPE) is a mature technology and has unique features that make it still applicable for use in niche applications within certain device technologies. It has given way in many areas, however, to various vapor-phase epitaxy techniques, for example, metalorganic vapor phase, molecular beam,

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_14

atomic layer, etc., epitaxies (MOVPE, MBE, ALE, etc.); see Sects. 14.2 and 14.3. When selecting an epitaxial growth technology for a particular material system and/or device application, the choice needs to take into account the basic principles of thermodynamics, kinetics, surface energies, etc., as well as practical issues of reproducibility, scalability, process control, instrumentation, safety, and capital equipment costs. A systematic comparison of the various epitaxy techniques suggests that no single technique can best satisfy the needs of all materials/devices combinations needed in microelectronics, optoelectronics, solar cells, thermophotovoltaics, thermoelectrics, semiconductor electrochemical devices, magnetic/magneto-optical devices, and microelectromechanical systems. LPE is still a good choice for many of these application areas [14.1, Chap. 10], [14.2, Chap. 14].

Scheel ([14.3, Chap. 17] and Foreword to [14.2]) has discussed the issues surrounding the use of LPE to produce compound semiconductors for III–V light-emitting diodes (LEDs), high-power devices in SiC and GaN, where device performance may be higher than in material grown by other techniques. He notes in [14.2] that LPE has lost its dominance but growth of near-perfect layers with extremely flat surfaces in homoepitaxy and near-lattice-matched heteroepitaxy is still being employed in several important niche areas, for example, HgCdTe (MCT), garnets, GaN, etc.

14.1.2 History and Status

LPE is basically a high-temperature solution growth technique [14.4] in which a thin layer of the required material is deposited onto a suitable substrate. Homoepitaxy is defined as growth of a layer of the same composition as the substrate, whereas heteroepitaxy is the growth of a layer of markedly different composition. A suitable substrate material would have the same crystal structure as the layer, have as close a match in lattice parameter as possible, and be chemically compatible with the solution and the layer. *Nelson* [14.5] is commonly thought to have developed the first LPE systems, in this case for producing multilayer compound semiconducting structures. In the following decades, a large technology base was established for III–V compound semiconductor lasers, LEDs, photodiodes, and solar cells. LPE has been applied to the growth of a wide range of semiconductors, for example, Si, Ge, SiGe alloys, SiC, GaAs, InP, GaP, GaSb, InAs, InSb (and their ternary and quaternary alloys), GaN, ZnSe, MCT, HgZnTe, and HgMnTe. It has also been used to produce a diverse range of oxide/fluoride compounds, such as high-temperature superconductors, garnets, para- and ferroelectrics, and for various other crystals for optical and magnetic applications. Dipping LPE is still in use to make magneto-optical isolators by epitaxially growing garnet layers on gadolinium gallium garnet substrates [14.2, Chap. 14].

It is probably true to say that most of these systems were first studied using LPE, where it was used in the demonstration, development, and commercialization of many device types, including GaAs solar cells, III–V LEDs and laser diodes, GaAs-based Gunn-effect and other microwave devices, and various IR detectors based on InSb and on MCT. Nevertheless, LPE does not appear in the research literature as often as MOVPE, MBE, ALE, etc., on work in these systems. However, it is still used in some industrial applications, including III–V LEDs, particularly those based on AlGaAs and GaP alloys, and IR detectors based on MCT. It often produces the highest sensitivity in devices, mainly due to it being a near-equilibrium process that can produce near-perfect layers.

Realistic industrial production data is difficult to obtain, but *Moon* [14.6] noted that the large majority of optoelectronic devices were still being grown by LPE at that time (1997), amounting to $\approx 4000\,\mathrm{m}^2$ per year. He estimated that despite the loss of market share to more advanced techniques, the total demand for LPE material was still increasing at $\approx 10\%$ per year. LPE was discontinued for many applications because of its perceived limitations with regard to control of layer thickness, alloy compositions, doping, interface smoothness, and difficulties in growing certain combinations of interest for heterostructure devices. LPE is normally dismissed for the production of superlattices, quantum wells, strained-layer structures, and heterojunctions with large lattice mismatch of chemical dissimilarities. It also suffers from a reputation for poor reproducibility, problems in scaling up in size or throughput, and difficulties in achieving abrupt interfaces between successive layers within structures.

14.1.3 Characteristics

LPE is characterized as a near-equilibrium-grown process, when compared to the various vapor-phase epitaxy techniques. Heat and mass transport, surface energies, interface kinetics, and growth mechanisms are different in LPE compared from those in vapor-phase epitaxy or bulk-growth techniques. These features produce both advantages and disadvantages of LPE, the former include:

- High growth rates. These are typically $0.1-1\,\mu\mathrm{m/h}$, $\approx 10-100$ times faster than in MOVPE or MBE. This feature is useful when thick layers or *virtual substrates* are required.

- Favorable segregation of impurities into the liquid phase. This can lead to lower residual or background impurities in the epitaxial layer [14.2, Chap. 10].
- Ability to produce very flat surfaces and excellent structural perfection (see Fig. 14.1 and [14.7]).
- Wide selection of dopants. Most solid or liquid elements can be added to a melt and incorporated in the layer, unlike in vapor-phase growth where development of volatile dopant precursors with suitable kinetics and sticking coefficients is a major undertaking.
- Suppression of certain types of defects. In general, LPE material has lower point defects (vacancies, interstitials, antisites) than material made by other techniques. For example, the Ga-rich conditions during GaAs LPE inhibits the formation of the As antisite defect that is responsible for the nonradiative losses in luminescent devices.
- Once the relevant phase diagram is established, growth can be made to occur over a wide range of temperatures.
- Absence of highly toxic precursors or byproducts.
- Low capital equipment and operating costs. A research LPE kit can be constructed for under $50 000.

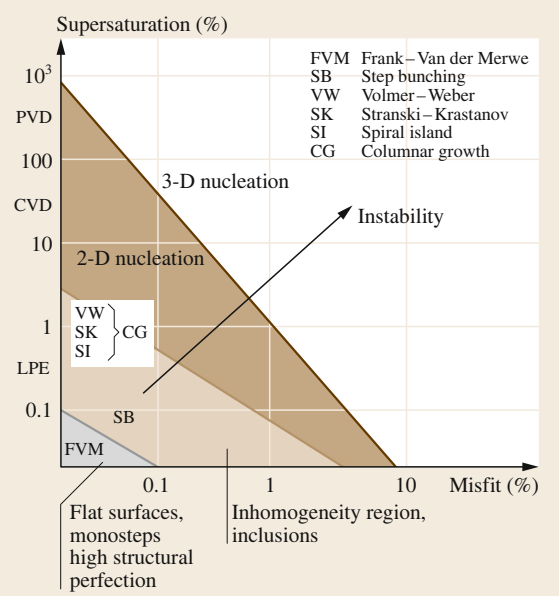


Fig. 14.1 The effects of both supersaturation (and the growth method) and misfit on the nucleation and growth regimes. Only at very low supersaturation in LPE using low-misfit substrates can really flat surfaces be expected. (After [14.7])

The main consideration in designing an LPE process is to determine accurate phase equilibria (S–L and/or S–L–V) of the required system [14.3, Chap. 2]. Solution modeling, extrapolations of existing phase equilibria, and semiempirical predictions are usually sufficient to guide developments in new systems/applications. The near-equilibrium nature of LPE provides for several important growth modes, for example, selective epitaxy (deposition through masks on a substrate) and epitaxial layer overgrowth (ELOG, where growth over a mask occurs), which are useful for defect reduction and new device structures.

14.1.4 Apparatus and Techniques

The basic requirement is to bring the substrate and growth solution into contact while the epitaxial layer grows, and then to separate them cleanly at the end of the growth cycle. The three main embodiments of the LPE growth method are tipping, dipping, and sliding boat (Fig. 14.2).

Figure 14.3 shows the tipping furnace system used by *Nelson* [14.5]. The boat, normally graphite or silica, sits in the work tube in the center of the tilted furnace such that the substrate is held at one end of the boat with the growth solution at the other end. Once the melt has been equilibrated, the temperature is slowly reduced and the furnace is tipped to roll the solution over the substrate. After a suitable time, the furnace is tipped back to the start position and the solution rolls off the

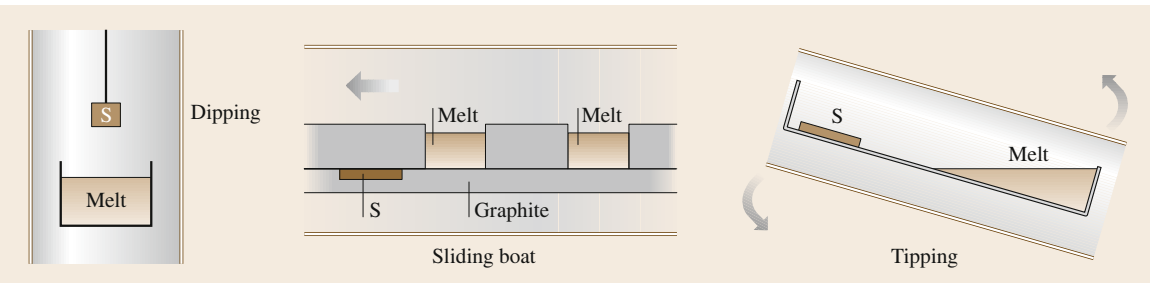


Fig. 14.2 Dipping, sliding boat, and tipping LPE arrangements. (After [14.8])

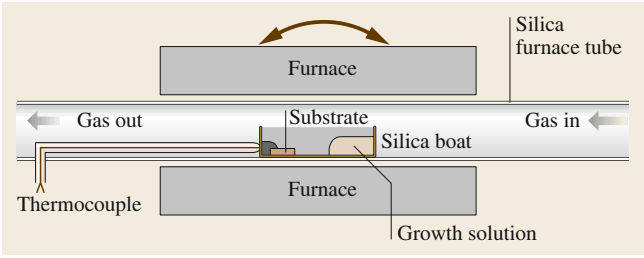


Fig. 14.3 Tipping LPE furnace. (After [14.5])

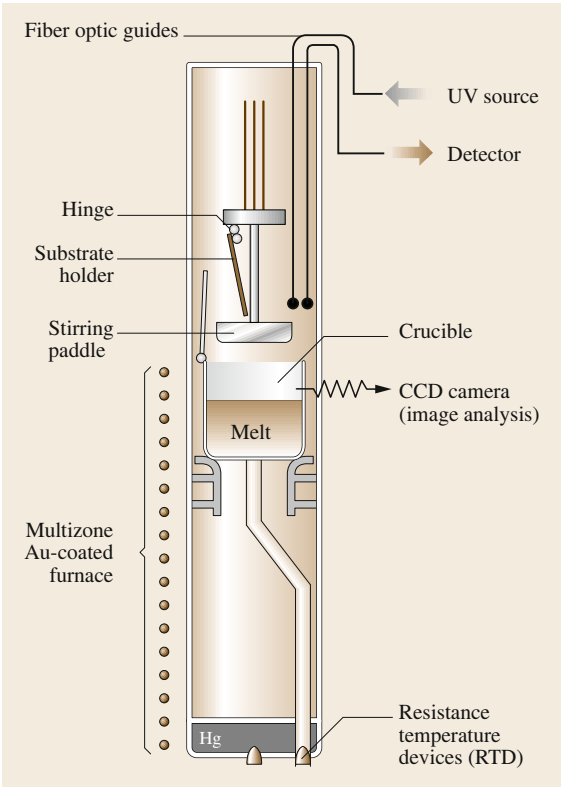


Fig. 14.4 Schematic diagram of a dipping LPE reactor showing the Te-rich melt, the mercury reservoir, and position of the sensors. (After [14.8])

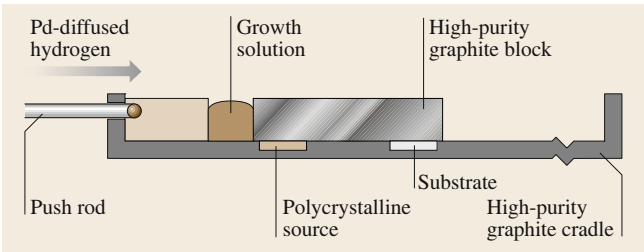


Fig. 14.5 Basic structure of graphite horizontal sliding boat LPE. (After [14.9])

grown layer. This is a relatively simple and cheap technique but has the limitations that solution removal is difficult and it is normally only suitable for single-layer growth.

Figure 14.4 shows the dipping system used for the growth of MCT [14.8]. The design and operation of a system for growth from Hg-rich melts is dominated by consideration of the high vapor pressure of Hg that comprises $\geq 90\%$ of the growth solution, which led to the evolution of a vertical high-pressure furnace design with a cooled reflux region. Such systems are capable of containing about 10–20 kg of melt at 550 °C for several years with no degradation in melt integrity or purity. The use of large melts results in a near-constant saturation temperature from run-to-run and ensures excellent reproducibility of layer characteristics.

Figure 14.5 shows the basic structure of a graphite sliding-boat system, which became the most popular and versatile of the three main methods [14.9]. The substrate sits in a recess in a slider supported by a base section. Growth solutions reside in wells in the upper section of the boat and can be repositioned over the substrate using a pushrod arrangement. The critical design feature of the boat to aid wipe-off is to control the gap between the top of the grown layer and the underside of the top section. If this is too large, melt retention occurs, but if it is too small the layer may be scratched. Multilayer growth is easily possible using this sliding-boat method, providing melt retention is kept to a minimum. The thin melts lead to suppression of thermal or solutal convection, and hence reduce enhanced edge growth. Scale-up has also been achieved in this method with several substrates (up to 16) growing three-layer GaAlAs structures in a single run.

In all the LPE methods, production of supersaturation in the growth solution drives the deposition of the layer on the substrate. This supersaturation can be produced by ramp cooling, step cooling, supercooled growth (a hybrid of the previous two techniques), two-phase growth, constant-temperature growth, or transient growth [14.10]. The choice between these various means will depend on the details of the particular material system and the precise requirement for the material. Another means of producing the required supersaturation is that of electroepitaxy (LPEE), in which an electric current is passed through the interface to stimulate layer growth. It is now thought that this occurs via an electromigration process rather than via Peltier cooling. Benefits claimed for the technique include reduced surface ripple, a reduction in certain microdefects, and an ability to grow millimeter-thick layers of GaAlAs with uniform composition.

14.1.5 Group IVs

Silicon and Silicon/Germanium

Ciszek [14.11] noted that high-quality Si layers had been grown on Si substrates at temperatures in the range 700–900 °C at a rate of 1 μm/min. The potential application was for solar cells, but because growth was on silicon, rather than a low-cost alternative, this was not considered to be a viable production process at that time.

However later, *Poortmans* [14.12] discussed the then current state of Si growth for thin-film solar cells. He notes that as LPE growth is close to thermal equilibrium and the Si atoms in the melt (700–900 °C) have a high diffusion coefficient, the layers have low defect density, low excess carrier recombination, leading to high lifetimes of ≈ 10 μs, and allows an in situ doping gradient to be achieved to increase diffusion length in the layer. He also discusses potential scale-up options by two methods. The first is the temperature difference method, which is a quasicontinuous means of growing Si on $10 \times 10\,\mathrm{mm}^2$ multicrystalline substrates at 0.3 μm/min, producing layers with lifetimes of 5–10 μs. The second method is a batch-type multisubstrate dipping method using substrate meltback to condition the melt. The University of Konstanz, Germany, [14.13] produced a system with 16 substrates, and an example 3 cm^2 solar cell produced a 10% efficiency.

Another report on Si for solar cells [14.14] quoted a conversion efficiency of 4.7% in Si grown from a Sn–Ga melt.

Alonso and *Winer* [14.15] grew SiGe alloys of various compositions from Si–Ge–In melts at temperatures between 640 and 900 °C. Layers were 1–5 μm thick and used to study Raman spectra features seen in material grown by MBE. The advantage of LPE-grown material was thought to lie in the random distribution of Si and Ge atoms, that is, no ordering is present, compared to the MBE-grown material. *Dost* [14.16] recently grew SiGe layers up to 14 mm diameter by LPEE with a magnetic field.

Liu et al. [14.17] grew Ge on Si for Ge-on-insulator applications, where layers showed no dislocations or stacking faults and were compatible with Si processing. More recently, *Tanaka* et al. [14.18] and *Toko* et al. [14.19] have studied the growth of Ge on Si (GoI) and on quartz, the latter as stripes on poly-Si islands for applications as thin-film transistors with high mobility.

Silicon Carbide

Dmitriev [14.20] described the production of high-quality 6H-SiC and 4H-SiC p–n junctions by LPE from Si melts. Layer thicknesses range from 0.2 to 100 μm with growth rates of 0.01–2 μm/min. Nitrogen is used as the donor impurity and aluminum, gallium, and boron as acceptor impurity elements. The material showed high carrier mobility and low deep-center concentrations. Initial attempts used a technique where molten silicon ran from an upper section of the crucible to a lower section where the SiC substrates were held fixed. Dipping was also used, in an attempt to grow material that was less stressed by Si melt solidification. Growth temperatures were 1500–1750 °C and layer thicknesses were 20–40 μm. Both these techniques produced material that was used successfully to make blue LEDs. A new version of LPE, so-called container-free LPE (CFLPE) based on the electromagnetic crucible technique, was also developed (Fig. 14.6). Liquid metal, that is, molten Si, is suspended in a high-frequency electromagnetic field at 1000 °C and the substrates, SiC, are placed on top of

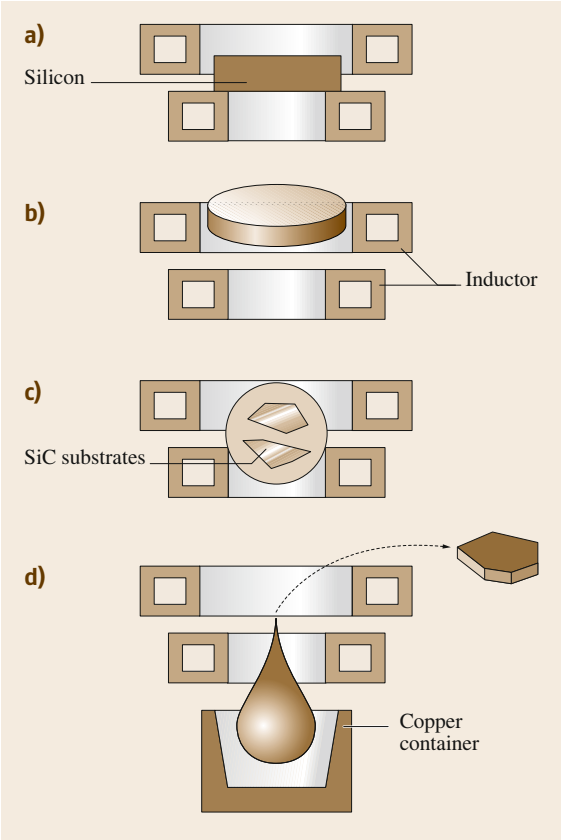


Fig. 14.6a–d Schematic of CFLPE: **(a)** solid Si in copper inductor, **(b)** solid Si suspended in the inductor, **(c)** liquid Si with SiC crystals suspended in the electromagnetic field, **(d)** Si is dropped into the container. (After [14.20])

the melt after heating to 1450 °C. A source of SiC is also placed at the bottom of the Si melt. Cooling of the solution was used to produce the epitaxial layer and the samples were then removed from the melt. The technique succeeded in producing p–n junctions by growing both layers in the same run. These formed the basis of several types of high-temperature devices (up to 500 °C), including diodes, FETs, bipolar transistors and dynistors, and optoelectronic devices such as green, blue, violet, and UV LEDs.

A more recent report [14.21] describes growth of SiC on Si substrates from a Sm:Co melt, producing an epitaxial layer initially followed by SiC whisker growth.

14.1.6 Group III–Vs

Arsenic- and Phosphorus-Based Materials

The majority of work in the area of III–V growth has been in GaAs and GaP, plus additions of As and Al. Following the earlier treatment by *Elwell* and *Scheel* [14.4], *Astles* [14.10] gave a comprehensive treatment of the LPE growth of GaAs and other III–V binaries and ternaries. He lists the advantages of LPE as high-luminescence efficiency due to the low concentration of nonradiative centers and deep levels, growth of ternary and quaternary alloys, controlled p- and n-type doping, multilayer growth with low interface recombination velocities, and good reproducibility and uniformity. Disadvantages included large areas required to be free of surface features (e.g., for photocathodes or ICs), very abrupt control of doping/composition profiles required (e.g., for microwave devices), accurate thickness control required (e.g., for microwave and quantum-well devices), and compositional grading between substrate and layer is inevitable.

All of the methods outlined above were attempted for the growth of GaAs and related materials. In addition, because LPE is a near-equilibrium technique that uses low supersaturation, nucleation is very sensitive to substrate lattice parameter and the growth rate is influenced by the substrate orientation. This enables localized growth in windows on the substrate surface and growth on nonplanar substrates with ribs or channels produced by preferential etching. The latter feature has been used to produce novel laser structures.

In fact, a vast array of both optoelectronic and microwave devices have been produced in LPE GaAs and related materials. The earliest were the GaAs Gunn devices and GaP/GaAsP LEDs. Later, GaAs/GaAlAs heterojunctions were produced for use in lasers, photocathodes, and solar cells. Other alloy systems such as GaInP for blue/green LEDs, GaInSb for improved Gunn devices, and GaInAs or GaAsSb for photocathodes were also studied. Later still came growth of ternaries, for example, GaInAsP (lattice matched to InP) for heterostructure optoelectronic devices. Finally, OEICs and buried heterostructure lasers were developed to exploit the potential for selective-area growth and anisotropy of growth rate.

Over the past several years, there have been research activities still progressing on LPE growth of some III–Vs. *Dhar* [14.22] grew a range of III–Vs on GaAs substrates, up to 8 μm thick. *Milanova* and *Vitanov* [14.23] reviewed solar-cell applications using GaAsN and InGaAsN. Addition of N into GaAs reduced the bandgap, but there was limited incorporation (0.15–0.6 at%) that nevertheless increased the carrier concentration and reduced the mobility compared to GaAs. *Mouleeswaran* et al. [14.24] grew thick GaAs layers on patterned GaAs substrates using selective-area growth while applying a DC current. Up to 268 × 318 μm^2 lateral films were produced for x-ray detector applications. This same group later [14.25] reported on Te-doped GaAs truncated pyramidal growth where the height was ≈ 9 times greater with the DC current. The dislocation density in the layer was reduced compared to that in the substrate. *Ashery* et al. [14.26] reported the growth of GaP/p-Si heterojunctions that produced moderate rectifying behavior. *Strak* et al. [14.27] grew 3–4 mm thick *bulk* GaAs by LPEE to produce large-diameter single-crystal layers with multielectrode systems. *Dost* [14.16] used a static magnetic field in his work on LPEE growth of GaAs/GaInAs, where growth rates were increased from 0.5 to 5 mm/day with application of a magnetic field. *Zytkiewicz* et al. [14.28] simulated the growth of GaAs by LPEE where Peltier cooling plus As electromigration produced a growth rate of 0.5 μm/min, which reduced with time as Joule heating took over.

III–V Antimonides

Commercially available substrates for epitaxy are limited in their lattice-constant spread and this imposes certain constraints in terms of lattice-matched growth and miscibility gaps. Ternary and quaternary alloy substrates with adjustable lattice parameters would open up new device applications. However, bulk-grown ternary alloys suffer from segregation and stress effects. An alternative approach is to grow very thick layers (> 50 μm) of these compounds for use as *virtual substrates* [14.29]. For III–V antimonides, where substrate and lattice-matching problems are acute, such thick layers are feasible by LPE due to the relatively fast growth rates (1–10 μm/min). The challenge for these virtual substrates is to produce lattice constants that are sufficiently different from those available using binary

substrates, without introducing an excessive level of defects.

Another interesting application of antimonides is that of InSb-based quantum dots [14.30]. The potential application here is in mid-IR lasers, LEDs, and detectors. In particular, there is a market for these materials as gas detectors based on IR absorption. The principle is that of rapid-slider LPE, in which a thin slit of melt is wiped across the substrate producing contact times of 0.5–200 ms. This produces low-dimensional structures such as quantum wells and quantum dots. InSb quantum dots were grown on InAs substrates at 465 °C with 10 °C supercooling and a 1 ms melt–substrate contact time. Both small (4 nm high and 20 nm in diameter) and large quantum dots (12 nm high and 60 nm in diameter) are produced.

Recently, *Compean-Jasso* et al. [14.31] grew Sn-doped GaSb at low temperatures (260 °C) to investigate the doping behavior – Sn was still an acceptor dopant, however. *Mondal* et al. [14.32] grew GaSbN on GaSb substrates with N incorporation up to 1.7%, thereby reducing the bandgap by 0.37 eV. *Sun* et al. [14.33] have grown InAsSb on InSb substrates with good full width at half-maximum values and suggested its use as an IR detector material at $\approx 8\,\mu$m cut-off.

Juarez-Diaz et al. [14.34] studied the growth of AlGaSb on GaSb substrates and found low surface roughness (39 Å), although some layers were strained and contained Te, C, and O as impurities. *Diaz-Reyes* et al. [14.35] grew InGaAsSb on GaAs substrates under lattice-matched conditions and studied the low-temperature luminescence due to the Zn dopant.

Group III Nitrides

LPE of GaN is difficult due to the low solubility of nitrogen in molten metals at atmospheric pressure. There are reports of growth of GaN from gallium and bismuth melts and in some instances the melt is replenished with nitrogen by introducing ammonia into the growth ambient, relying on a so-called VLS (vapor–liquid–solid) growth mechanism that essentially combines LPE with CVD (chemical vapor deposition). Another report [14.36] notes the use of Na fluxes as a solvent. *Klemenz* and *Scheel* [14.37] used a dipping mode at 900 °C with sapphire, $LiGaO_2$, $LiAlO_2$, and CVD GaN on sapphire substrates.

Iwahashi et al. [14.38] grew GaN layers on GaN/sapphire substrates ($10\times10\times0.5\,mm^3$) using a Na flux LPE system. The FWHM reduced from 3248 to 231 arcsec. *Hussy* et al. [14.39] modified the melt density to make problematic particles float on the melt to prevent their incorporation into the layer and thereby produce macrodefects. *Azizi* et al. [14.40] grew GaN onto AlN substrates for the first time. Cathodoluminescence measurements showed that the Al incorporation is orientation dependent. *Naritsuka* et al. [14.41] have recently grown GaN on GaN substrates using LPEE to increase the growth rate.

Other Topics

Doping with rare-earth elements (Dy, Er, Hl. Nd, Pr, Yb, Y, etc.) in the AlGaAs, InGaAs, InGaAsSb, and InGaAsP systems can lead to impurity gettering effects that radically reduce background doping and junction saturation currents and increase carrier mobilities and minority carrier lifetimes. Such rare-earth doping in InAsSb LEDs [14.42] increases the luminescence by 10–100 times.

There is no fundamental limit to the number of components in mixed alloy layers produced by LPE. For example, AlGaInPAs layers have been grown on GaAs by LPE [14.43]. Each additional element adds an extra degree of freedom for tailoring the properties of the layer, although more detailed phase equilibria data or models are required to determine accurate melt compositions and temperatures. However, as more constituents are added the melt becomes more dilute and more nearly approaches ideal behavior.

The low supersaturation of LPE makes selective modes of epitaxy feasible. A substrate can be masked (using SIO_2, Si_3N_4, TiN, etc.) and patterned with openings that serve as sites for preferential nucleation. In epitaxial lateral overgrowth (ELO), the selectively seeded material overgrows the mask. This technique has been used for defect filtering, stress reduction, substrate isolation, and buried mirrors and electrodes [14.1]. ELO is difficult with vapor-phase methods, aspect ratios (width to thickness of selectively grown material) are small, whereas that can be 100 in LPE. This could have potential for light-emitting diodes [14.1]. Another interesting application of selective LPE is the growth of pyramidal AlGaAs microtips for scanning near-field optical microscopy.

Another variant of the basic LPE process is that of LPEE, where application of an electric current through the growth interface can enhance growth rates for producing thick ternary layers [14.1]. Selective LPEE on patterned, tungsten-masked GaAs substrates can produce inverted pyramid-shaped crystals that can be used to make very high efficiency LEDs [14.1].

Mauk et al. [14.44] have reported on a massive scaling up of the LPE growth of thick ($> 50\,\mu$m) AlGaAs on 75 mm diameter GaAs substrates. The method produces a two orders of magnitude improvement in areal throughput compared to conventional horizontal sliding boat systems and has applications for LEDs, thermophotovoltaic devices, solar cells, and detectors. A large rectangular aluminum chamber is used instead

of a silica tube (Fig. 14.7). A modified sliding-boat arrangement is used with a top heater mounted above the boat cap, in conjunction with a heat exchanger/pedestal that acts as a cold finger to impose a vertical temperature gradient, which is the driving force for growth. These $> 50\,\mu m$ thick layers were then produced as free-standing layers bonded to glass.

14.1.7 Group II–VIs

Widegap Compounds

While vapor-phase methods are normally used, LPE has been used to grow some widegap Zn-based II–VI compounds [14.45]. This work was driven by the view at that time that blue LEDs could be made economically in ZnSe. The aim was to produce p–n junctions directly by LPE via growth at 950–650 °C, that is, much lower than the bulk crystal growth, to reduce the number of defects. The problems included the high Se vapor pressure, necessitating a closed-tube approach and the need to maintain the ZnSe substrates in the upper portion of the vertically held melt during deposition. Nevertheless, 10–20 μm thick layers were grown in 2 h. Growth at 950 °C produced smoother surfaces, but the layers contained more deep levels and impurities compared with those grown at the lower temperatures. Addition of a separate Zn vapor pressure source improved the properties, showing p-type conductivity, and doping with Au, Na, and Li was also attempted. All these produced p-type material but there was no n-type material reported. Later, the same group [14.46] reported growth of p-type ZnSe doped with Na_2Se from which p–n junctions were fabricated after Ga diffusion from a Zn solution, to produce the n-type layer. Blue light was emitted at a wavelength of 471 nm.

Astles [14.47] reviewed the work done on LPE of CdTe-based compounds. Most studies were carried out from Te-rich solutions in the temperature range 500–900 °C. Layers are p-type as-grown or n-type if doped with In or Al. Growth rates are typically 0.5 μm/°C at 500 °C. Growth from Bi-rich melts was also studied and this was found to improve melt wipe-off and surface morphology. Buffer layer growth of CdZnTe layers was used by *Pelliciari* et al. [14.48] as impurity barrier layers. Both CdMnTe [14.49] and HgCdMnTe [14.50] have also been grown by LPE. The latter compound was used to produce mesa diodes for room-temperature 1.3–1.8 μm applications. *Dost* [14.16] recently used a 0.45 T magnetic field in his LPEE growth of CdTe, producing 2 mm thick samples.

MCT (Mercury Cadmium Telluride)

The situation regarding LPE of MCT has been reviewed by *Capper* et al. in 1997 [14.8] and more recently by *Capper* [14.51]. In 1997, LPE had emerged as the predominant materials growth technology for the fabrication of both first- and second-generation MCT IR focal plane arrays (FPAs). The technology had advanced to the point where material could be routinely grown for high-performance photoconductive (PC), photovoltaic (PV), and laser detector devices in the entire 2–18 μm spectral region. Two different technical approaches were pursued, with virtually equal success: growth from Hg solutions and growth from Te solutions. One major advantage of the Hg-solution technology was its capability to produce layers of excellent surface morphology due to the ease of melt decanting. Two additional unique characteristics are now widely recognized as essential for the fabrication of high-performance double-layer heterojunction (DLHJ) detectors by LPE: low liquidus temperature (< 400 °C), which made the cap-layer growth step feasible, and ease of incorporating both p-type and n-type temperature-stable impurity dopants, such as As, Sb, and In, during growth.

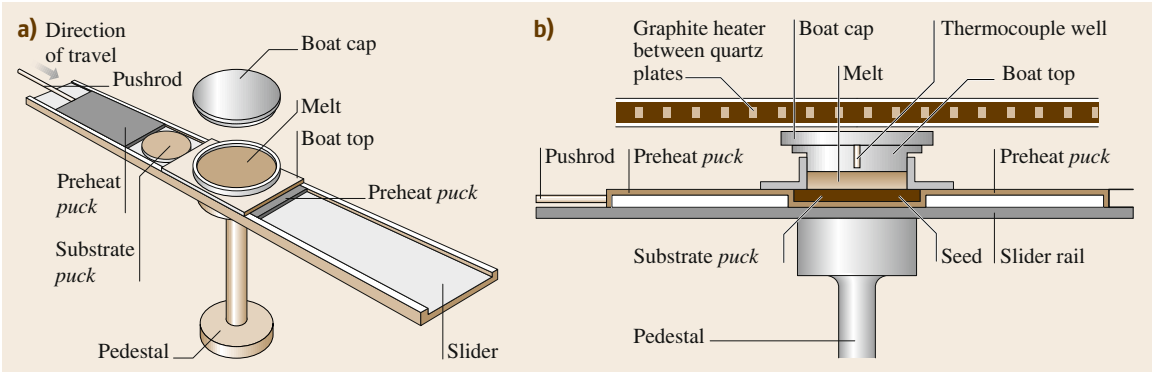


Fig. 14.7a,b Schematic of 3-inch LPE apparatus, **(a)** perspective view, **(b)** side view. (After [14.44])

Figure 14.4 shows the dipping system used for the growth of MCT [14.8]. Large melts allow the production of layers up to 30 cm^2 with excellent compositional and thickness uniformity and allow dopant impurities to be accurately weighed for incorporation into layers and to maintain stable electrical characteristics over a long period of time. Four layers (30 cm^2 each) with a total area of 120 cm^2 can be grown in a single run [14.52]. *Norton* et al. [14.53] also scaled up for the growth of cap layers from Hg-rich solutions, each reactor capable of growth on four 24 cm^2 base layers per run.

While layers grown from Hg-rich solutions are easily doped with group VB elements with high solubility, layers grown from tellurium-rich solutions are not. Group VB dopants have low solubility and are not 100% active electrically. Group IIIB elements, indium in particular, are easily incorporated from both solutions. Indium doping from tellurium-rich melts, however, has one advantage in that the segregation coefficient is near unity.

Astles [14.47] reviewed the experimental data of Te-rich LPE growth at 460–550 °C. As an example, to compare growth parameters for Te solutions with those for Hg solutions, consider the growth of LWIR MCT ($x = 0.2$) at 500 °C from both Te and Hg solutions. The x_{Cd} for Te-rich solutions is 8.3×10^{-3}, while x_{Cd} for Hg-rich solutions is 2.6×10^{-4}. This is one of the difficulties encountered in LPE growth from Hg-rich solutions. Use of large melts, however, overcomes the Cd depletion problem.

A typical sliding-boat arrangement was described by *Capper* et al. [14.54]. The LPE boat is made from purified, high-density, small-grain, electronic-grade graphite. The precompounded growth solution is placed in a growth well, and crushed HgTe is placed in a blind well. The HgTe acts as a buffering source of the volatile elements to help stabilize the growth solution composition and also to deliver an overpressure to control the metal vacancy level during annealing. The control over the impurity levels in the major constituent elements is a crucial part of the control of the overall process.

Surface morphology is controlled at two levels: microtexture and long-range variation. The micro-texture is a result of misalignment of the substrate crystal plane with the growing surface. Deviations $> 0.1°$ from the ⟨111⟩B plane lead to significant surface texture. Growth on accurately orientated substrates gives a specular surface on which atomic-scale growth features can be seen using atomic force microscopy (AFM). Figure 14.8 shows a classical Frank–Read site on an as-grown LPE layer surface [14.55].

In addition to the normal MW (3–5 µm) and LW (8–12 µm) IR detectors made in this LPE material, work has also been carried out in the SW/MW (1–3 µm) region [14.2, Chap. 10]. Using avalanching produces signal gains up to 10, with a sensitivity of ≈ 10 photons rms, attesting to the extremely high material quality. Figure 14.9 shows an example of its application in active IR imaging, combined with a laser for range finding. This SW material is also being used currently in several astronomical applications, where again the high level of material quality is essential for the correct operation of the device (see Chap. 34 in this book).

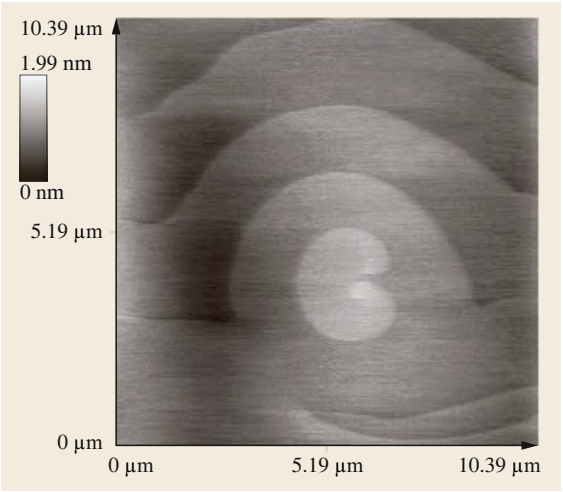


Fig. 14.8 AFM image of a Frank–Read growth spiral. (After [14.55])

Schirmacher et al. [14.56] recently reported on MCT gown by LPE for higher operating temperature (HOT) devices using Au as the p-type dopant, used to reduce the number of native mercury vacancies known to act as recombination centers. By lowering the p-type dopant level and improving the passivation they achieved Rule 07 performance, a normalized measure of the dark current, when operating the detectors at ≈ 140 K.

Other Narrowgap II–IV Compounds

HgZnTe was first proposed as an alternative detector material to MCT due to its superior hardness and its high energies for Hg vacancy formation and dislocation formation [14.57]. *Rogalski* [14.58] reviewed the LPE growth of HgZnTe and notes that Te-rich growth is favored due to the low solubility of Zn in Hg and the high Hg partial pressure. He also commented that the same factors apply to the growth of HgMnTe. *Becla* et al. [14.49] grew HgMnTe in a two-temperature, closed-tube tipping arrangement at 550–670 °C onto CdMnTe bulk substrates and CdMnTe LPE layers previously grown on CdTe substrates. *Rogalski* [14.58]

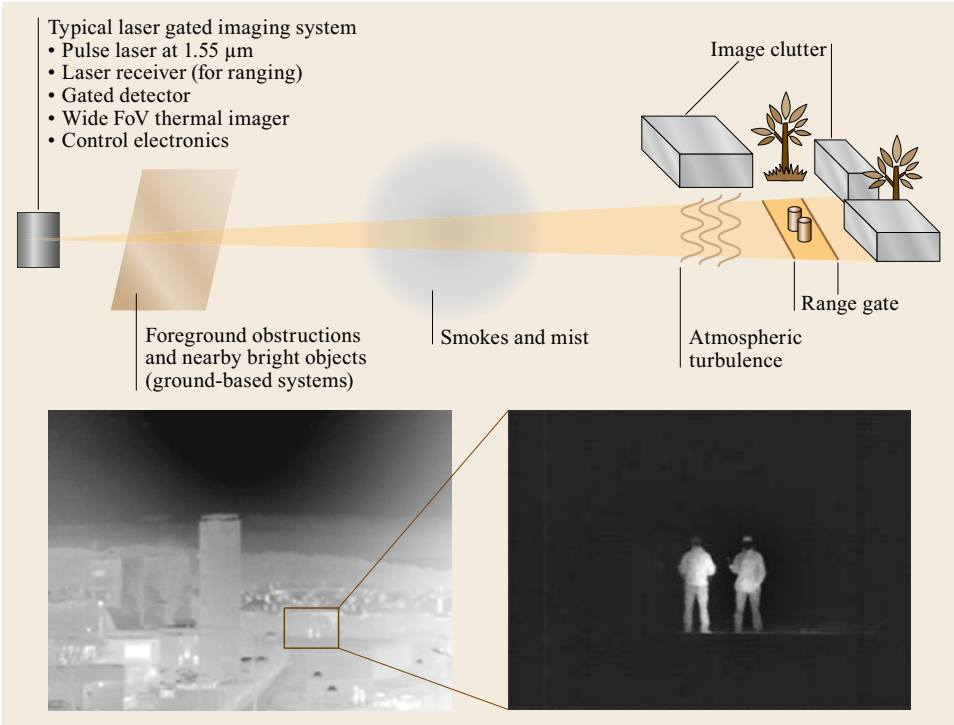


Fig. 14.9 Advantages of laser-gated imaging in conjunction with other sensors together with a typical pair of images. (Courtesy of Leonardo MW Ltd)

also reviewed the status of PC and PV detectors in both HgZnTe and HgMnTe.

14.1.8 Garnets

Comments have already been made earlier in this chapter regarding the growth of garnets by LPE. *Gornert* et al. [14.2, Chap. 14] have recently described the growth of garnets with large Faraday rotation for magneto-optical applications. Bi-substituted gadolinium, gallium garnet (GGG) layers are grown on 1-inch diameter GGG substrates from Pb/Bi/B/O melts that show low misfit, low effective anisotropy, low defect levels, and correct crystal orientation. LPE is noted as producing devices with the highest sensitivity compared to material produced by other techniques.

In a more recent report [14.59], Ce-doped multicomponent garnets for scintillating screens for x-ray imaging was described. Both YAG and GGG substrates were used.

14.1.9 Oxides/Fluorides

In a study of ZnO growth onto ZnO substrates *Ehrentraut* et al. [14.60] reported that AFM showed a monolayer step height. More recently, *Kobayashi* et al. [14.61] grew Al-doped ZnO (at 800 °C) and found subsequent well-defined excitonic luminescence.

Nasu et al. [14.62] grew YBCO on $NdGaO_3$ substrates and studied the orientation versus growth temperature. Work on Zn-substituted $LiNbO_3$ waveguides was described by *Dubs* et al. [14.63], where high crystalline perfection and extremely flat surfaces (0.2–0.3 µm) were produced. *Lu* et al. [14.64] worked on Yb:Zn $LiNbO_3$ in a K_2O flux and found that the segregation coefficient and surface roughness increased with dopant concentration. *Pena* et al. [14.65] grew Yb:CaF_2 layers on CaF_2 substrates up to a few hundred micrometers thick with structural and spectroscopy properties as good as the substrates and similar to bulk crystals, respectively.

14.1.10 Atomically Flat Surfaces

Chernov and *Scheel* [14.66] argued that far from the perceived drawback of LPE of producing rough surfaces, it may be uniquely suited to providing atomically flat, singular surfaces over distances of several micrometers. These surfaces could have applications in surface physics, catalysis, and improved homogeneity of layers and superlattices of semiconductors and superconductors.

In support of this view, Fig. 14.8 shows an AFM image of a Frank–Read growth spiral on the surface of an MCT layer grown by LPE in this author's laboratory [14.55].

14.1.11 Conclusions

LPE was generally the first epitaxial technique applied to most systems of interest in micro- and optoelectronics. It is now generally a mature technology, several optoelectronic, IR detectors, and other device types are still being made in LPE material, although some developments are still taking place. LPE has several advantages over the various vapor-phase epitaxial techniques, such as high growth rates, favorable impurity segregation, ability to produce flat faces, suppression of certain defects, absence of toxic materials, and low cost. There is much less emphasis on LPE in the current literature than on the vapor-phase methods, but LPE continues to seek out and develop in several niche markets where vapor-phase techniques are not suitable.

14.2 Metal Organic Chemical Vapor Deposition

14.2.1 Introduction and Background

The technique of MOCVD was first introduced in the late 1960s for the deposition of compound semiconductors from the vapor phase. The pioneers of the techniques, *Manasevit* and *Simpson* [14.67] were interested in a method for depositing optoelectronic semiconductors such as GaAs onto different substrates such as spinel and sapphire. The near-equilibrium techniques such as LPE and chloride VPE were not suitable for nucleation onto a chemically very different surface to the compound being deposited. These pioneers found that if they used combinations of an alkyl organometallic for the group III element and a hydride for the group V element, then films of GaAs could be deposited onto a variety of different surfaces. Thus, the technique of MOCVD was born but it was not until the late 1980s that MOCVD became a production technique of any significance. This success depended on painstaking work with improving the impurity of the organometallic precursors and hydrides. By this time, the effort was on high-quality epitaxial layers on lattice-matched substrates, in contrast with the early work. The high-quality epitaxial nature of the films was emphasized by changing the name of the growth method to metalorganic vapor phase epitaxy (MOVPE) or organometallic VPE (OMVPE). All these variants of the name can be found in the literature and in most cases they can be used interchangeably. However, MOCVD can also include polycrystalline growth that cannot be described as epitaxy. The early niche applications of MOVPE were with GaAs photocathodes, GaAs HBT lasers, and GaInAsP lasers and detectors for 1.3 μm optical fiber communications.

The characteristics of MOCVD that have taken it from a research curiosity to production have been in the simplicity of delivery of the reactive vapors and the versatility of selecting different alloy compositions, dopants, and layer thickness. These basic attributes have enabled the same basic technique to be used for narrow bandgap semiconductors such as the infrared detector materials $Cd_xHg_{1-x}Te$ and GaInSb and now for wide bandgap semiconductors such as GaN and ZnO. Indeed, the success of GaInN in the 1990s for high brightness blue LEDs has now led to this being the most popular material produced by MOCVD for high-brightness white LEDs for lighting [14.68, 69]. The early strength of MOCVD was its ability to grow onto different substrates but later abandoned in favor of the more conventional homoepitaxy; however, the nitrides rely on heteroepitaxy onto sapphire and SiC substrates bringing MOCVD back to its roots with the early work of Manasevit. The goal of more recent work has been to achieve a reduced density of dislocations in the GaN layer using patterned substrates [14.70]. This versatility with substrate materials presents MOCVD with the ultimate challenge of mating high-performance optoelectronic materials with silicon substrates to combine the best of optoelectronic and electronic performance.

This section of the chapter will cover the key elements of the MOCVD process from the physical characteristics of the precursors, through reactor design to getting the right materials properties for high-performance devices.

14.2.2 Basic Reaction Kinetics

The precursors for III–V MOCVD are generally a simple alkyl for the group III source and a hydride for the group V source. Both have the essential properties of being volatile in a suitable carrier gas stream (usually hydrogen) and being chemically stable at ambient temperature. These precursors are normally mixed outside the reaction chamber, introduced into the reaction chamber through a suitable injector arrangement and directed onto a hot substrate. This is schematically shown in Fig. 14.10. The details of reactor design will be discussed later in this chapter. The reaction of the precursors to yield the III–V compound on the substrate can occur either in the hot vapor above the surface or on the hot surface. The stoichiometric reaction for GaAs

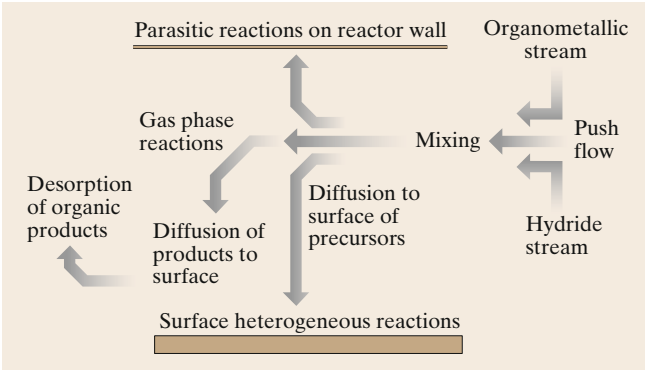


Fig. 14.10 Schematic of MOCVD process from mixing of gas streams to reaction on the substrate surface

growth is given as follows

$$(CH_3)_3Ga + AsH_3 \rightarrow GaAs + 3\,CH_4\,. \qquad (14.1)$$

This reaction has been the most widely studied of all the MOCVD reactions and was one of the original processes reported by *Manasevit* and *Simpson* [14.67]. One reason why it works so well is that the hydrogen required to satisfy the CH_3 radical bond is supplied from the arsine hydride and at normal growth temperatures (around 700 °C) avoids unsatisfied carbon bonds that could lead to carbon incorporation. In reality, this very simple picture covers a complexity of reaction steps that have to take place and has been discussed in great detail by *Chernov* [14.71]. However, some of the important reaction steps will be described here as an introduction to the kinetics of GaAs MOCVD.

The schematic shown in Fig. 14.11 gives some of the important reaction steps that have been identified

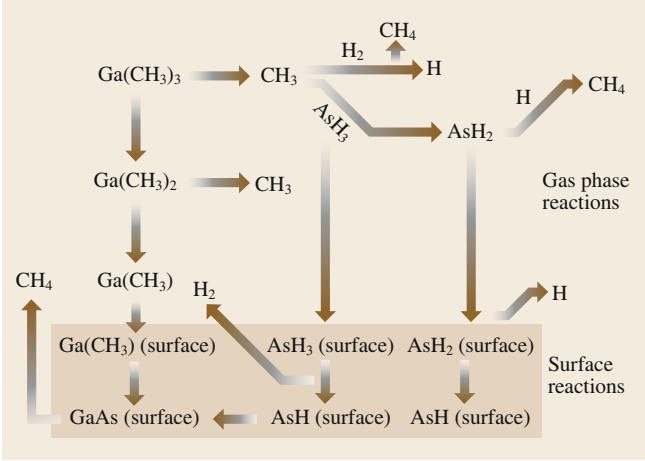


Fig. 14.11 Reaction scheme for epitaxial deposition of GaAs showing the most important vapor phase and surface reaction steps

for the reaction of GaAs. This gives some insight into the complexity of the reaction kinetics and it is worth remembering that this is a relatively straightforward reaction for MOCVD. Fortunately, one does not have to understand every step in the process before attempting to grow a layer and goes some way to explaining why MOCVD has developed along very empirical lines. An understanding of the reaction kinetics does, however, enable some of the features of MOCVD to be understood, particularly when these relatively simple precursors are replaced by more complex precursors.

It can be seen from Fig. 14.11 that the reaction process is started by a gas-phase homolysis of TMGa ($(CH_3)_3Ga$) to yield dimethylgallium and methyl radicals (CH_3). There are two important roles that the methyl radicals can take and this is generally important in all MOCVD processes for deposition of III–V semiconductors:

1. Methyl radicals can react with the ambient hydrogen carrier gas to yield stable methane and hydrogen radicals.
2. Methyl radicals can react with arsine (AsH_3) to yield stable methane and AsH_2.

Both these steps can initiate the decomposition of arsine either through a methyl radical or a hydrogen radical removing a hydrogen atom from the arsine; these reaction steps can be seen in Fig. 14.11. There are some important consequences of these initial reaction steps and one you may have already spotted is that the hydrogen carrier gas is not included in the stoichiometric reaction (reaction 1) but can play a part in the reaction process. Another consequence is that although it is highly unlikely that the two (or more) precursors have the same thermal decomposition characteristics, they can be used together to react at the same temperature through radical reaction mechanisms. The latter has been illustrated in many studies of different systems using Erzats reaction tube experiments and further information on these can be found in *Stringfellow* [14.72]. For example, in the case of GaAs, the pyrolysis of TMGa will occur at 500 °C but for arsine decomposition, a temperature of around 700 °C is required.

The stepwise removal of ligands from TMGa and arsine can occur in the vapor or on the surface. Figure 14.11 shows both alternatives and the dominant path will depend on both the surface temperature and vapor temperature above the substrate. A general rule is that more of the reaction process will occur on the surface at lower temperatures when the vapor reaction rate will be slower and the surface adsorption will be more efficient. It is also important to note that there are two different adsorption sites on the surface, the Ga sites that will

take the anion species and the As sites that will take the cation species.

Part of the early success of GaAs and the alloy GaAlAs was due to the absence of unwanted vapor-phase reactions that could lead to the formation of polymer chains that would contaminate the layer. This was not the case with InP where the liquid alkyl source TEIn was used instead of the solid TMIn. This led to a mist in the reactor resulting from the formation of an adduct between the precursors that subsequently polymerized [14.73]. This was overcome through the use of low-pressure reactors and this has remained a feature of most production reactors today. However, the preferred precursor is TMIn and the problems of uncertain supply from a solid source has been overcome by a variety of different innovative methods. One example is to form a liquid with a stable amine and the source will then behave in the same way as any other liquid precursor as discussed by *Frigo* et al. [14.74].

The details of the reaction kinetics will change with substrate temperature but as the temperature is increased a point will be reached where the rate of epitaxial growth will no longer be determined by the overall reaction rate but will be determined by the supply of precursors to the substrate. This will be reflected by a depletion of the precursor concentration immediately above the surface and a gradient in precursor concentration toward the undepleted free stream. The limitation to the rate of epitaxial growth then becomes the rate of diffusion through the depleted boundary layer to the substrate. This is called transport-limited growth and is characterized by high growth rates and only a weak dependence of growth rate on substrate temperature. This is shown schematically in Fig. 14.12. The plot is of ln(growth rate) versus $1/T$ because of the expected Arrhenius relationship in the rate constants. This really only applies to the low-temperature (kinetic regime) where the growth rate can be expressed as

$$\text{growth rate} = A \exp\left(-\frac{E_a}{RT}\right) , \qquad (14.2)$$

where A is a constant and E_a is an activation energy, R is the gas constant. It is unlikely that E_a can be attributed to the activation energy for a single reaction step but is still useful in characterizing the overall kinetics where different precursors are being tried.

In the transport-limited regime, there will still be a small dependence of growth rate on temperature due to the increase in diffusion rate with temperature and this is illustrated in Fig. 14.12. Most MOCVD growth processes will take place in the transport-limited regime where it is easier to control growth rate. However, there are a number of growth processes that will occur at lower temperatures in order to control the

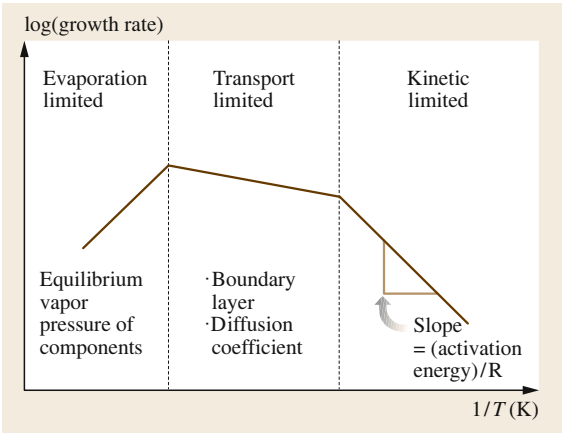


Fig. 14.12 Schematic of the three different temperature regimes for growth of a compound semiconductor by MOCVD

properties such as native defect concentration of the epitaxial films. This is generally the case with II–VI semiconductors, but can also apply to the formation of thermodynamically unstable III–V alloys.

In the high-temperature regime, the growth rate decreases with temperature as the equilibrium vapor pressure of the constituent elements in the film will increase and give desorption rates similar to the deposition rate, leading to significant loss of material through evaporation to the gas stream.

14.2.3 Precursors

The choice of precursors is not confined to simple alkyls and hydrides but can extend to almost any volatile organometallic as a carrier for the elemental components of a film. In the case of II–VI semiconductors, it is usual to use an alkyl for both the group II and the group VI elements. Hydrides have been used as Se and S sources but prereaction makes it difficult to control the growth process and in particular can make it difficult to incorporate dopants. The use of combined precursor sources has been extensively researched but is not in common use for epitaxial device-quality material. One reason for this is the difficulty in controlling the precursor ratio that is needed to control the stoichiometry of the material.

The important properties of precursors, and their selection, can be generalized and provide a basis for optimizing the MOCVD process. These properties can be summarized as follows:

1. Saturated vapor pressure (SVP) should ideally be in the range of 1−10 mbar in the range of temperatures of 0−20 °C.

2. Stable for long periods at room temperature.
3. Will react efficiently at the desired growth temperature.
4. The reaction produces stable leaving groups.
5. Avoids unwanted side reactions such as polymerization.

According to the Clausius–Clapeyron equation, the SVP of a liquid is given by an exponential relationship

$$\mathrm{SVP} = \exp\left(-\frac{\Delta G}{\mathrm{R}T}\right) , \tag{14.3}$$

where ΔG is the change in Gibbs free energy on evaporation, R is the gas constant, and T the temperature of the liquid in the bubbler. This can be expressed as the heat of evaporation ΔH and the entropy for evaporation ΔS, where $\Delta G = \Delta H - T\Delta S$; this gives the familiar form of the SVP equation

$$\begin{aligned}\mathrm{SVP} &= \exp\left(-\frac{\Delta H}{\mathrm{R}T} + \frac{\Delta S}{\mathrm{R}}\right) \\ &= \exp\left(\frac{\Delta S}{\mathrm{R}}\right)\exp\left(-\frac{\Delta H}{\mathrm{R}T}\right) .\end{aligned} \tag{14.4}$$

This is of the form

$$\log_{\mathrm{e}}(\mathrm{SVP}) = A - \frac{B}{T} , \tag{14.5}$$

where A and B are constants given by $A = \Delta S/\mathrm{R}$ and $B = \Delta H/\mathrm{R}$. Manufacturers of the precursors will generally give the SVP data in the form of the constants A and B in (14.5). It can also be given in the form

$$\log_{10}(\mathrm{SVP}) = A' - \frac{B'}{T} . \tag{14.6}$$

To convert the constants in (14.6) to (14.5), just multiply by ln 10. Some examples of the SVP constants, along with the calculated SVP at 20 °C, for a number of typical precursors are shown in Table 14.1.

Table 14.1 List of precursors with the vapor pressure constants according to (14.6)

Precursor	*A*	*B*	SVP at 20 °C (mmHg)
TMGa	8.07	1703	182
TEGa	8.08	2162	5.0
TMAl	8.22	2134	8.7
TEAl	9.0	2361	0.02
TMIn	10.52	3014	1.7
TEIn	8.94	2815	1.2
Solution TMIn	10.52	2014	1.7
DMZn	7.80	1560	300
DEZn	8.28	2109	12
DMCd	7.76	1850	28.2
DES	8.184	1907	47
DMSe	9.872	2224	191
DESe	8.20	2020	20.2
DMTe	7.97	1865	40.6
DIPTe	8.29	2309	2.6

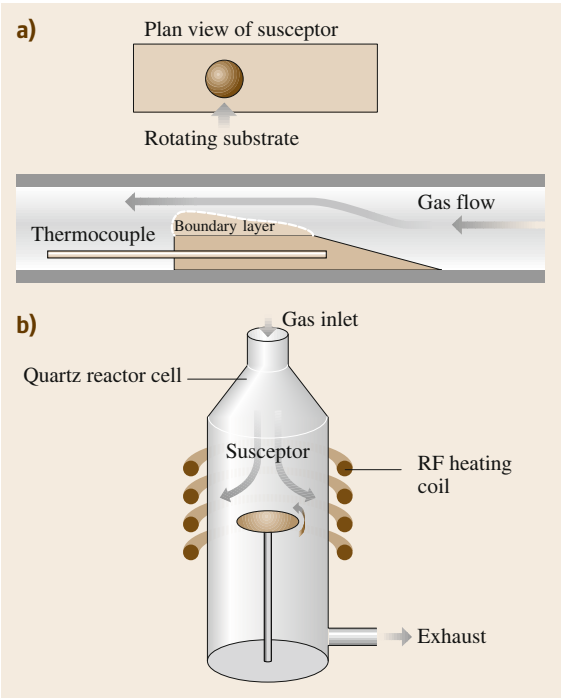


Fig. 14.13a,b Schematic of (a) horizontal and (b) vertical reactor cells

14.2.4 Reactor Cells

The design of reactor cells has formed a very important part of the development of MOVPE and has been crucial in scaling laboratory processes to large-scale production [14.75–77]. The original research reactors fell into one of two groups, either the vertical reactor or horizontal reactor. These reactor designs are schematically shown in Fig. 14.13. The substrate is placed onto a graphite susceptor that is heated by either RF coupling via a coil surrounding the reactor, a resistance heater underneath the susceptor or lamps placed underneath the susceptor. The reactor wall can be water-cooled or gas-cooled to minimize reaction and deposition onto these surfaces. Either of these reactor cells could be operated at atmospheric or reduced pressure. For reduced-pressure operation, the reactor-cell pressure would be typically a tenth of an atmosphere but a wide range of different pressures have been successfully used. Reduced pressure will increase the gas velocity and help to overcome the effects of free convection from a hot

substrate. The forced-convection parameter that is often quoted is the Reynolds number and is proportional to gas velocity. A high Reynolds number will ensure streamline flow and at low Reynold's number the buoyancy effects of the hot substrate will take over and the gas flow will be dominated by free convection (characterized by the Grashof number) and become disorganized with recirculation cells. In transport-limited growth, the exact nature of the gas flow will determine the uniformity of deposition and can also affect the defect concentration in the films due to particulates and reaction products being swept back across the growing surface. It is not normal to achieve the very high flow velocities and Reynold's numbers associated with turbulent flow in an MOVPE chamber; however, the disorganized flow due to free convection is often, wrongly, referred to as *turbulent flow*.

Another reason for using high flow velocities is to overcome the effects of depletion of the precursor concentration at the downstream end of the deposition region. For transport-limited growth, the growth rate is limited by the rate of diffusion from the free stream to the substrate. This region is called the boundary layer [14.78] and increases in thickness going downstream from the leading edge of the susceptor, as shown in Fig. 14.14. Some horizontal reactors are designed with a tilt in the susceptor so that the free cross-sectional area decreases and hence the flow velocity increases going downstream. This helps to flatten the boundary layer and ensure better uniformity.

Maintaining a boundary layer has a cost in high utilization of expensive precursors and gases as the flow throughput is typically very high and the proportion of precursors reacted in the region of the substrates is very low. This became a more serious problem when reactors were scaled to hold multiple wafers for production. These problems have been resolved with different solutions in the vertical and horizontal reactor configurations with the Emcore (now Veeco) turbo-disc reactor, the Thomas Swan (now Aixtron) showerhead reactor, and the Aixtron planetary reactor. A fourth type of reactor design was introduced by EMF, called the Titan, and works on the principal of introducing the precursors separately into the reactor chamber. This approach proved difficult to control and has not been developed further. A common feature to all these production reactor designs is the substrates are rotated so the concentration of precursors and reactants arriving at the substrate do not have to be uniform across the surface as a portion of the substrate will alternately experience high and low concentrations that will average out. Each of these reactor designs is schematically shown in Fig. 14.15.

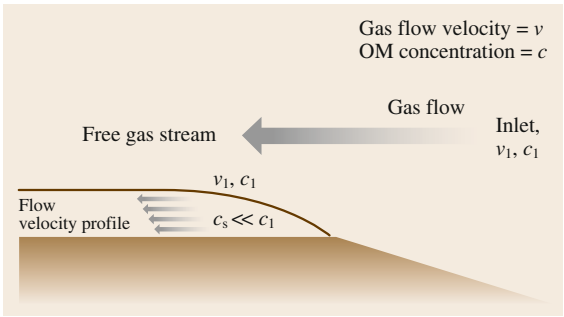


Fig. 14.14 Schematic of a boundary layer in a horizontal reactor

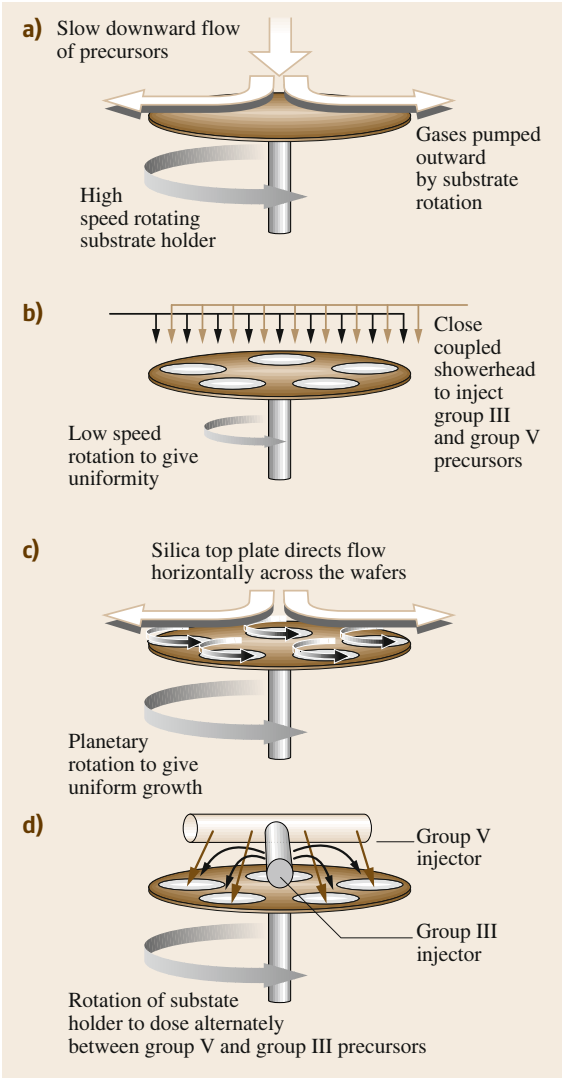


Fig. 14.15a–d Schematic of four different production reactor designs: **(a)** the Emcore (Veeco) Turbo-disc, **(b)** the Aixtron close-coupled showerhead, **(c)** the Aixtron planetary, and **(d)** the EMF Vector flow reactor

The turbo-disc reactor, schematically shown in Fig. 14.15a, is a vertical reactor configuration but the boundary layer is kept to a narrow region above the susceptor by high-speed rotation that pumps the gas radially outward due to viscous drag. The rotation speeds are up to 2000 rpm in order to create this lateral flow of the constituents above the substrate. This is continuously replenished from the slower downward gas stream and results in excellent uniformity of deposition across the wafers and a high utilization of the reactant gases *Tompa* et al. [14.79]. The reactor pressure is typically around 100 mbar.

The showerhead reactor is another vertical reactor arrangement but takes a different approach to overcoming free-convection currents and poor uniformity [14.76]. The precursors are introduced through a water-cooled showerhead placed just above the susceptor. The susceptor is rotated but typically at much lower speeds than for the turbo-disc reactor. The precursor distribution can be balanced across the width of the reactor to give a uniform supply of precursors.

The planetary reactor is a horizontal flow arrangement where the reactants enter at the center of rotation of the susceptor and flow outward. This is an example of a fully developed flow where depletion of the reactants is occurring as the gases move away from the center and this will be accentuated by a decrease in the mean flow velocity as the gases move outward [14.80]. This would normally give very poor uniformity but the planetary rotation mechanism will rotate each wafer on the platen so that it will sample alternately high and low concentrations giving uniform deposition. This approach has the advantage of a high utilization of the precursors and the ability to extend the design to very large reaction chambers for multiple wafers from 2 to 8 inches diameter.

14.2.5 III–V MOCVD

This section will consider the range of III–V materials grown by MOCVD and the precursors used. Most of the III–V semiconductors can be grown from organometallics of the group III element and hydrides of the group V element. Exceptions to this will be noted where appropriate.

Arsenides and Phosphides

The most commonly studied alloy system is $Al_{1-x}Ga_xAs$ that is used for LEDs and laser diodes from the near infrared to the red part of the visible spectrum. This is a well-behaved alloy system with only a small change in lattice parameter over the entire composition range and covering a range of bandgaps from 1.435 eV for GaAs to 2.16 eV for AlAs. One problem with this alloy is the sensitivity of aluminum to oxygen, which makes it extremely difficult to grow high-quality AlAs. Just 1 ppm of oxygen contamination will result in $10^{20}\,cm^{-3}$ incorporation of oxygen into $Al_{0.30}Ga_{0.7}As$ [14.81]. In addition to the normal MOCVD precautions of using ultrahigh-purity hydrogen carrier gas and ensuring that the moisture in the system is removed, the hydrides and organometallics also need to have extremely low oxygen content. Precursor manufacturers have tended to keep to the simple alkyl precursors but to find innovative ways of reducing the alkoxide concentrations.

Alternative group V precursors have been sought due to the high toxicity of arsine and phosphine. These hydride sources also suffer the hazard that they are stored in high-pressure cylinders and any leakage could result in the escape of large quantities of toxic gas. Alternative alkyl group V sources have been extensively researched but only two precursors have proved to be suitable for high-quality epitaxial growth, tertiarybutylarsine (TBAs), and tertiarybutylphosphine (TBP) [14.82, 83]. These precursors only have one of the hydrogen ligands replaced with an alkyl substituent but they are liquid at room temperature rather than high-pressure gases. In the reactor chamber, the likely reaction path is to form the hydride by a process called beta-hydrogen elimination. This entails one of the hydrogen atoms from the methyl groups satisfying the bond to As (or P) with a butene leaving group as shown below [14.43]

$$C_4H_9AsH_2 \rightarrow C_4H_8 + AsH_3\,. \qquad (14.7)$$

This process is more likely to dominate at the normal growth temperature for transport-limited growth and effectively yields the arsine precursor that can then react in the normal way. In the search for alternative alkyl precursors, this proved to be an important factor as the fully substituted alkyl arsenic sources tended to incorporate large concentrations of carbon, degrading the electrical properties of the film. The importance of the group V hydride was discussed in Sect. 3.2 and it can be understood why TBA and TBP (for the phosphorous alloys) have proved to be good alternatives to the hydrides. However, it is fair to say that these have never been widely utilized due to much higher cost than the hydrides and poor availability.

An alternative in improving the safety has been investigated and relies on the same principle of reducing the toxic gas pressure in the event of a system leak. This alternative stores the hydride in a reversible adsorption system [14.84]. The adsorption system keeps the hydride at subatmospheric pressure and requires pumping to draw-off the hydride when needed hence making it

inherently safer. One major advantage to this system, in addition to the inherent improvement in safety, is that the precursors and hence the precursor chemistry are unchanged in the reactor cell.

Other alloys commonly grown using MOCVD include $In_{0.5}Ga_{0.5}P$ that has a bandgap of approximately 2 eV and is lattice matched to GaAs. The quaternary alloy GaInAsP enables lattice matching to InP substrates while controlling the bandgap in the 1.3 and 1.55 μm bands used for long-range fiber-optic telecommunications.

More recent research in the arsenides and phosphides has focused around the nanostructuring of the materials through self-assembled quantum dots (QD) [14.85, 86] and nanowires [14.87, 88]. This has enabled bandgap engineering to overcome limitations of lattice-matched structures. For example, stacked lasers of InAs QDs produce emission beyond 1300 nm [14.86]. The mechanism is to take advantage of the inherent strain created when depositing InAs onto GaAs where the large lattice mismatch results in only 2 monolayers to be grown as a 2-D layer before strain relaxation creates three-dimensional (3-D) growth. This mechanism is called the Stranski–Krastanov process and results in a 2-D array of self-assembled QDs. These can be overgrown with an epitaxial layer of GaAs to cap the QDs and to allow another layer of InAs QDs to be grown on top [14.85]. Further refinements in the MOCVD of QD arrays uses a strain relaxation layer on top of the InAs QD with GaAsSb [14.86] to maintain the Type-I quantum transition to longer wavelengths.

The formation of nanowires has enabled greater relaxation of the lattice-matching requirement and is gaining attention as a means of making optoelectronic structures onto silicon substrates. The approach is to use a vapor–liquid–solid (VLS) mechanism where a thin gold film is heated to produce nanoscale gold droplets that are then exposed to the precursor gases. These will react in the normal way but the Ga (or In) and As (or P) released from the precursors will dissolve into the gold nano-droplet and start to nucleate a GaAs film at the gold/substrate interface. By changing the gas phase composition during the deposition, it is possible to grow junctions and heterostructures in the vertical nanowires. *Heurlin* et al. [14.88] have grown tandem InP p–n junctions in nanowires onto silicon substrates for photovoltaic devices. These structures have the added advantage that the nanowires will also help to capture the light, requiring less of the semiconductor material to achieve light capture than would be required for a conventional planar device. *Gu* et al. [14.89] grew GaAs/AlGaAs core–shell nanowires and assessed their optical absorption properties showing enhanced absorption compared with a planar structure. The core–shell structure has the advantage of passivating the large surface area created by the nanowires and is an alternative to the vertical junction structure.

Antimonides

The antimonides cover an important range of bandgaps from the near infrared to the mid-infrared bands, up to 5 or 6 μm. These compounds and alloys can be used in infrared detectors, thermophotovoltaic (TPV) devices, and high-speed transistors. The growth of the antimonides is more complex than for the arsenides and phosphides because the hydride, stibine, is not very stable at room temperature so the use of alkyl precursors has been a more natural choice. Another factor that has influenced the growth of the antimonides is the lower thermodynamic stability and decomposition of substrates such as InSb above about 400 °C. It is also desirable to grow the films at a much lower temperature than for the arsenides and phosphides in order to keep the native defect concentration low for controlled n-type and p-type doping. However, the antimonides have an advantage in the reaction chemistry that they do not incorporate carbon as readily as in the arsenides and phosphides and there is greater flexibility over the choice of antimony precursors [14.90]. The easiest choice is to use trimethylantimony (TMSb) or to use larger alkyl groups such as triethyl and triisopropyl to reduce the reaction temperature as required. It is also possible to reduce the reaction temperature by replacing TMGa with triethylgallium (TEGa) [14.91]. This approach has been particularly advantageous when growing aluminum-containing alloys where carbon incorporation can be a problem, but is reduced using TEGa. This serves to illustrate the flexibility of MOCVD and has given more scope for design of precursors, not envisaged in the early days of MOCVD.

Nitrides

The nitrides, mainly GaN and the alloy GaInN, have brought MOCVD into prominence as a manufacturing technology with the success of high-brightness blue LEDs making large-screen full-color LED displays a reality. This has extended in recent years to high efficiency and high-brightness white LEDs making huge inroads into the lighting market [14.92, 93]. The precursors used for the nitrides are standard with the methyl alkyls (TMGa and TMIn) for the group III elements and ammonia for the nitrogen. The key technological barriers to obtaining device-quality GaN were in overcoming the problems associated with heteroepitaxy onto the nonlattice-matched substrate, sapphire, or SiC and in controlling the p-type doping [14.94]. Some research has been carried out with homoepitaxy onto GaN substrates but the very high pressures needed in

bulk crystal growth of GaN will restrict the size of available substrates.

The heteroepitaxial problems have been overcome with a two-stage growth. High-quality GaN requires growth temperatures in excess of 1000 °C, much higher than is needed for the arsenides and phosphides. At these temperatures, the nucleation onto sapphire is poor and large facetted islands grow before complete coalescence of the film occurs. This not only leads to very poor surface morphology but also a high dislocation density where the islands coalesce. The two-stage growth overcomes this by growing a uniform nucleation (or buffer) layer onto the sapphire at 600 °C. This is then heated to the normal growth temperature where a thicker GaN film is then grown. The whole nucleation process can be monitored in situ using laser reflectometry and an example is shown in Fig. 14.16 [14.95]. It can be seen that approximately 50–100 nm of GaN is grown as a smooth layer but during heating this changes to a rough layer. A remarkable part of the process is that during the high-temperature growth, shown in Fig. 14.16 by the interference oscillation, there is a recovery in the surface morphology, resulting in smooth, device-quality layers. In some cases, the GaN buffer layer is substituted with an AlN layer [14.96], but in all cases a two-stage growth process is required. The growth of device layers is achieved by controlling the bandgap through growth of the alloys GaInN or GaAlN. The bandgap of GaN is 3.4 eV, which is in the UV, so the color of the LED is determined by the alloy composition where increasing the In content will reduce the bandgap and push the emission wavelength from blue to green. However, this is not an easy alloy to form due to the different stabilities of the GaN and InN bonds [14.97]. In fact, it cannot be grown to any useful In content at temperatures above 1000 °C, so the temperature has to be reduced to around 800 °C. Even with this compromise, the different lattice parameters of GaN and InN result in poor solubility and In content above 40% is not practical. In principle, it is possible to prepare LEDs of any color by just changing the In content in the alloy, but in practice the high-brightness diodes can only be prepared from nitrides covering the blue to green portion of the spectrum. The high brightness white LEDs are formed from violet LEDs combined with phosphors to provide the color balance at the longer wavelengths [14.93]. The growth of Al-containing alloys is desirable to achieve wider bandgaps for applications such as UV LEDs and solar-blind detectors. The problems are similar to those of the In, Ga alloys but in the case of Al, Ga, the stability of AlN is much greater than that of GaN, which will tend to form AlN clusters. The growth of different alloy compositions, particularly the growth of higher In content GaInN, has stimulated some research on alternative precursors for nitrogen as ammonia pyrolysis is not very efficient at temperatures below 800 °C. One of the favorite candidates is dimethylhydrazine that will react readily with TMGa at temperatures down to 400 °C.

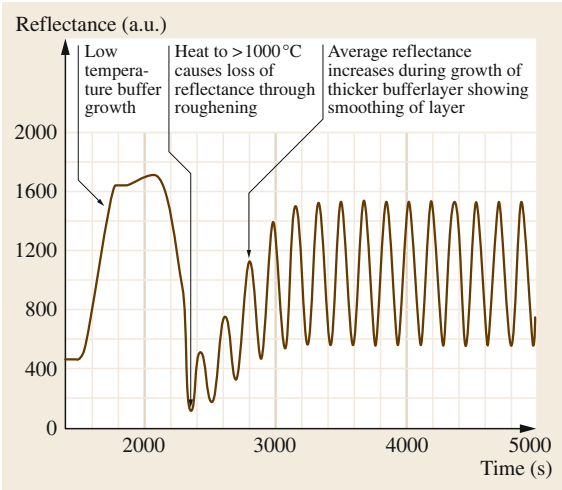


Fig. 14.16 Laser reflectometry of GaN nucleation and growth

The reaction of the ammonia with the group III alkyls to form adducts that can then polymerize is a problem associated with the high growth temperature. This requires special care in the introduction of the precursors, control of gas flows and wall temperatures. A failure to adequately control these parasitic reactions will lead to poor growth efficiency, higher defect concentration in the GaN layer, and poor dopant control. The dopants used for n-type and p-type GaN are Si from silane and Mg from dicyclopentadienylmagnesium. The n-type doping has proved to be fairly straightforward but Mg doping results in the formation of Mg–H bonds that passivate the acceptor state. This problem was solved by annealing the epitaxial films after growth to remove the hydrogen. This is possible due to the thermal stability of GaN and the high mobility of hydrogen in the lattice. A further problem with p-type doping is that the Mg acceptor has an ionization energy between 160 and 250 meV and only about 10% of the chemically introduced Mg is ionized at room temperature [14.98].

Despite the materials challenges of GaN and its alloys, MOCVD has enabled the production of a wide range of devices based on these alloys over the past two decades, from high-power transistors to laser diodes. Both these examples have required improvements in the material quality and a reduction in the relatively high dislocation densities. In fact, the potential for nitrides is enormous as the quaternary GaInNAs can be tuned to

around 1 eV with just 4% nitrogen and is a challenger for InP-based materials for the 1.3 μm telecommunications lasers and as an additional infrared junction in four junction solar cells [14.99].

14.2.6 II–VI MOCVD

The MOCVD of II–VI semiconductors is carried out at much lower temperatures than for their III–V counterparts and this has stimulated a wide range of research on alternative precursors, growth kinetics, and energy-assisted growth techniques such as photo-assisted. The basic principles are the same as for III–V MOCVD and, in general, the same reaction chambers can be used but the lower growth temperatures have led to the development of new precursors, particularly for the group VI elements. Hydrides are, in general, not used now but early work on ZnSe and ZnS used hydrogen selenide and hydrogen sulfide [14.100]. A strong prereaction occurred between the hydrides and dimethylzinc that could result in deposition at room temperature, but as with III–V MOCVD prereactions can make it difficult to control the defect chemistry and the doping. These II–VI compounds and their alloys have been investigated as blue emitter materials with similar bandgaps to GaInN. Alternatively, ZnTe is a potential green emitter and the narrower bandgap tellurides are used for infrared detectors. More recently, MOCVD of CdTe has been gaining attention for photovoltaic solar energy [14.101]. As this is polycrystalline growth on glass substrates, this will not be discussed further here but in Chap. 43 by Stuart Irvine. To date, the only commercial application of II–VI MOCVD has been for the fabrication of HgCdTe alloys for infrared detectors. However, the processes used are quite different to standard MOCVD and require different designs of reactor cells, as will be shown in the next section.

MOCVD of HgCdTe

HgCdTe is one of the few direct bandgap semiconductors suitable for infrared detection in the important 10 μm band. The alloy only has a 0.3% mismatch over the entire composition range and will cover the entire infrared spectrum from the near infrared with CdTe to the far infrared (HgTe is a semimetal so there is no lower limit to the bandgap). The main difficulty for growing HgCdTe by MOCVD has been the very high equilibrium vapor pressure of Hg over the alloy even at relatively low temperatures. For example, MBE has to be carried out at temperatures below 200 °C. A further difficulty created by the instability of HgTe is that the tellurium-rich phase boundary, which represents the minimum Hg pressure to achieve growth, has a high concentration of doubly ionized metal vacancies that make the material p-type. At typical MOCVD growth temperatures for HgCdTe, 350–400 °C, the equilibrium vapor pressure for Hg would have to be close to the SVP for liquid Hg in order to keep the metal vacancy concentration below the impurity background. This is clearly not realistic in MOCVD as the walls of the reaction chamber would have to be heated to the same temperature as the substrate to avoid mercury condensation and this, in turn, would cause pyrolysis of the precursors before they arrived at the substrate. Fortunately, it is possible to grow HgCdTe film on the tellurium-rich phase boundary where the Hg source and reactor walls can be held at around 250 °C. At this temperature, the pyrolysis of the Cd source, dimethylcadmium (DMCd), and the tellurium source, diisopropyltellurium (DIPTe) are sufficiently slow to minimize the premature reaction.

A schematic of an MOCVD reactor suitable for the growth of HgCdTe is shown in Fig. 14.17. The main features are the heating of the reactor wall and the internal source of liquid Hg. This is the only MOCVD process that uses an elemental source and is only possible because of the high vapor pressure of liquid Hg. Hg is unusual in that a higher partial pressure can be achieved using the elemental source than an organometallic source and is the only metal where this is true. Another feature of this reactor cell is the injection of DMCd past the Hg source to avoid a radical exchange reaction between DMCd and Hg [14.102].

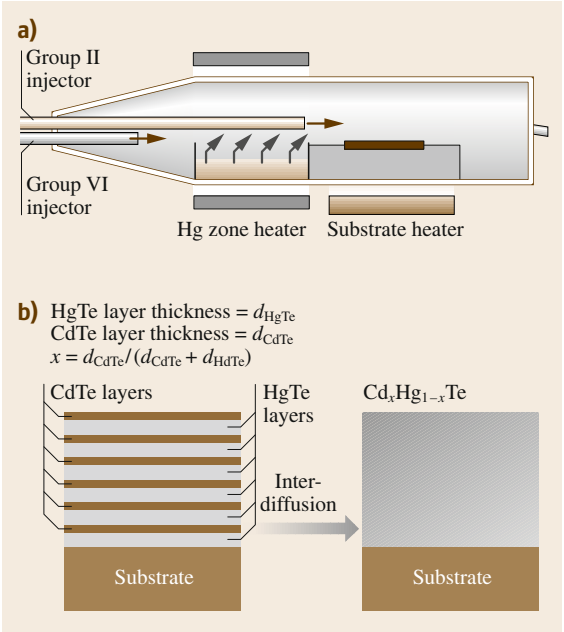


Fig. 14.17 (a) Schematic of MOCVD cell for HgCdTe growth, (b) schematic of IMP

The alloy composition needs to be precisely controlled in order to control the detection wavelength of the infrared detector. For a 10 μm detector, the proportion of Cd on the metal sublattice is 21% and needs to be controlled to better than 0.5%. The reaction rates for CdTe and HgTe require different optimum flow rates, which makes simultaneous alloy control very difficult. This problem was overcome using the interdiffused multilayer process (IMP) where alternate layers of HgTe and CdTe were grown and the flow rate optimized for each [14.103]. This process relied on a very rapid Cd/Hg interdiffusion in the HgCdTe alloy, which enables complete homogenization of the alloy at the growth temperature. The composition is now simply controlled by the relative thicknesses of HgTe and CdTe as shown schematically in Fig. 14.17b.

The IMP process has enabled controlled doping of the homogenized HgCdTe layers leading to production of heterojunction infrared detectors manufactured by MOCVD [14.104] by introducing the dopant into the CdTe layers. A concern for using dopant precursors such as AsH_3 has been the incorporation of hydrogen, which passivates the substitutional As dopant. Work by *Mitra* et al. [14.105] using the precursor *tert*-Butylarsine, followed by a 230 °C anneal under Hg vapor to remove Hg vacancies, yielded higher minority carrier lifetime than vacancy-doped HgCdTe. More recent work has concentrated on using tris(dimethylamino)arsine to minimize hydrogen incorporation. In a paper by *Hails* et al. [14.106], it was argued that a high-temperature anneal followed by a low-temperature Hg anneal helped to break the remaining As–H bonds in a HgCdTe layer as opposed to the alternative explanation of amphoteric As substitution. This procedure can result in As activation in excess of 50%.

Sulfides and Selenides

Low-temperature growth of ZnSe (below 400 °C) has been achieved using di-*tert*-butyl selenide (DTBSe) with dimethylzinc (DMZn) or the adduct DMZn.TEN. The amine, triethylamine (TEN) was first introduced to suppress gas-phase reactions with the hydrides H_2Se and H_2S. However, this adduct has the additional advantage of reducing the SVP of DMZn and making it easier to manage. The significance of keeping the growth temperature low is to avoid deep-level native defects that act as trapping sites for donor and acceptors. The p-type doping of ZnSe with nitrogen was problematic due to hydrogen passivating the dopant. Incorporation of nitrogen from a variety of precursor sources such as amines and azides up to concentrations of $10^{18}\,cm^{-3}$ could readily be achieved but active dopant concentration, in general, remained below $10^{15}\,cm^{-3}$. This was in contrast with the success of MBE growth of ZnSe that was doped from a nitrogen-plasma source. *Fujita* and *Fujita* [14.107] overcame this problem by ex-situ annealing in a similar manner to GaN but the weaker crystal lattice gives less scope for this treatment compared with GaN.

An alternative approach for low-temperature growth of ZnSe was to use short-wavelength light and UV wavelengths to illuminate the growing surface and promote surface photocatalytic reactions [14.108]. It was shown by *Irvine* et al. [14.109] that the reaction kinetics did not depend on the Se precursor, giving similar growth rates for DMSe, DESe, and DIPSe, but depended on a hydrogen radical reaction that was initiated by a surface decomposition of the group II precursor. Although this was effective in growing epitaxial films of ZnSe at temperatures well below 400 °C, it was clear that hydrogen incorporation was a natural consequence of the reaction mechanism.

MOCVD of Group II Oxides

Resurgence of interest in ZnO and related materials such as ZnMgO and ZnCdO has arisen because of the success of GaInN as a blue emitter and the potential for further developments with UV laser diodes and a solid-state replacement for domestic lighting. GaInN white light LEDs are ubiquitous but the efficiency of the phosphors would improve if they were excited with UV rather than blue photons. However, the interest in ZnO and other group II–O compounds extends beyond emitters and the flexibility in the morphology of growth makes these oxides attractive as nanowire templates [14.110].

All the early work on ZnO, MOCVD used oxygen or water vapor as the oxygen source. These react strongly at room temperature with DMZn and DEZn. Although reasonably good-quality ZnO films have been deposited with this approach, it is unlikely that it will lead to high-quality epitaxial growth or good doping control. Essentially, prereaction in all the III–V and II–VI semiconductors has been a barrier to obtaining device-quality material. The favored alternative oxygen precursors are the alcohols: isopropanol and tertiarybutanol. For higher temperatures, N_2O is a suitable precursor. In general, for epitaxial growth on sapphire or ZnO substrates it is necessary to grow at temperatures above 600 °C but for polycrystalline, transparent conducting oxides (TCOs) these precursors can react at temperatures as low as 300 °C. It is possible to readily dope ZnO n-type using TMAl but as with ZnSe it has been difficult to achieve p-type doping. Some encouraging results have been obtained using ammonia [14.111], but this work is still at an early stage of development and must be solved before electroluminescent devices

can be made. This is proving to be another class of materials where the versatility of MOCVD has a lot of potential for innovative solutions.

14.2.7 Conclusions

This section of the chapter has covered the basic principals of MOCVD and reviewed the range of III–V and II–VI semiconductor epilayers that can be grown by this method. This can be contrasted with LPE and MBE where each will have its own strengths and weaknesses for a particular material or application. The strength and the weakness of MOCVD is in its complexity. With the right precursors, it is possible to deposit almost any inorganic material but in many cases the reaction mechanisms are not well understood and the development is empirical with the researcher spoilt by a very wide choice. This is not to deny the very considerable successes that have led to major industries in compound semiconductors that has been epitomized in the past 10 years with the large-scale productionization of GaN and the plethora of large LED displays that would not have been possible without MOCVD. It has been shown in the last few years that MOCVD is very attractive for producing nanowires of compound semiconductors leading to many more possibilities for materials selection and applications. Without the pioneering work of Manasevit and Simpson who demonstrated the potential to grow so many of these materials in the early years, and the fortuitous ease with which GaAs/AlGaAs could be grown. Without this, we might not have tried so hard with the more difficult materials and hopefully we will see many more innovations in the future with MOCVD.

14.3 Molecular Beam Epitaxy (MBE)

14.3.1 Introduction and Background

MBE is conceptually a very simple route to epitaxial growth, in spite of the technology required, and it is this simplicity that makes MBE such a powerful technique. It can be thought of as a refined form of vacuum evaporation, in which neutral atomic and molecular beams from elemental effusion sources impinge with thermal velocities on a heated substrate under ultrahigh vacuum (UHV). Because there are no interactions within or between the beams, only the beam fluxes and the surface reactions influence growth, giving unparalleled control and reproducibility. Using MBE, complex structures can be grown by atomic layer with precise control of thickness, alloy composition, and intentional impurity (doping) level. UHV confers two further advantages; cleanliness, because the partial pressures of impurities are so low, and compatibility with in-situ analytical techniques – essential for understanding the surface reaction kinetics. The basic elements of an MBE system are schematically shown in Fig. 14.18. A number of reviews [14.112–114] and books [14.115, 116] have discussed the physics, chemistry, technology, and applications of MBE.

The technique that became known as MBE evolved from surface kinetic studies of the interaction of silane (SiH_4) beams with Si [14.117] and of Ga and As_2 beams with GaAs [14.118]. Cho and coworkers, who first used the term MBE, demonstrated that MBE was a viable technique for the growth of III–V material for devices, leading the way for a worldwide expansion of effort.

Much early MBE equipment had a single vacuum chamber for loading, deposition, and analysis, which led to prolonged system pumpdown between growths. The technique required significant improvements in vacuum conditions before very high quality thin films were grown. *Cho* surrounded the effusion cells with a liquid nitrogen cryopanel to give thermal isolation and reported the MBE growth of thin films of n- and p-type GaAs for device purposes [14.119]

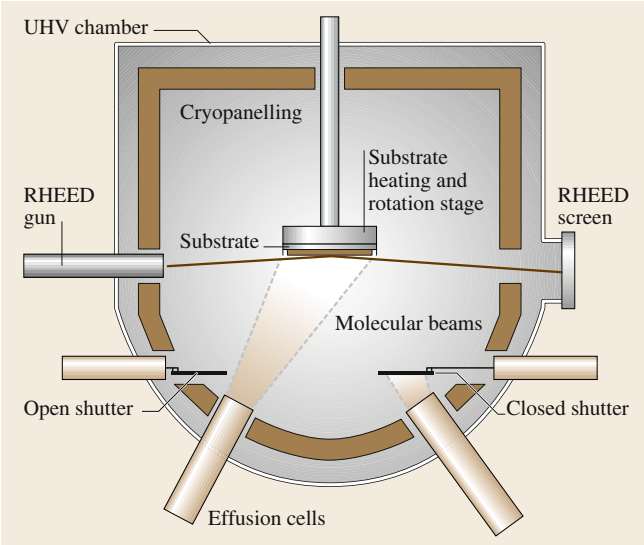


Fig. 14.18 Schematic diagram of an MBE growth chamber, showing the effusion cells and shutters, the substrate stage, and the arrangement of the RHEED system

and of GaAs/AlGaAs heterostructures [14.120]. The introduction of a substrate-exchange load lock [14.121] drastically reduced pumpdown times and reduced contamination of the deposition chamber. The installation of extensive internal liquid-nitrogen-cooled cryopanels [14.122] substantially increased pumping of oxygen-containing species and permitted the growth of AlGaAs with superior quality. *Tsang* [14.123] demonstrated lasers with threshold current densities superior to those grown by LPE. Uniform growth over a 2-inch diameter wafer was achieved by the introduction of a rotating substrate holder capable of 5 rpm [14.68]. Advanced forms of these features are now standard on commercial MBE systems, many of which feature a modular design. The technology of MBE is now mature, with increasing numbers of ever larger high-throughput, multiwafer production MBE machines in widespread use since the early 1990s.

MBE has been used to grow a wide range of materials, including semiconductors, superconductors, metals, oxides, nitrides, and organic films. In almost all cases, there is a drive to produce structures with ever smaller dimensions, whether for higher performance devices, quantum confinement or, more recently, nanotechnology. This is longest established in III–V semiconductors, from the GaAs/AlGaAs superlattice [14.124] through quantum wells (QWs) and modulation doping to quantum wires and quantum dots [14.125]. Such low-dimensional structures form the basis of the QW lasers and p-HEMTs produced in huge volumes by MBE for optoelectronic and microwave applications. The combination of precise growth control and in-situ analysis makes MBE the pre-eminent technology to meet such demands.

14.3.2 Reaction Mechanisms

One of the first and most important of the in-situ analytical techniques to be used in MBE is reflection high-energy electron diffraction (RHEED). From an early stage [14.126], it was used to determine the surface structure of the clean substrate and growing layer. This revealed that, in general, all surfaces are reconstructed, that is, have a lower symmetry than the bulk, and Cho was the first to propose that the twofold periodicity observed in the [$\bar{1}$ 10] direction on the (001) surface was the result of dimerization of As atoms on the arsenic-terminated surface, which was confirmed many years later by scanning tunneling microscopy [14.127]. RHEED is a forward-scattering technique and therefore more compatible with the MBE arrangement of normally incident fluxes than the back-scattering geometry of LEED.

A further application of RHEED is the in-situ measurement of growth rate. It was found [14.128, 129] that the intensity of any diffraction feature oscillated with a period corresponding to the growth of a single monolayer (ML), that is, a layer of Ga + As, in the [001] direction on a (001) substrate. These oscillations arise from surface morphological changes during two-dimensional (2-D) layer-by-layer growth, the Frank–van der Merwe mode, and a typical result for GaAs is shown in Fig. 14.19. The exact origin of the oscillations is still a subject of debate [14.114] but the technique was found to be applicable to many other material systems, including elemental semiconductors, metals, insulators, superconductors, and even organic compounds.

Thus RHEED provided information on surface reconstruction and quantitative measurements of growth dynamics, which could be combined with that from other in-situ techniques and related to theoretical treatments. In the early 1970s, the surface chemistry involved in the MBE growth of III–V compounds was studied extensively using a combination of RHEED and modulated molecular beam mass spectrometry [14.130] or temperature-programmed desorption [14.131].

Detailed information was obtained on surface reactions involving gallium and arsenic. It was found that the dissociation of GaAs results in the desorption of As_2, and not of As or As_4 [14.132], and that a significant amount of desorption takes place at MBE growth temperatures. The sublimation of elemental arsenic, as from an effusion cell, results in the formation of As_4 alone [14.130]. It was also shown that growth from Ga + As_2 is a first-order reaction (Fig. 14.20), whereas growth from Ga + As_4 is second order. These results suggest that in MBE growth Ga sticks to available As sites and chemisorption of As_2 occurs on available Ga atoms. Chemisorption of As_4 occurs with 2 As_4 molecules interacting on adjacent Ga atoms. The sticking coefficient of As_4 is observed to be less than or equal to 0.5, whereas the sticking coefficient of As_2 can be equal to 1, in agreement with this model. This implies that maximum coverage will be less than 100%, since single Ga sites cannot be occupied, and that As vacancies will be introduced into material grown using As_4. This was thought to be responsible for the higher deep level concentrations observed, which were found to be reduced when As_2 was used or at higher growth temperatures. It is now known that above 580 °C, As_4 dissociates to As_2 on the surface [14.133].

Detailed models were established for the growth of GaAs [14.134, 135] and AlAs [14.136]. With minor modifications, these models are also valid for ternary alloys such as AlGaAs [14.137]. It was established that growth rate depends solely on the net group III flux (i. e., incident flux minus desorbing flux), and that

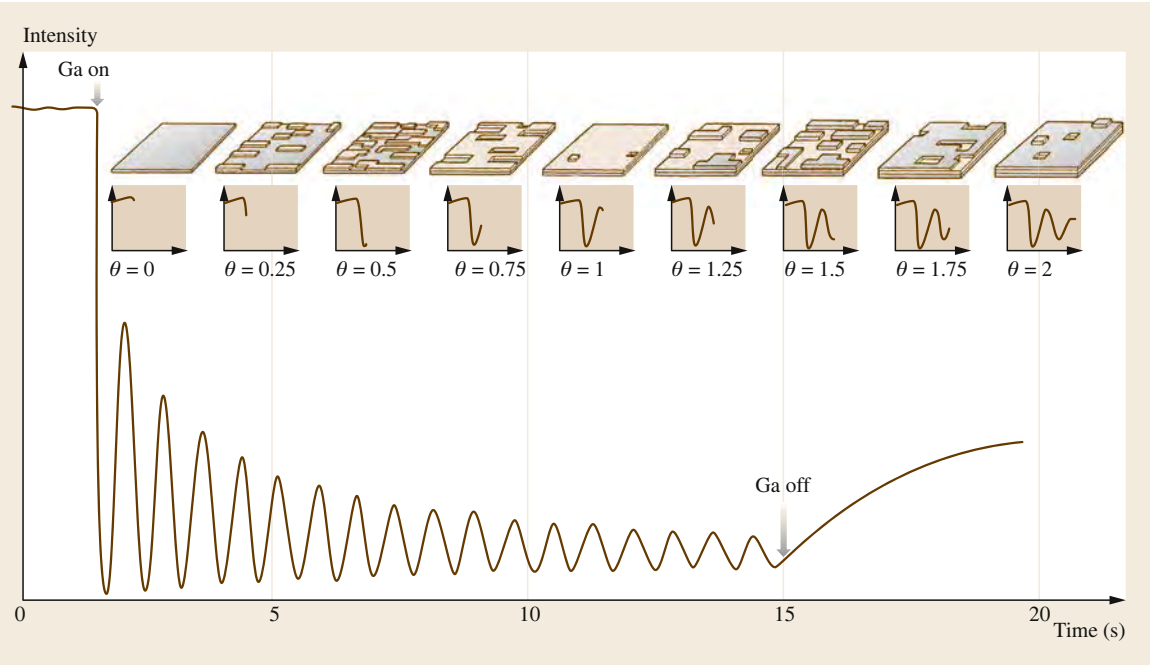


Fig. 14.19 Diagram illustrating the development of RHEED intensity oscillations during 2-D layer-by-layer growth, together with an actual experimental plot. (Courtesy of B.A. Joyce)

ternary alloy composition can be controlled by adjusting the ratio of the group III fluxes provided account is taken of the thermal stability of both of the binary compounds that make up the alloy. The group V element need only be supplied in excess. The situation is more complicated for alloys containing both arsenic and phosphorus, since the presence of one group V element influences the sticking coefficient of the other. *Foxon* et al. [14.138] found that phosphorus has a much lower incorporation probability than arsenic.

More recent studies have made use of in-situ STM and more powerful theoretical treatments to consider nucleation and growth at the atomistic level, but the basic models are still sound.

14.3.3 MBE Growth Systems

The UHV system required for MBE is of conventional stainless steel construction, with an ultimate or background vacuum of less than 5×10^{-11} torr achievable with a clean system after baking, and with the liquid-nitrogen cryopanels filled. At such a pressure, molecules have a typical mean free path of 10^6 m and so only suffer collisions with the internal surfaces of the system. It would take several hours to build up a monolayer of impurity on the wafer surface.

Oil-free pumping is used to eliminate the possibility of contamination by hydrocarbon backstreaming, typically rough pumping is with sorption pumps and UHV pumping is with ion pumps and titanium sublimation pumps. Diffusion pumps or turbomolecular pumps can be used, however, provided suitable cold-traps are fitted, and such pumps are required for the higher gas loads involved in GSMBE and CBE. A two- or three-stage substrate entry load-lock and preparation chamber isolated by gate valves is used to minimize the system

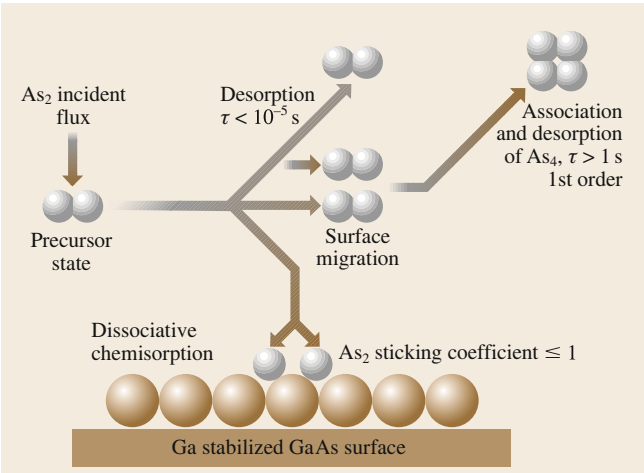


Fig. 14.20 Surface chemistry of the growth of GaAs from As_2 and Ga. (Courtesy of B.A. Joyce)

exposure to air. In modular systems, further deposition and analysis chambers may be added to the system and samples are transferred via the preparation chamber under UHV conditions.

The growth chamber substrate stage is surrounded by a large liquid-nitrogen-cooled cryopanel, which has a high pumping speed for H_2O, CO, O_2, and other condensable species. This is arranged so that the heated (500–700 °C) substrate is not directly exposed to thermal sources other than the molecular beams themselves, and impurities emanating from any other source can only reach the substrate after suffering at least one collision with, and probable adsorption by, a surface at liquid-nitrogen temperatures. The cryopanel also reduces contamination arising from outgassing from the walls of the chamber that are exposed to radiation from the effusion cells.

Control over the composition and doping levels of the epitaxial layers is achieved by precise temperature control of the effusion cells and the use of fast-acting shutters in front of these cells. In most systems, the growth rate is about 1 monolayer per second, and the shutter operation time of 0.1 s thus corresponds to less than a monolayer of growth. A growth rate of 1 monolayer per second is closely approximate to 1 μm/h for GaAs/AlGaAs.

Although often referred to as K-cells, the solid-source effusion cells used in MBE growth have a large orifice so as to obtain a high flux at reasonable temperatures. A true Knudsen cell has a very small outlet orifice compared to the evaporating surface so that an equilibrium vapor pressure, typically 10^{-3} torr, is maintained within the cell. The diameter of this orifice is less than one-tenth of the molecular mean free path, which is typically several centimeters. Under these conditions, a near approximation to ideal Knudsen effusion is obtained from the cell, giving molecular flow with an approximately cosine distribution. The flux from such a cell can be calculated quite accurately, but a high temperature is required to produce a reasonable growth rate, for example a gallium Knudsen cell would need to be at 1500 °C to produce the same flux as an open-ended effusion cell at 1000 °C. The lower operating temperature helps to reduce impurities in the flux and puts a lower thermal load on the system. The beam from an open-ended cell may not be calculable with any degree of accuracy but it is highly reproducible. Once calibrated via growth rate, normally by in-situ measurement, the flux can be monitored using an ion gauge located on the substrate stage.

A number of effusion cells can be fitted to the growth chamber, generally in a ring facing toward the substrate with the axis of each cell at an angle of 20–25 °C to the substrate normal. Simple geometrical considerations, therefore, dictate the best possible uniformity that can be achieved with a stationary substrate [14.139]. Associated with each cell is a fast action refractory metal shutter with either pneumatic or solenoid operation.

Uniformity of growth rate for a binary compound can be achieved by rotating the substrate at speeds as low as a few rpm, but compositional uniformity of ternary or quaternary alloys requires rotation of the substrate at speeds of up to 120 rpm (normally rotation is timed so that one rotation corresponds to the growth of one monolayer). The mechanical requirements for a rotating substrate stage in a UHV system are quite demanding, as no conventional lubricants can be used on the bearings or feedthroughs and yet lifetime must exceed several million rotations. Magnetic rotary feedthroughs have largely replaced the earlier bellows type.

The need to rotate the substrate to give uniformity also leads to complications in substrate-temperature measurement. The substrate is heated by radiation from a set of resistively heated tantalum foils behind the substrate holder, and both the heater and the thermocouple are stationary. Without direct contact between the thermocouple and the wafer, the *indicated* thermocouple temperature will be very different from the *actual* substrate temperature. Some form of calibration can be obtained by using a pyrometer, although problems of window coating, emissivity change, and of substrate transparency below the bandgap impose limits on the accuracy of such measurements. Alternatively, a number of *absolute* temperature measurements can be obtained by observing transitions in the RHEED pattern, which occur at reasonably fixed temperatures. However, such transitions occur in the lower temperature range and extrapolation to higher growth temperatures is not completely reliable. If the substrate is indium bonded to a molybdenum block (using the surface tension of the indium to hold the substrate), then inconsistencies in wetting can lead to variations in temperature across the substrate. Most modern systems and all production machines use *indium-free* mounting, which avoids these problems. However, the substrate is transparent to much of the IR radiation from the heater, putting a higher thermal load on the system.

The substrate-preparation techniques used prior to MBE growth are very important, as impurities on the surface provide nucleation sites for defects. Historically, various chemical clean and etch processes were used, but wafers are now usually supplied *epi-ready*, with a volatile oxide film on the surface that protects the surface from contamination and can be thermally removed within the UHV chamber. RHEED is used to confirm the cleanliness of the surface prior to growth.

Historically, one of the major problems in MBE was the presence of macroscopic defects, with a typical density of 10^3–$10^5\,cm^{-2}$, although densities below $300\,cm^{-2}$ were reported for ultraclean systems [14.140]. Defects are generally divided into two types: small hillocks or pits and oval defects. Such defects are a serious obstacle to the growth of material for integrated circuits, and considerable effort was devoted to the problem. Oval defects are microtwin defects originating at a local imperfection, oriented in the [14.141] direction and typically 1–10 μm in length [14.142]. There are several possible sources of these defects, including foreign impurities on the substrate surface due to inadequate substrate preparation or to oxides from within the system, and possibly from the arsenic charge or the condensate on the cryopanels. The fact that oval defects were not seen when graphite crucibles were used but were common with PBN crucibles suggests that gallium oxide from the gallium melt is a major source of such defects since oxides would be reduced by the graphite. *Chai* and *Chow* [14.143] demonstrated a significant reduction in defects by careful charging of the gallium source and prolonged baking of the system. The irregular hillocks and pits seen in MBE-grown material were probably produced by microdroplets of gallium spitting from the effusion cells on to the substrate surface [14.144]. Gallium spitting can be caused by droplets of Ga that condense at the mouth of the effusion cell, fall back into the melt and explode, ejecting droplets of liquid Ga, or by turbulence in the Ga melt due to uneven heating that causes a sudden release of vapor and droplets.

Continuous developments in the design of group III effusion cells for solid-source MBE have largely eliminated the problem of macroscopic defects. Large-area Ta foil K-cell heaters reduced the uneven heating of the PBN crucible; the use of a *hot-lipped* or two-temperature group III cell, designed with a high-temperature front end to eliminate the condensation of gallium metal at the mouth of the cell, significantly reduced the spitting of microdroplets. Combined with careful procedures and the use of an arsenic cracker, cell defect densities as low as $10\,cm^{-2}$ have been reported [14.145].

There have also been developments in the design of group V cells. A conventional arsenic effusion cell produces a flux of As_4 [14.134], but the use of a thermal cracker to produce an As_2 flux resulted in the growth of GaAs with better optical properties and lower deep-level concentrations [14.146]. In the case of phosphorus, growth from P_2 was strongly preferred to that from P_4 for several reasons [14.147] and this was one reason behind the development of GSMBE described below. However, the use of phosphine requires suitable pumping and safety systems. The group V cracker cell has two distinct zones. The first comprises the As or P reservoir and produces a controlled flux of the tetramer; this passes through the second – high-temperature – zone, where dissociation to the dimer occurs. Commercial high-capacity cracker cells, some including a valve between the two zones to allow fast switching of group V flux, have been developed for arsenic and phosphorus.

Commercial MBE systems have increased throughput with multiwafer substrate holders, cassette loading and UHV storage, and preparation chambers linked to the growth chamber with automated transfer, while increased capacity effusion cells have reduced the downtime required for charging. In some cases, additional analytical and processing chambers have been added to permit all UHV processing of the device structure.

14.3.4 Gas Sources in MBE

A number of MBE hybrids were developed that combined the advantages of UHV deposition and external gas or metal-organic sources to produce a versatile technique that has some advantages over MBE and MOVPE. These techniques utilize the growth chambers developed for MBE and pumping systems with a high continuous throughput, typically liquid-nitrogen-trapped diffusion pumps or turbomolecular pumps. Layers are deposited from molecular beams of the precursor materials introduced via gas-source cells that are essentially very fine leak controllers. As in MBE, there are no interactions within or between beams and the precise control of beams using fast-acting gas-line valves is, therefore, translated into precise control of the species arriving at the substrate. Shutters are not generally required; atomically sharp interfaces and monolayer structures can be defined as a consequence of submonolayer valve switching times.

Panish [14.148] investigated the use of cracked arsine and phosphine in the epitaxial growth of GaAs and InP, later extending this work to the growth of GaInAsP, and suggested the name gas source MBE (GSMBE). The major advantage of gaseous group V sources was that the cracker cells produced controllable fluxes of the dimers As_2 and P_2, giving improved control of the As:P ratio. The MBE growth of high-quality GaAs from cracked arsine and elemental gallium was demonstrated at the same time by *Calawa* [14.149].

The extension to gaseous group III sources was made by *Veuhoff* et al. [14.150], who investigated the MOCVD of GaAs in a simple MBE system using trimethylgallium (TMGa) and uncracked arsine. Further study showed that cracking of arsine at the substrate surface was negligible, leading to the conclusion

that unintentional cracking of the arsine had taken place in the inlet system. The acronym metal-organic MBE (MOMBE) was used to describe GaAs growth from TMGa and cracked arsine in a modified commercial MBE system [14.151].

The growth of device-quality GaAs, InGaAs, and InP from alkyl sources of both group III and group V elements was demonstrated by *Tsang* [14.152], who used the alternative acronym chemical beam epitaxy (CBE). The use of group V alkyls, which had much poorer purity than the hydrides, was undertaken for safety reasons [14.153]. Material quality was improved when cracked arsine and phosphine were used [14.154]. RHEED observations indicated that reconstructed semiconductor surfaces could be produced prior to growth as for MBE [14.155] and GaAs/AlGaAs quantum well structures were demonstrated that were comparable with those grown by MBE or MOVPE [14.156].

Almost all this work was with III–Vs, where MOMBE/CBE was seen to have several significant advantages over MBE while retaining many of its strengths, including in-situ diagnostics. The use of vapor group III sources would avoid the morphological defects associated with effusion cells, and higher growth rate and greater throughput could be achieved. Both group III and group V sources were external, allowing for easy replacement without the need to break vacuum. Flux control with mass flow controllers (MFCs) and valves would improve control of changes in composition or doping level, since flow could be changed faster than effusion-cell temperature. Abrupt changes could be thus achieved that would require switching between two preset effusion cells in MBE (a problem when the number of cells was limited by geometry). It also offered improved long-term flux stability and greater precursor flexibility. As this was still a molecular-beam technique, precise control of layer growth and abrupt interfaces would be retained, without any of the gas-phase reactions, boundary-layer problems, or depletion of reagents associated with MOVPE. Other advantages included improved InP quality using a P_2 flux, lower growth temperatures, and selective area epitaxy.

There was, however, a price to pay in system complexity, with the need for gas handling and high-volume pumping arrangements added to the expensive UHV growth chamber. These would have been acceptable if CBE had demonstrated clear advantages but there were a number of other issues. The standard Al and Ga precursors used in MOVPE (trimethylaluminum and trimethylgallium) produced strongly p-type material when used in CBE, due to the incorporation of C as an acceptor. Triethylgallium proved to be a viable Ga source, but alloy growth was more complicated; no universally acceptable Al source was found, while InGaAs growth was found to be strongly temperature dependent. The surface chemistry involved with metal-organic sources proved complex and the temperature dependence of surface reactions not only restricted growth conditions, but also had a serious impact on uniformity and reproducibility, particularly for quaternary alloys such as GaInAsP [14.157]. The lack of suitable gaseous dopant sources, particularly for Si, was a further handicap [14.158], but the deliberate use of C for p-type doping proved a success and this was transferred to MBE and MOVPE. Carbon diffuses significantly less than the *standard* MBE and MOVPE dopants, Be and Zn, respectively [14.159], and proved an ideal dopant for thin highly doped layers such as the base region in heterojunction bipolar transistors (HBTs) and for p-type Bragg reflector stacks in vertical cavity surface emitting laser structures (VCSELs).

GSMBE remains important, not in the III–V field where, with some exceptions [14.160], the development of high-capacity group V cracker cells provided an easier route to an As_2 and P_2 flux, but in the III-nitride field. There are two major routes to nitride MBE: active nitrogen can be supplied by cracking N_2 in an RF or ECR plasma cell, or ammonia can be injected and allowed to dissociate on the substrate surface. In contrast, CBE has not demonstrated a sufficient advantage over its parent technologies to be commercially successful, particularly as both MOVPE and MBE have continued to develop as production techniques.

14.3.5 Growth of III–V Materials by MBE

Although they were among the earliest materials to be grown by MBE, GaAs-based alloys retain great importance, with MBE supplying materials for the mass production of optoelectronic and microwave devices and leading research into new structures and devices.

GaAs/AlGaAs

AlGaAs is an ideal material for heterostructures, since AlAs has a greater bandgap than GaAs and the two have negligible mismatch ($\approx 0.001\%$). The growth rate in MBE depends on the net group III flux, with $1\,\mu m/h$ corresponding to a flux of 6.25×10^{14} Ga (or Al) atoms $cm^{-2}\ s^{-1}$. At low-growth temperatures, all incident group III atoms are incorporated into the growing film, together with sufficient arsenic atoms to maintain stoichiometry, and excess arsenic atoms are desorbed. However, III–V compounds are thermally unstable at high temperatures. Above $\approx 600\,K$ [14.134] arsenic is preferentially desorbed, so an excess arsenic flux is required to maintain stoichiometry. At higher temperatures, loss of the group III element becomes significant so that the growth rate is less than would be expected for

the incident flux. This is particularly important for the growth of AlGaAs, where growth temperatures above 650 °C are generally used to give the best optical properties. The Ga flux has to be significantly increased above that used at lower temperatures in order to maintain the required composition of the alloy.

Typical growth temperatures for MBE of GaAs are in the range 580–650 °C and material with high purity and low deep-level concentrations has been obtained in this temperature range [14.161, 162]. The commonly used dopants, Be (p-type) and Si (n-type), show excellent incorporation behavior and electrical activity at these temperatures and at moderate doping levels. As was noted above, for highly doped layers Be has largely been replaced by C, which diffuses significantly more slowly. At doping levels above $\approx 5\times10^{17}$, Si occupies both Ga (donor) and As (acceptor) sites, producing electrically compensated material with a consequent reduction in mobility. It is still predominantly a donor, however, and is the best available n-type dopant. The electrical properties of GaAs also depend on the As/Ga flux ratio, since this influences the site occupancy of dopants. The optimum As/Ga ratio is that which just maintains As-stabilized growth conditions, which can be determined using RHEED observations of surface reconstruction.

MBE-grown GaAs is normally p-type, the dominant impurity being carbon [14.163]. The carbon concentration was found to correlate with CO partial pressure during growth [14.164] – CO is a common background species in UHV, being synthesized at hot filaments. The lowest acceptor levels commonly achieved are of the order of $5\times10^{13}\,\mathrm{cm}^{-3}$ and such layers can be lightly doped to give n-type material with high mobilities. However, very high purity GaAs has been produced by adjusting the operating conditions for an arsenic cracker cell [14.141], which suggests that carbon contamination originates from hydrocarbons in the As charge. Unintentionally doped GaAs was n-type with a total impurity concentration of $< 5\times10^{13}\,\mathrm{cm}^{-3}$ and a peak mobility of $4\times10^{5}\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ at 40 K, the highest reported for n-type GaAs. The 77 K mobility of $> 200\,000\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ is comparable with that for the highest purity GaAs grown by LPE [14.165].

MBE is capable of the growth of such high-quality material for structures whose physical dimensions are comparable to the wavelength of an electron (or hole) so that quantum size effects are important. Such structures have typical layer thicknesses from 100 Å down to 2.8 Å, the thickness of a monolayer of GaAs, and have doping and composition profiles defined on an atomic scale. These low-dimensional structures have become very important in III–V device technology, one example being the modulation-doped heterojunction.

The precise control of growth that is possible is demonstrated by the very high mobilities obtained for modulation-doped structures grown by MBE. The modulation-doped GaAs/AlGaAs heterojunction has a band structure that causes carriers from the highly doped AlGaAs to be injected into a thin undoped region in the GaAs (Fig. 14.21a). Since the carriers are now separated from the donor atoms that normally scatter them, and confined in a quantum well as a two-dimensional electron gas (2DEG), very high mobilities can be achieved. By suitable design of the structure and the use of lightly doped AlGaAs, a GaAs/AlGaAs 2DEG structure was produced with a peak mobility of $1.1\times10^{7}\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ at 1.3 K and 4 K mobility of $4.5\times10^{6}\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ and with a sheet carrier concentration of $10^{12}\,\mathrm{cm}^{-2}$ [14.166]. GaAs layers grown in the same machine typically exhibited 77 K mobilities of the order of $100\,000\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$. Similar results can be obtained by delta doping the AlGaAs, that is, by confining the Si donor atoms to a single monolayer in the

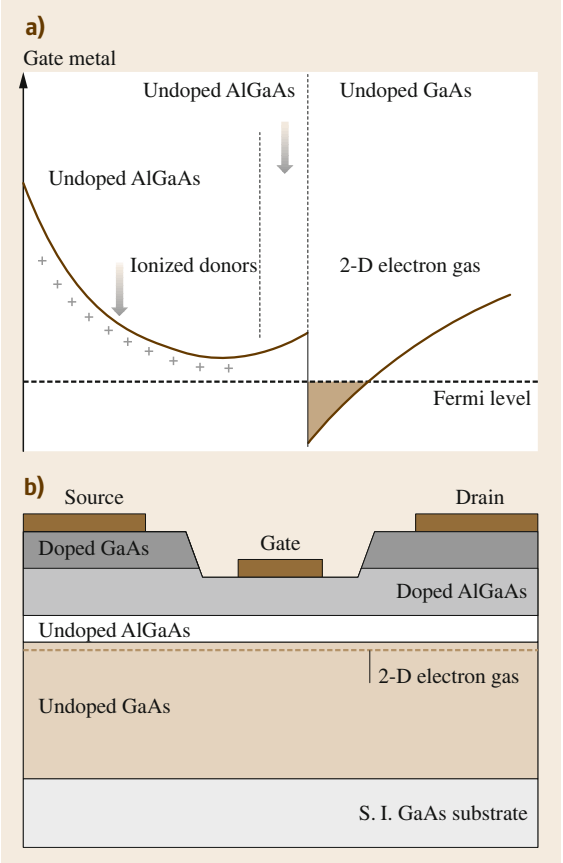


Fig. 14.21a,b Schematic diagrams of the band structure **(a)** and physical arrangement **(b)** of the high electron mobility transistor (HEMT)/modulation doped FET

AlGaAs separated by a few nanometer from the well, a technique with a wide range of applications [14.167]. The major application of this structure for device purposes is the high electron mobility transistor (HEMT), also known as the modulation-doped FET (MODFET), an FET in which the carriers are confined to the two-dimensional layer (Fig. 14.21b). The main advantage of the HEMT is not the increase in mobility, which is modest at room temperature for a practical device, but the very low noise when it is operated as a microwave amplifier, due to the reduction in impurity scattering. The HEMT is an essential component of many microwave systems, including mobile phones and satellite TV receivers.

MBE became used as a production technology for GaAs-based devices because of the excellent uniformity and reproducibility possible, and because the extensive UHV load-lock system of a production machine permits a large number of runs to be undertaken on a continuous basis. This was demonstrated as early as 1991 by a number of manufacturers producing both HEMTs and GaAs/AlGaAs lasers [14.168]. By 1994 some 5 million MQW lasers for compact disc applications were produced per month by MBE. Other devices included HEMTs and InAs Hall sensors. A comparison of production costs made at this time [14.169] showed that capital costs for MBE were very similar to those for MOVPE, which required costly safety systems for the hydride gases, while MBE costs per wafer were somewhat less than for MOVPE. The growth in demand for HEMTs for mobile phones drove a further expansion in production MBE, although MOVPE remained dominant in InP-based optoelectronic devices.

InGaAs

InGaAs is generally grown at lower temperatures than GaAs/AlGaAs because of the higher rate of In desorption. Both InGaAs and the quaternary InGaAsP can be lattice matched to InP substrates; heterostructures in these materials form the basis of the 1.3 and 1.55 μm optoelectronic devices used in telecommunications. InAs has a lattice mismatch of $\approx 7\%$ with GaAs; thin films of InGaAs can be deposited pseudomorphically but as the In content is increased the critical thickness falls so that little more than a ML of InAs can be grown in this way. Thicker films will be relaxed by the formation of dislocations, limiting their usefulness for devices. The pseudomorphic HEMT (or pHEMT) replaces the GaAs channel of the conventional HEMT with a strained InGaAs channel in order to take advantage of the greater carrier confinement and superior electron transport properties. The maximum In content of the channel is limited by the need to prevent relaxation.

Under certain growth conditions, InAs or InGaAs islands are spontaneously formed. It is generally accepted that the growth of InAs on GaAs (001) follows a version of the Stranski–Krastanov mode, which implies that following the deposition of ≈ 1.7 ML of InAs in a 2-D pseudomorphic form (sometimes referred to as the wetting layer), coherent 3-D growth is initiated by a very small increment (≤ 0.1 ML) of deposited material, to relax the elastic strain introduced by the lattice mismatch. The QDs rapidly reach a saturation number density, which is both temperature and In flux dependent, with a comparatively narrow size (volume) distribution. The actual process is rather more complicated, not least as a result of alloying with the GaAs substrate, and is the subject of much current research [14.114].

These islands can be embedded in a layer of GaAs to form self-assembled QDs, which have become a topic of immense interest due to the potential application of QDs in a wide range of devices, especially lasers. The volume fraction of QDs in an active layer can be increased by building up a 3-D array; the strain field induced around each dot influences not only the inter-dot spacing but also the capping layer growth so that subsequent layers of dots are aligned ([14.125] and references therein). QD lasers offer a route to long-wavelength emission from GaAs-based devices [14.170].

III–Nitrides

The growth of group III nitrides has been dominated by MOVPE since the demonstration of a high-brightness blue-emitting InGaN-AlGaN double heterostructure LED by Nichia [14.171] and the subsequent development of other optoelectronic devices, including laser diodes also emitting in the blue [14.172]. MBE has made a significant contribution to more fundamental studies and to the growth of nitrides for high-power and microwave devices. Once again, the wide range of in-situ diagnostic techniques available has been important.

For heteroepitaxial growth on the most commonly used substrates, sapphire and silicon carbide, several parameters strongly influence the quality of material produced. These include substrate cleaning, initial nitridation, the nucleation and coalescence of islands involved in the low-temperature growth of a buffer layer, and subsequent annealing at a higher temperature. The polarity (nature of the outermost layer of atoms) of {0001}-oriented hexagonal structure films also has a crucial influence on material quality, but both N- or Ga-polarity can occur with MBE growth on sapphire substrates [14.173]. Under typical growth conditions with MOVPE, however, Ga-polarity material is exclusively produced.

Several of these problems can be resolved using GaN templates obtained by growing thick layers on to suitable substrates using MOVPE and then exploiting the advantages of MBE to produce the functional layer on the GaN template. These advantages include well-controlled layer-by-layer growth and lower growth temperatures than those used for MOVPE so that InGaN phase separation and In desorption are less problematic and precise quantum wells can be grown. No post-growth thermal annealing is required to activate the p-type dopant. In this way films have been produced with smooth surface morphology and high performance, although MBE was still limited to low-power LEDs until the recent demonstration of laser diodes [14.174].

In the use of active nitrogen from plasma sources, the III/V flux ratio at the substrate during growth is also a critical parameter. GaN layers grown with a low III/V flux ratio (N-stable growth) display a facetted surface morphology and a tilted columnar structure with a high density of stacking faults. Smooth surfaces are only obtained under Ga-rich conditions, where not only is there a dramatic reduction of surface roughness, but significant improvements in structural and electrical properties are also observed. This is, of course, the exact opposite of the growth of most III–V compounds, such as GaAs. In the case of nitrides, it is thought that Ga-rich conditions (close to the point where Ga droplets are formed) promote step flow growth, whereas N-stable growth promotes the nucleation of new islands. In contrast, growth from NH_3 is smoother under N-rich conditions [14.175].

III–V-Nitrides

The *dilute nitrides* are III–V–N materials such as GaAsN and GaInNAs, where the N concentration is $\leq 2\%$. Replacing a small fraction of As atoms with smaller N atoms reduces both the lattice constant, and the bandgap. Adjusting the composition of GaInNAs allows the bandgap, band alignment, lattice constant and strain to be tailored in a material that can be lattice matched with GaAs [14.176]. This offers strong carrier confinement and thermal stability compared to InP-based devices operating at 1.3 and 1.55 μm, and allows GaAs VCSEL technology to be exploited at these wavelengths [14.177].

These materials are grown in a metastable regime at a low-growth temperature because of the miscibility gap in the alloys, so a less-stable precursor than NH_3 is needed. A nitrogen-plasma source provides active N without the incorporation of hydrogen during growth associated with hydride sources, thus avoiding the deleterious formation of N–H bonds. Material with excellent crystallinity and strong PL at 1.3 μm can be obtained by optimizing growth conditions and using post-growth rapid thermal annealing [14.178]. A number of challenges remain, particularly in the higher N material required for longer wavelengths, including the limited solubility of N in GaAs and nonradiative defects caused by ion damage from the N plasma source. These challenges may be met by using GaInNAsSb the addition of Sb significantly improves the epitaxial growth and the material properties, and enhanced luminescence is obtained at wavelengths longer than 1.3 μm [14.179].

14.3.6 Conclusions

MBE, historically seen as centered on GaAs-based electronic devices, has broadened its scope dramatically in both materials and devices. In addition to the materials described above, MBE has been used to grow epitaxial films of a wide range of semiconductors, including other III–V materials such as InGaAsP/InP and GaAsSb/InAsSb; silicon and silicon/germanium; II–VI materials such as ZnSe; dilute magnetic semiconductors such as GaAs:Mn [14.180] and other magnetic materials. It has also been used for the growth of metals, including epitaxial contact for devices, oxides [14.181]; and organic films [14.182]. Two clear advantages possessed by MBE are the wide range of analytical techniques compatible with a UHV system and the precise control of growth to less than a monolayer, which confer unrivalled capability in the growth of quantum dots and other nanostructures.

Acknowledgments. The author P.C. wishes to thank the management of Leonardo MW Ltd for permission to publish this chapter.

References

14.1 P. Capper, M. Mauk (Eds.): *Liquid Phase Epitaxy of Electronic, Optical and Optoelectronic Materials* (Wiley, Chichester 2007)

14.2 P. Capper, P. Rudolph (Eds.): *Crystal Growth Technology: Semiconductors and Dielectrics* (Wiley-VCH, Weinheim 2010)

14.3 H.J. Scheel, P. Capper (Eds.): *Crystal Growth Technology: From Fundamentals and Simulation to Large-Scale Production* (Wiley-VCH, Weinheim 2008)

14.4 D. Elwell, H.J. Scheel: *Crystal Growth from High-Temperature Solutions* (Academic Press, New York 1975)

14.5 H. Nelson: RCA Review **24**, 603 (1963)
14.6 R.L. Moon: J. Cryst. Growth **170**, 1 (1997)
14.7 H.J. Scheel: In: *The Technology of Crystal Growth and Epitaxy*, ed. by H.J. Scheel, T. Fukuda (Wiley, Chichester 2003)
14.8 P. Capper, T. Tung, L. Colombo: In: *Narrow-Gap II-VI Compounds for Optoelectronic and Electromagnetic Applications*, ed. by P. Capper (Chapman Hall, London 1997)
14.9 M.B. Panish, I. Hayashi, S. Sumski: Appl. Phys. Lett. **16**, 326 (1970)
14.10 M.G. Astles: *Liquid Phase Epitaxial Growth of III-V Compound Semiconductor Materials and Their Device Applications* (IOP Publishing, Bristol 1990)
14.11 T. Ciszek: In: *The Technology of Crystal Growth and Epitaxy*, ed. by H.J. Scheel, T. Fukuda (Wiley, Chichester 2003)
14.12 J. Poortmans: Epitaxial thin film crystalline silicon solar cells on low cost silicon carriers. In: *Thin Film Solar Cells: Fabrication, Characterization and Applications*, ed. by J. Poortmans, V. Arkhipov (Wiley, Chichester 2006)
14.13 K. Peter, R. Kopecek, P. Fath, E. Bucher, C. Zahedi: Sol. Energy Mater. Sol. Cells **74**, 219 (2002)
14.14 F. Abdo, A. Fave, H. El Omari, M. Lemiti, C. Bernard, A. Pisch: Moroc. J. Condens. Matter **11**, 15 (2009)
14.15 M.I. Alonso, K. Winer: Phys. Rev. B **39**, 10056 (1989)
14.16 S. Dost: Growth of bulk crystals of semiconductors by electroepitaxy, Proc. 17th Int. Conf. Cryst. Growth Epitaxy, Warsaw (2013)
14.17 Y. Liu, M.D. Deal, J.D. Plummer: Appl. Phys. Lett. **84**(14), 2563 (2004)
14.18 T. Tanaka, M. Tanaka, M. Itakura, T. Sadoh, M. Miyao: Thin Solid Films **518**(6), S170 (2010)
14.19 K. Toko, T. Tanaka, T. Sadoh, M. Miyao: Thin Solid Films **518**(6), S179 (2010)
14.20 V.A. Dmitriev: In: *Properties of Silicon Carbide*, EMIS Datareview, ed. by G.L. Harris (IEE, London 1995) p. 214
14.21 S.-C. Wang, P.K. Nayak, Y.-L. Chen, J.C. Sung, J.-L. Huang: Ceram. Int. **37**(8), 3671 (2011)
14.22 S. Dhar: Bull. Mater. Sci. **28**(4), 349 (2005)
14.23 M. Milanova, P. Vitanov: Dilude nitride GaAsN and InGaAsN layers grown by low-temperature liquid-phase epitaxy. In: *Solar Cells – New Aspects and Solutions*, ed. by L.A. Kosyachenko (InTech, Rijeka 2011), doi:10.5772/21800
14.24 D. Mouleeswaran, T. Koyama, Y. Hayakawa: J. Cryst. Growth **321**(1), 85 (2011)
14.25 D. Mouleeswaran, T. Koyama, A. Tanaka, Y. Hayakawa: J. Cryst. Growth **362**, 238 (2013)
14.26 A. Ashery, A.A.M. Farag, M. Zeama: Superlatt. Microstruct. **66**, 136 (2014)
14.27 P. Strak, S. Krukowski, Z.R. Zytkiewicz: Bulk GaAs growth by contactless liquid phase electroepitaxy, Proc. 17th Int. Conf. Cryst. Growth Epitaxy, Warsaw (2013)
14.28 Z.R. Zytkiewicz, P. Strak, S. Krukowski: J. Cryst. Growth **318**, 351 (2011)
14.29 Y. Mao, A. Krier: Mater. Res. Soc. Symp. Proc. **450**, 49 (1997)
14.30 A. Krier, Z. Labadi, A. Manniche: J. Phys. D Appl. Phys. **32**, 2587 (1999)
14.31 V.H. Compean-Jasso, F. de Anda, V.A. Mishurnyi, A.Y. Gorbatchev, T. Prutskij, Y. Kudrievtsev: Thin Solid Films **548**, 168 (2013)
14.32 A. Mondal, T.D. Das, N. Halder, S. Dhar, J. Kumar: J. Cryst. Growth **297**, 4 (2006)
14.33 C.H. Sun, Q.W. Wang, F. Qiu, Y.F. Lv, H.Y. Deng, S.H. Hu, N. Dai: J. Alloy. Compd. **535**, 39 (2012)
14.34 G. Juarez-Diaz, J. Diaz-Reyes, J. Martinez-Juarez, M. Galvan-Arellano, J.A. Balderas-Lopez: Mater. Sci. Semicond. Process. **15**(5), 472 (2012)
14.35 J. Diaz-Reyes, P. Rodriguez-Fragoso, J.G. Mendoza-Alvarez: J. Lumin. **134**, 126 (2013)
14.36 H. Yamane, M. Shimada, T. Sekiguchi, F.J. DiSalvo: J. Cryst. Growth **186**, 8 (1998)
14.37 C. Klemenz, H.J. Scheel: J. Cryst. Growth **211**, 62 (2000)
14.38 T. Iwahashi, Y. Kitaoka, F. Kawamura, M. Yoshimura, Y. Mori, T. Sasaki, R. Armitage, H. Hirayama: Jpn. J. Appl. Phys. **46**(10), L227 (2007)
14.39 S. Hussy, P. Berwian, E. Meissner, J. Friedrich, G. Müller: J. Cryst. Growth **311**, 62 (2008)
14.40 M. Azizi, E. Meissner, J. Friedrich, G. Müller: J. Cryst. Growth **322**, 74 (2011)
14.41 S. Naritsuka, D. Kambayashi, H. Takakura, M. Tomita, T. Maruyama: C-plane GaN selective growth by liquid phase electroepitaxy under atmospheric pressure, Proc. 17th Int. Conf. Cryst. Growth Epitaxy, Warsaw (2013)
14.42 A. Krier, H.H. Gao, V.V. Sherstinov: IEE Proc. Optoelectron. **147**, 217 (2000)
14.43 E.R. Rubstov, V.V. Kuznetsov, O.A. Lebedev: Inorg. Mater. **34**, 422 (1998)
14.44 M.G. Mauk, Z.A. Shellenbarger, P.E. Sims, W. Bloothoofd, J.B. McNeely, S.R. Collins, P.I. Rabinowitz, R.B. Hall, L.C. DiNetta, A.M. Barnett: J. Cryst. Growth **211**, 411 (2000)
14.45 J.-I. Nishizawa, K. Suto: In: *Widegap II-VI Compounds for Optoelectronic Applications*, ed. by H.E. Ruda (Chapman Hall, London 1992)
14.46 F. Sakurai, M. Motozawa, K. Suto, J.-I. Nishizawa: J. Cryst. Growth **172**, 75 (1997)
14.47 M.G. Astles: In: *Properties of Narrow Gap Cadmium-Based Compounds*, EMIS Datareview, ed. by P. Capper (IEE, London 1994) p. 13
14.48 B. Pelliciari, J.P. Chamonal, G.L. Destefanis, L.D. Cioccio: Proc. SPIE **865**, 22 (1987)
14.49 P. Belca, P.A. Wolff, R.L. Aggarwal, S.Y. Yuen: J. Vac. Sci. Technol. A **3**, 116 (1985)
14.50 S.H. Shin, J. Pasko, D. Lo, W.E. Tennant, J.R. Anderson, M. Gorska, M. Fotouhi, C.R. Lu: Mater. Res. Soc. Symp. Proc. **89**, 267 (1987)
14.51 P. Capper: Liquid phase epitaxy of MCT. In: *Mercury Cadmium Telluride: Growth, Properties and Applications*, ed. by P. Capper, J. Garland (Wiley, Chichester 2011) p. 95
14.52 T. Tung, L.V. DeArmond, R.F. Herald, P.E. Herning, M.H. Kalisher, D.A. Olson, R.F. Risser, A.P. Stevens, S.J. Tighe: Proc. SPIE **1735**, 109 (1992)
14.53 P.W. Norton, P. LoVecchio, G.N. Pultz, J. Hughes, T. Robertson, V. Lukach, K. Wong: Proc. SPIE **2228**,

73 (1994)
14.54 P. Capper, J. Gower, C. Maxey, E. O'Keefe, J. Harris, L. Bartlett, S. Dean: In: *Growth and Processing of Electronic Materials*, ed. by N.M. Alford (IOM Communications, London 1998)
14.55 A. Wasenczuk, A.F.M. Willoughby, P. Mackett, E.S. O'Keefe, P. Capper, C.D. Maxey: J. Cryst. Growth **159**, 1090 (1996)
14.56 W. Schirmacher, R. Wollrab, H. Lutz, T. Schallenberg, J. Wendler, J. Ziegler: J. Electron. Mater. **43**(8), 2778 (2014)
14.57 R.S. Patrick, A.-B. Chen, A. Sher, M.A. Berding: J. Vac. Sci. Technol. A **6**, 2643 (1988)
14.58 A. Rogalski: *New Ternary Alloy Systems for Infrared Detectors* (SPIE, Bellingham, Washington 1994)
14.59 Y.V. Zorenko, V. Gorbenko, V. Savchyn, T. Zorenko, A. Fedorov: Development of novel scintillating screens based on the single crystalline films of Ce doped multi-component $(Gd,Y,Lu,La)_3(Al,Ga,Sc)_5O_{12}$ garnets by LPE method, Proc. 17th Int. Conf. Cryst. Growth Epitaxy, Warsaw (2013)
14.60 D. Ehrentraut, H. Sato, M. Miyamoto, T. Fukuda, M. Niki, K. Maeda, I. Niikura: J. Cryst. Growth **287**(2), 367 (2006)
14.61 J. Kobayashi, N. Ohashi, H. Sekiwa, I. Sakaguchi, M. Miyamoto, Y. Wada, Y. Adachi, K. Matsumoto, H. Haneda: J. Cryst. Growth **311**(19), 4408 (2009)
14.62 H. Nasu, Y. Tamada, S. Kubo: J. Cryst. Growth **310**(13), 3214 (2008)
14.63 C. Dubs, J.P. Ruske, J. Krausslich, A. Tunnermann: Opt. Mater. **31**(11), 1650 (2009)
14.64 Y. Lu, P. Dekker, J.M. Dawes: J. Cryst. Growth **311**(5), 1441 (2009)
14.65 A. Pena, P. Camy, A. Benayad, J.L. Doualan, C. Maurel, M. Olivier, V. Nazabal, R. Moncorge: Opt. Mater. **33**(11), 1615 (2011)
14.66 A.A. Chernov, H.J. Scheel: J. Cryst. Growth **149**, 187 (1996)
14.67 H.M. Manasevit, W.I. Simpson: J. Electrochem. Soc. **116**, 1725 (1969)
14.68 S. Pimputkar, J.S. Speck, S.D. DenBaars, S. Nakamura: Nat. Photonics **3**, 180–182 (2009)
14.69 S. Nakamura, S. Pearton, G. Fasol: *The Blue Laser Diode: The Complete Story* (Springer, Berlin 2013)
14.70 Y.H. Kim, H. Ruh, Y.K. Noh, M.D. Kim, J.E. Oh: J. Appl. Phys. **107**, 063501 (2010)
14.71 A.A. Chernov: Kinetic processes in vapor phase growth. In: *Handbook of Crystal Growth*, ed. by D.T.J. Hurle (Elsevier, Amsterdam 1994)
14.72 G.B. Stringfellow: J. Cryst. Growth **115**, 1 (1991)
14.73 S.J. Bass, M.L. Young: J. Crystal Growth **68**, 311 (1984)
14.74 D.M. Frigo, W.W. van Berkel, W.A.H. Maassen, G.P.M. van Mier, J.H. Wilkie, A.W. Gal: J. Cryst. Growth **124**, 99 (1992)
14.75 T. Bergunde, M. Dauelsberg, L. Kadinski, Y.N. Makarov, M. Weyers, D. Schmitz, G. Strauch, H. Jurgensen: J. Cryst. Growth **170**, 66 (1997)
14.76 X. Zhang, I. Moerman, C. Sys, P. Demeester, J.A. Crawley, E.J. Thrush: J. Cryst. Growth **170**, 83 (1997)
14.77 A.G. Thompson, R.A. Stall, W. Kroll, E. Armour, C. Beckham, P. Zawadzki, L. Aina, K. Siepel: J. Cryst. Growth **170**, 92 (1997)
14.78 L. Stock, W. Richter: J. Cryst. Growth **77**, 144 (1986)
14.79 S. Tompa, M.A. McKee, C. Beckham, P.A. Zwadzki, J.M. Colabella, P.D. Reinert, K. Capuder, R.A. Stall, P.E. Norris: J. Cryst. Growth **93**, 220 (1988)
14.80 P.M. Frijlink, J.L. Nicolas, P. Suchet: J. Cryst. Growth **107**, 166 (1991)
14.81 D.W. Kisker, J.N. Miller, G.B. Stringfellow: Appl. Phys. Lett. **40**, 614 (1982)
14.82 D. Ritter, H. Heinecke: J. Cryst. Growth **170**, 149 (1997)
14.83 C.A. Larson, N.I. Buchan, S.H. Li, G.B. Stringfellow: J. Cryst. Growth **93**, 15 (1988)
14.84 M.W. Raynor, V.H. Houlding, H.H. Funke, R. Frye, J.A. Dietz: J. Cryst. Growth **248**, 77 (2003)
14.85 M. Akaishi, T. Okawa, Y. Saito, K. Shimomura: J. Cryst. Growth **310**, 5069 (2008)
14.86 A. Hospodková, M. Zíková, J. Pangrác, J. Oswald, K. Kuldová, J. Vyskočil, E. Hulicius: J. Cryst. Growth **370**, 303 (2013)
14.87 J. Johansson, L.S. Karlsson, K.A. Dick, J. Bolinsson, B.A. Wacaser, K. Deppert, L. Samuelson: J. Cryst. Growth **310**, 5102 (2008)
14.88 M. Heurlin, P. Wickert, S. Falt, M.T. Borgstrom, K. Deppert, L. Samuelson, M.H. Magnusson: Nano Lett. **11**, 2028 (2011)
14.89 Z. Gu, P. Prete, N. Lovergine, B. Nabet: J. Appl. Phys. **109**, 064314 (2011)
14.90 R.M. Biefeld, R.W. Gedgridge Jr.: J. Cryst. Growth **124**, 150 (1992)
14.91 C.A. Wang, S. Salim, K.F. Jensen, A.C. Jones: J. Cryst. Growth **170**, 55 (1997)
14.92 S. Choi, H.J. Kim, S.-S. Kim, J. Liu, J. Kim, J.-H. Ryou, R.D. Dupuis, A.M. Fischer, F.A. Ponce: Appl. Phys. Lett. **96**, 221105 (2010)
14.93 S.T. Tan, X.W. Sun, H.V. Demir, S.P. DenBaars: IEEE Photonics J **4**, 613 (2012)
14.94 S. Nakamura: Jpn. J. Appl. Phys. **30**, 1620 (1991)
14.95 A. Stafford, S.J.C. Irvine, Z. Bougrioua, K. Jacobs, I. Moerman, E.J. Thrush, L. Considine: J. Cryst. Growth **221**, 142 (2000)
14.96 S. Qu, S. Li, Y. Peng, X. Zhu, X. Hu, C. Wang, X. Chen, Y. Gao, X. Xu: J. Alloy. Compd. **502**, 417 (2010)
14.97 S. Keller, S.P. DenBaars: J. Cryst. Growth **248**, 479 (2003)
14.98 B. Monemar, P.P. Paskov, G. Pozina, C. Hemmingsson, J.P. Bergman, T. Kawashima, H. Amano, I. Akasaki, T. Paskova, S. Figge, D. Hommel, A. Usui: Phys. Rev. Lett. **102**, 235501 (2009)
14.99 Y. Kim, K. Kim, T.W. Kim, L.J. Mawst, T.F. Kuech, C.Z. Kim, W.-K. Park, J. Lee: Sol. Energy **102**, 126 (2014)
14.100 B. Cockayne, P.J. Wright: J. Cryst. Growth **68**, 223 (1984)
14.101 S.J.C. Irvine, V. Barrioz, D. Lamb, E.W. Jones, R.L. Rowlands-Jones: J. Cryst. Growth **310**, 5198 (2008)
14.102 W. Bell, J. Stevenson, D.J. Cole-Hamilton, J.E. Hails: Polyhedron **13**, 1253 (1994)

14.103 J. Tunnicliffe, S.J.C. Irvine, O.D. Dosser, J.B. Mullin: J. Cryst. Growth **68**, 245 (1984)
14.104 C.D. Maxey, J.C. Fitzmaurice, H.W. Lau, L.G. Hipwood, C.S. Shaw, C.L. Jones, P. Capper: J. Electron. Mater. **35**, 1275 (2006)
14.105 P. Mitra, T.R. Schimert, Y.L. Tyan, A.J. Brouns, F.C. Case: Mat. Res. Soc. Symp. Proc. **299**, 155 (1994)
14.106 J.E. Hails, S.J.C. Irvine, D.J. Cole-Hamilton, J. Geiss, M.R. Houlton, A. Graham: J. Electron. Mater. **37**, 1291 (2008)
14.107 S. Fujita, S. Fujita: J. Cryst. Growth **145**, 552 (1994)
14.108 S. Fujita, A. Tababe, T. Sakamoto, M. Isemura, S. Fujita: J. Cryst. Growth **93**, 259 (1988)
14.109 S.J.C. Irvine, M.U. Ahmed, P. Prete: J. Electron. Mater. **27**, 763 (1988)
14.110 B.O. Jung, Y.H. Kwon, D.J. Seo, D.S. Lee, H.K. Cho: J. Cryst. Growth **370**, 314 (2013)
14.111 J. Wang, G. Du, B. Zhao, X. Yang, Y. Zhang, Y. Ma, D. Liu, Y. Chang, H. Wang, H. Yang, S. Yang: J. Cryst. Growth **255**, 293 (2003)
14.112 A.Y. Cho: J. Cryst. Growth **150**, 1 (1995)
14.113 C.T. Foxon: J. Cryst. Growth **251**, 1 (2003)
14.114 B.A. Joyce, T.B. Joyce: J. Cryst. Growth **264**, 605 (2004)
14.115 A.Y. Cho (Ed.): *Molecular Beam Epitaxy* (AIP, New York 1994)
14.116 E.H.C. Parker (Ed.): *The Technology and Physics of Molecular Beam Epitaxy* (Plenum, New York 1985)
14.117 B.A. Joyce, R.R. Bradley: Philos. Mag. **14**, 289 (1966)
14.118 J.R. Arthur: J. Appl. Phys. **39**, 4032 (1968)
14.119 A.Y. Cho: J. Vac. Sci. Technol. **8**, 31 (1971)
14.120 A.Y. Cho: Appl. Phys. Lett. **19**, 467 (1971)
14.121 J.W. Robinson, M. Ilegems: Rev. Sci. Instrum. **49**, 205 (1978)
14.122 P.A. Barnes, A.Y. Cho: Appl. Phys. Lett. **33**, 651 (1978)
14.123 W.T. Tsang: Appl. Phys. Lett. **34**, 473 (1979)
14.124 L.L. Chang, L. Esaki, W.E. Howard, R. Ludeke: J. Vac. Sci. Technol. **10**, 11 (1973)
14.125 H. Sakaki: J. Cryst. Growth **251**, 9 (2003)
14.126 A.Y. Cho: J. Appl. Phys. **41**, 2780 (1970)
14.127 M.D. Pashley, K.W. Haberern, J.M. Woodall: J. Vac. Sci. Technol. B **6**(4), 1468 (1988)
14.128 J.J. Harris, B.A. Joyce, P.J. Dobson: Surf. Sci. **103**, L90 (1981)
14.129 J.H. Neave, B.A. Joyce, P.J. Dobson, N. Norton: Appl. Phys. A **31**, 1 (1983)
14.130 C.T. Foxon, M.R. Boudry, B.A. Joyce: Surf. Sci. **44**, 69 (1974)
14.131 J.R. Arthur: Surf. Sci. **43**, 449 (1974)
14.132 C.T. Foxon, J.A. Harvey, B.A. Joyce: J. Phys. Chem. Solids **34**, 1693 (1973)
14.133 E.S. Tok, J.H. Neave, J. Zhang, B.A. Joyce, T.S. Jones: Surf. Sci. **374**, 397 (1997)
14.134 C.T. Foxon, B.A. Joyce: Surf. Sci. **50**, 434 (1975)
14.135 C.T. Foxon, B.A. Joyce: Surf. Sci. **64**, 293 (1977)
14.136 A.Y. Cho, J.R. Arthur: Prog. Solid State Chem. **10**(3), 157 (1975)
14.137 C.T. Foxon, B.A. Joyce: J. Cryst. Growth **44**, 75 (1978)
14.138 C.T. Foxon, B.A. Joyce, M.T. Norris: J. Cryst. Growth **49**, 132 (1980)
14.139 M.A. Herman, H. Sitter: *Molecular Beam Epitaxy*, Springer Series in Material Science, Vol. 7 (Springer, Berlin, Heidelberg 1988)
14.140 J. Saito, K. Nambu, T. Ishikawa, K. Kondo: J. Cryst. Growth **95**, 322 (1989)
14.141 C.R. Stanley, M.C. Holland, A.H. Kean, J.M. Chamberlain, R.T. Grimes, M.B. Stanaway: J. Cryst. Growth **111**, 14 (1991)
14.142 M. Bafleur, A. Munoz-Yague, A. Rocher: J. Cryst. Growth **59**, 531 (1982)
14.143 Y.G. Chai, R. Chow: Appl. Phys. Lett. **38**, 796 (1981)
14.144 C.E.C. Wood, L. Rathburn, H. Ohmo, D. DeSimone: J. Cryst. Growth **51**, 299 (1981)
14.145 S. Izumi, N. Hayafuji, T. Sonoda, S. Takamiya, S. Mitsui: J. Cryst. Growth **150**, 7 (1995)
14.146 J.H. Neave, P. Blood, B.A. Joyce: Appl. Phys. Lett. **36**(4), 311 (1980)
14.147 C.R. Stanley, R.F.C. Farrow, P.W. Sullivan: In: *The Technology and Physics of Molecular Beam Epitaxy*, ed. by E.H.C. Parker (Plenum, New York 1985)
14.148 M.B. Panish: J. Electrochem. Soc. **127**, 2729 (1980)
14.149 A.R. Calawa: Appl. Phys. Lett. **38**(9), 701 (1981)
14.150 E. Veuhoff, W. Pletschen, P. Balk, H. Luth: J. Cryst. Growth **55**, 30 (1981)
14.151 N. Putz, E. Veuhoff, H. Heinicke, H. Luth, P.J. Balk: J. Vac. Sci. Technol. B **3**(2), 671 (1985)
14.152 W.T. Tsang: Appl. Phys. Lett. **45**(11), 1234 (1984)
14.153 W.T. Tsang: J. Vac. Sci. Technol. **B3**(2), 666 (1985)
14.154 W.T. Tsang: Appl. Phys. Lett. **49**(3), 170 (1986)
14.155 T.H. Chiu, W.T. Tsang, J.E. Cunningham, A. Robertson: J. Appl. Phys. **62**(6), 2302 (1987)
14.156 W.T. Tsang, R.C. Miller: Appl. Phys. Lett. **48**(19), 1288 (1986)
14.157 J.S. Foord, C.L. Levoguer, G.J. Davies, P.J. Skevington: J. Cryst. Growth **136**, 109 (1994)
14.158 M. Weyers, J. Musolf, D. Marx, A. Kohl, P. Balk: J. Cryst. Growth **105**, 383 (1990)
14.159 R.J. Malik, R.N. Nottenberg, E.F. Schubert, J.F. Walker, R.W. Ryan: Appl. Phys. Lett. **53**, 2661 (1988)
14.160 F. Lelarge, J.J. Sanchez, F. Gaborit, J.L. Gentner: J. Cryst. Growth **251**, 130 (2003)
14.161 H. Morkoç, A.Y. Cho: J. Appl. Phys. **50**, 6143 (1979)
14.162 R.A. Stall, C.E.C. Wood, P.D. Kirchner, L.F. Eastman: Electron. Lett. **16**, 171 (1980)
14.163 R. Dingle, C. Weisbuch, H.L. Stormer, H. Morkoç, A.Y. Cho: Appl. Phys. Lett. **40**, 507 (1982)
14.164 G.B. Stringfellow, R. Stall, W. Koschel: Appl. Phys. Lett. **38**, 156 (1981)
14.165 H.G.B. Hicks, D.F. Manley: Solid State Commun. **7**, 1463 (1969)
14.166 C.T. Foxon, J.J. Harris, D. Hilton, J. Hewett, C. Roberts: Semicond. Sci. Technol. **4**, 582 (1989)
14.167 K. Ploog: J. Cryst. Growth **81**, 304 (1987)
14.168 H. Tanaka, M. Mushiage: J. Cryst. Growth **111**, 1043 (1991)
14.169 J. Miller: III-V Rev. **4**(3), 44 (1991)
14.170 D. Bimberg, M. Grundmann, F. Heinrichsdorff, N.N. Ledentsov, V.M. Ustinov, A.R. Korsh, M.V. Maximov, Y.M. Shenyakov, B.V. Volovik, A.F. Tsatsalnokov, P.S. Kopiev, Z.I. Alferov: Thin Solid Films **367**, 235 (2000)

14.171 S. Nakamura, T. Mukai, M. Senoh: Appl. Phys. Lett. **64**(13), 1689 (1994)
14.172 S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku, Y. Sugimoto: Jpn. J. Appl. Phys. **35**, L74 (1996)
14.173 H. Morkoç: J. Mater. Sci. Mater. Electron. **12**, 677 (2001)
14.174 S.E. Hooper, M. Kauer, V. Bousquet, K. Johnson, J.M. Barnes, J. Heffernan: Electron. Lett. **40**(1), 33 (2004)
14.175 N. Grandjean, M. Leroux, J. Massies, M. Laügt: Jpn. J. Appl. Phys. **38**, 618 (1999)
14.176 M. Kondow, K. Uomi, A. Niwa, T. Kitatani, S. Watahiki, Y. Yazawa: Jpn. J. Appl. Phys. **35**, 1273 (1996)
14.177 H. Riechert, A. Ramakrishnan, G. Steinle: Semicond. Sci. Technol. **17**, 892 (2002)
14.178 M. Kondow, T. Kitatani: Semicond. Sci. Technol. **17**, 746 (2002)
14.179 J.S. Harris, S.R. Bank, M.A. Wistey, H.B. Yuen: IEE Proc. Optoelectron. **151**(5), 407 (2004)
14.180 H. Ohno: J. Cryst. Growth **251**, 285 (2003)
14.181 H.J. Osten, E. Bugiel, O. Kirfel, M. Czernohorsky, A. Fissel: J. Cryst. Growth **278**(1), 18 (2005)
14.182 F.-J. Meyer zu Heringdolf, M.C. Reuter, R.M. Tromp: Nature **412**, 517 (2001)

15. Narrow Bandgap II–VI Semiconductors: Growth

Peter Capper

The field of narrow bandgap II–VI semiconductors is dominated by the compound $Hg_{1-x}Cd_xTe$ (MCT), although some Hg-based alternatives to this ternary have also been suggested. The fact that MCT is still the preeminent infrared (IR) material stems, in part, from the fact that by varying the x value material can be made to cover all the IR regions of interest. In addition, the direct band transitions are responsible for large values of absorption coefficients, allowing quantum efficiencies to approach 100%. Long minority carrier lifetimes result in low thermal noise allowing high-performance detectors to be made at the highest operating temperatures of comparable wavelengths. This chapter covers the growth of MCT by various bulk growth techniques, used mainly for first-generation infrared detectors; by liquid phase epitaxy, used mainly for second-generation infrared detectors; and by metal-organic vapor phase and molecular beam epitaxies, used mainly for third-generation infrared detectors.

15.1 **Bulk Growth Techniques** 344
15.1.1 Phase Equilibria 344
15.1.2 Crystal Growth 345
15.1.3 Material Characterization 347
15.2 **Liquid Phase Epitaxy (LPE)** 349
15.2.1 Hg-rich Growth 349
15.2.2 Te-rich Growth 350
15.2.3 Material Characteristics 351
15.2.4 Device Status 352
15.3 **Metal-Organic Vapor Phase Epitaxy (MOVPE)** 353
15.3.1 Substrate Type and Orientation 356
15.3.2 Doping 357
15.3.3 In situ Monitoring 357
15.4 **Molecular Beam Epitaxy (MBE)** 358
15.4.1 Double-Layer Heterojunction Structures 359
15.4.2 Multilayer Heterojunction Structures 359
15.4.3 MCT and CdZnTe Growth on Silicon 360
15.5 **Alternatives to MCT** 361
References 362

The field of narrow bandgap II-VI semiconductors is dominated by the compound $Hg_{1-x}Cd_xTe$ (MCT) [15.1–10], although some Hg-based alternatives to this ternary have been suggested and are discussed by *Rogalski* [15.11]. The fact that MCT is still the preeminent infrared (IR) material stems, in part, from the fact that by varying the x value material can be made to cover all the IR regions of interest. In addition, the direct band transitions are responsible for large values of absorption coefficients, allowing quantum efficiencies to approach 100%. Long minority carrier lifetimes result in low thermal noise allowing high-performance detectors to be made at the highest operating temperatures of comparable wavelengths. These three major advantages all stem from the energy band structure of the material and they apply to whichever device architecture is being used. It is thought that MCT is the third most studied semiconductor after silicon and gallium arsenide. This chapter covers the growth and characterization of MCT, mainly concentrating on the x region between 0.2 and 0.4 where the majority of applications are satisfied, and some of the Hg-based alternative ternary systems first described by Rogalski. Detectors made from these materials will be described in Chap. 34 in this Handbook.

In the first section, the growth of MCT by various bulk techniques is reviewed. These include solid-state recrystallization (SSR), Bridgman (plus the accelerated crucible rotation technique, ACRT) and the traveling heater method (THM), employing a Te zone. Despite the major advances made over the last three decades in the various epitaxial processes, which are discussed in subsequent sections, bulk techniques for MCT are still used for some photoconductive devices and for some niche applications, for example large-area optical components, and for large-size CdTe-based materials for substrates for MCT [15.10, Chaps. 3 and 4]. MCT material grown by LPE is currently in production for large-area focal plane arrays (FPAs) for both terrestrial and space thermal imaging applications. More complex,

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_15

fully doped heterostructures are being grown by metal-organic vapor phase epitaxy (MOVPE) and molecular beam epitaxy (MBE) for FPAs and novel nonequilibrium device structures.

For a fuller treatment of the topic matter in this chapter, plus further details of the characterization of MCT by various optical, electrical and structural means, the reader is referred to [15.1, Sect. A], [15.2, Chaps. 1–11], [15.3, Chaps. 1, 2, 10–12], [15.4, Chap. 7], *Baker* [15.5], [15.6, Chap. 7], and [15.7–10], including a useful list of books/papers in the Foreword of [15.10] by *Reine*.

15.1 Bulk Growth Techniques

Despite the large amount of progress that has taken place in various epitaxial growth techniques for MCT, bulk growth methods still play a key role in preparing device-quality material and in producing substrates for these epitaxial techniques. A critical aspect of these techniques is the establishment of a thorough knowledge of the relevant phase equilibria, and a brief description of these for MCT is now given.

15.1.1 Phase Equilibria

The understanding of two types of equilibria is of critical importance. One is the solid compound in equilibrium with the gaseous phase (vapor growth) and the other is solid–liquid–gaseous equilibria (growth from liquid/melt). These phase equilibria also help in understanding postgrowth heat treatments, either during cooldown to room temperature or during annealing processes to adjust the stoichiometry, and hence the electrical properties of compounds. Solid–liquid–gaseous equilibria are described by three variables, temperature (T), pressure (P) and composition (x). It is, however, far easier to understand the interrelations between these parameters by using two-dimensional (2-D) projections such as T–x, P–T and P–x plots.

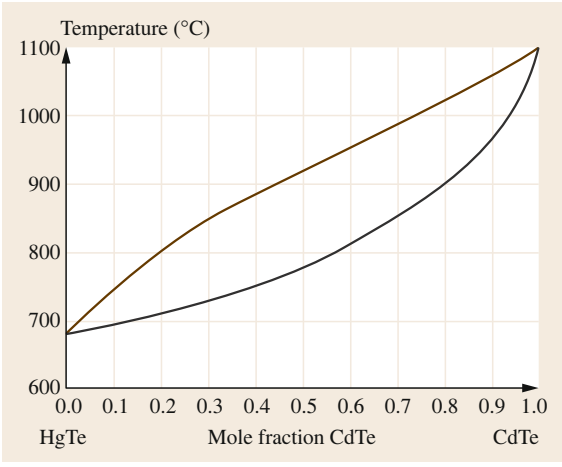


Fig. 15.1 T–x phase diagram for the pseudobinary CdTe-HgTe. (After [15.1])

A great deal of work has been carried out on the various phase equilibria ([15.1] contains reviews in this area). The liquidus and solidus lines in the pseudobinary HgTe-CdTe system are shown in Fig. 15.1. The wide separation between the liquidus and solidus, leading to marked segregation between the constituent binaries, CdTe and HgTe, was fundamental in the development of all the bulk growth techniques applied to this system. In addition to the solidus–liquidus separation high Hg partial pressures are also influential both during growth and postgrowth heat treatments. A full appreciation of the $P_{(Hg)}$–T diagram, shown in Fig. 15.2, is therefore essential. The curves are the partial pressures of Hg along boundaries for solid solutions of composition x where the solid solution is in equilibrium with another condensed phase as well as the vapor phase. For $x = 0.1$ and $10^3/T = 1.3$ (K^{-1}) MCT exists for Hg pressures between ≈ 0.1 (Te-saturated) and ≈ 7 (Hg-saturated) atm. Even at $x = 0.95$ and Te-saturated

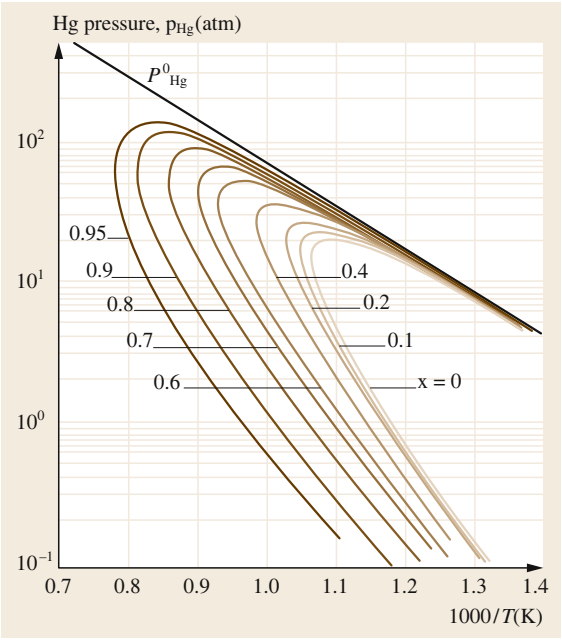


Fig. 15.2 P–T diagram for $Hg_{1-x}Cd_xTe$ (MCT). (After [15.1])

conditions, Hg is the predominant vapor species. It should also be noted that no solid solution contains exactly 0.5 atomic fraction Te. These features are highly significant in controlling the native defect concentrations and hence electrical properties in MCT.

15.1.2 Crystal Growth

Growth from the melt has constituted the vast majority of work on bulk growth of MCT, and material grown by bulk methods is still in use, particularly for photoconductive detectors. Several historical reviews have been published [15.1, 12–14]. *Micklethwaite* [15.12] and *Kruse* [15.13] gave detailed discussions on the growth techniques and material characterization prior to 1980. This situation up to 2005 was also summarized by this author [15.4, Chap. 7], [15.10, Chap. 1]. Many techniques were tried but three prime techniques survived these early attempts: SSR, Bridgman and THM [15.11]; some of which are still in production, for example Bridgman in the UK.

Solid-State Recrystallization (SSR)

This technique is used for ternary systems where there is a wide separation between solidus and liquidus, for example MCT. Other names that have been used for this process are quench anneal (QA) and cast recrystallize anneal (CRA). The term anneal is used in the first case to define a high-temperature grain-growth process, while in CRA it is a low-temperature process to adjust stoichiometry. In the basic technique, pure elements are loaded into a cleaned silica ampoule and the charge is melted and rocked to ensure complete mixing. Charges are then normally quenched rapidly to room temperature in the vertical orientation. This produces a dendritic structure that is reduced/removed by the recrystallization step that occurs at temperatures just below the melting point for many days. Grain growth occurs and microinhomogeneities in composition are removed.

Alternatives to the basic SSR process have included *slush* growth [15.1], high-pressure growth [15.16], incremental quenching [15.17], and horizontal casting [15.18]. In the *slush* process an initial homogeneous charge is prepared and then held across the liquidus-solidus gap with the lower-end solid and the upper-end liquid. High-pressure growth (30 atm He gas) was used in an attempt to reduce structural defects. The *slush* technique was used in production in the USA. Other developments were made in the basic process (e.g., [15.19, 20]) but details available in the open literature are sketchy, due to proprietary constraints. *Tregilgas* [15.21] has given a detailed review of the SSR process.

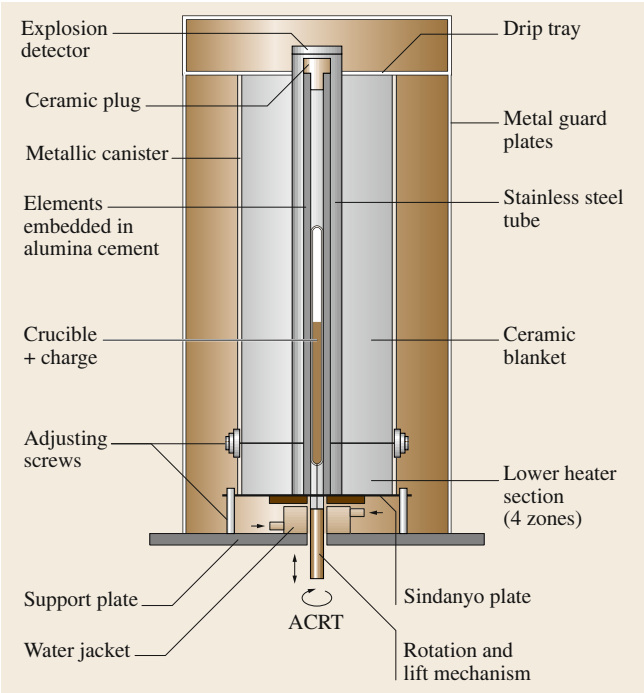


Fig. 15.3 Schematic of Bridgman growth kit for use with Hg-based compounds. (After [15.15])

Bridgman (plus ACRT)

In the standard Bridgman process, elemental Cd, Hg and Te are loaded into a clean silica ampoule, homogenized by melting/rocking and then frozen slowly from one end in a vertical system (Fig. 15.3, [15.15]) to produce a large-grained ingot [15.22]. There is marked segregation of CdTe with respect to HgTe in the axial direction, due to the wide gap between the solidus and liquidus, but this leads to an advantage of the Bridgman process over other techniques, i. e., material in both the main ranges of interest (3–5 and 8–12 µm, i. e., $x = 0.3$ and 0.2, respectively) is produced in a single run. The accelerated crucible rotation technique

Fig. 15.4 Comparison of 20, 40, 50 and 70 mm MCT slices. (Courtesy of Leonardo MW Ltd)

(ACRT) in which the melt is subjected to periodic acceleration/deceleration provided a means of stirring melts contained in sealed, pressurized ampoules.

The first report of ACRT in MCT Bridgman growth was given in [15.23]. These effects were developed and reviewed in [15.1, 4, 22]. Crystals were produced up to 20 mm in diameter and with x values up to 0.6 in the tip regions of some crystals. Currently in this author's laboratory crystals of 20 mm diameter, 200 mm in length weighing 0.5 kg are in routine production, while R&D activities have produced several 40 mm, 50 mm (up to 2.5 kg in weight) and even a 70 mm diameter crystal (Figs. 15.4–15.6 [15.10]). Extensions to the basic ACRT process enable material up to $x \approx 0.93$ (0.91 μm) to be produced in the first-to-freeze regions of crystals (Fig. 15.6 [15.10, 24]). These near-IR slices are highly transmissive and show a high degree of compositional uniformity ($\Delta x \approx 0.0005$) [15.24]. The use of ACRT produced material that enabled the fabrication of small two-dimensional arrays of photodiodes; this material was also used to demonstrate nonequilibrium detector operation and in the production of SPRITE detectors ([15.3, Chap. 11]). *Higgins* et al. [15.25] recently reported the history of their

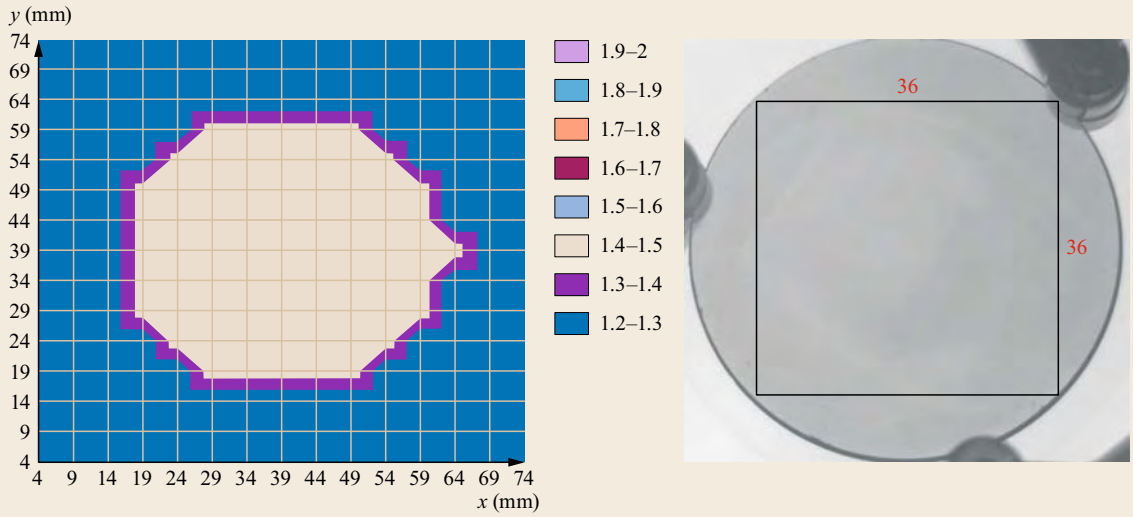


Fig. 15.5 Radial distribution of wavelength (in μm at room temperature) in 50 mm diameter high-x slice from 30% start

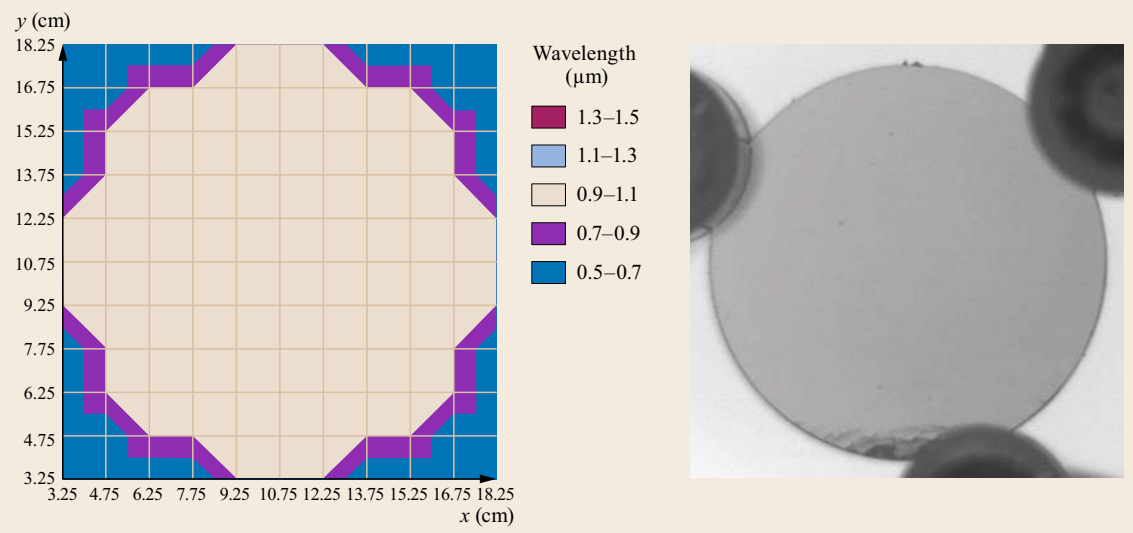


Fig. 15.6 Radial distribution of wavelength (in μm at room temperature – 0.91 μm ≡ $x \approx 0.93$) in 20 mm diameter slice

detector materials engineering (DME) process; in essence they grew crystals vertically at ≈ 700 °C with a piece of high-x MCT floating on the surface of the melt to provide replenishment and overcome the segregation effects. In addition, to increase the growth rate they reduced the boundary layer thickness by adding ACRT. (The authors note that the technique was conceived in 1981, in the same year as the patent [15.23], but not implemented until 1982, i. e., some years after Bridgman/ACRT crystals were being grown in this author's laboratory.) Crystals of 15 mm diameter were grown with good compositional uniformity, low Te precipitate density and high purity.

Traveling Heater Method (THM)

Triboulet [15.26] (and coworkers) developed this technique for MCT where diameters up to 40 mm were accomplished and x values up to 0.7 [15.27] for optical communication devices. *Durand* et al. [15.28] employed seeds to produce large oriented crystals, an advantage of this technique over other bulk methods. *Gille* et al. [15.29] adopted a slightly different approach. A pre-THM step was used to produce a quenched Te-rich (53–60%) MCT melt. The entire growth procedure took several months but gave uniform material. This group also used rotation in the horizontal growth by THM with some success [15.30]. Two groups also applied ACRT to THM and obtained enhanced material properties [15.31, 32].

15.1.3 Material Characterization

Compositional uniformity can be expressed in both axial and radial directions as $\Delta A(x)$ and $\Delta R(x)$, although the combination of the two determines the yield of useful material. Assessment is normally by Fourier transform infrared (FTIR) spectrometry. The recrystallization process (in SSR) improves $\Delta A(x)$ from ± 0.05 to ± 0.02 [15.13] and removes the microinhomogeneity in x in both directions [15.33]. Independent control of axial and radial variations was claimed by *Nelson* et al. [15.34] to achieve ± 0.002 over 95% of slices up to 25 mm in diameter. *Colombo* et al. [15.17] quote $\Delta R(x)$ of 0.005–0.01 in incremental-quenched material. *Su* et al. [15.35] report ± 0.008 for $\Delta A(x)$ and $\Delta R(x)$ in vertically cast material. *Galazka* et al. [15.36] showed that reduced gravity gives smaller $\Delta R(x)$, and they also noted that constitutional supercooling at the first-to-freeze end is also reduced and the growth is diffusion limited.

Bartlett et al. [15.37] grew SSR and Bridgman crystals over a wide range of growth rates and concluded that $\Delta R(x)$ was due to the combination of a concave (to the melt) growth interface and density-driven convective flow within a boundary layer close to the interface. This, added to the normal segregation of the low melting point HgTe to the center of the concave interface, led to the observed variation. It was also concluded that crystals with simultaneous small $\Delta A(x)$ and $\Delta R(x)$ were not possible; slow growth gave low $\Delta R(x)$ while fast growth minimized $\Delta A(x)$. In Bridgman/ACRT material, an 80 mm length at $x = 0.21$ with both $\Delta A(x)$ and $\Delta R(x)$ within ± 0.002, linked to a flat interface produced by strong Ekman flow in the interface region, was reported by *Capper* [15.22]. 20 mm diameter ACRT crystals have been grown that exhibited $\Delta R(x) = \pm 0.003$ at $x = 0.22$ [15.22] and $\Delta R(x) = \pm 0.0005$ at $x = 0.93$ [15.10].

Triboulet et al. [15.14] quoted values of ± 0.02 (over 3 cm at $x = 0.2$) and ± 0.002 for $\Delta A(x)$ and $\Delta R(x)$ respectively, for THM growth at 0.1 mm h^{-1}. For the very slow-grown material produced by *Gille* et al. [15.29], variations in composition were within experimental error for $\Delta R(x)$, while $\Delta A(x)$ was ± 0.005. *Royer* et al. [15.31] found improved radial uniformity with the addition of ACRT to THM but *Bloedner* and *Gille* [15.32] noted both $\Delta A(x)$ and $\Delta R(x)$ were only as good as non-ACRT crystals.

With regard to electrical properties, as-grown material from SSR, ACRT Bridgman (with $x < 0.3$) and THM is highly p-type in nature, believed to be due to metal vacancies. These types of material can all be annealed at low temperatures, in the presence of Hg, to low n-type levels, indicating that the p-type character is due to metal vacancies. However, Bridgman material is n-type as-grown [15.22] and the residual impurity donor level is found to be $< 5 \times 10^{14}$ cm^{-3}. *Higgins* et al. [15.20, 25] showed that n-type carrier concentrations can be $< 10^{14}$ cm^{-3} at $x \approx 0.2$ in melt-grown material. A great deal of work was done on the annealing behavior of SSR material, see *Tregilgas* [15.21]. n-Type levels of 2×10^{14} cm^{-3} after Hg annealing $x = 0.2$ and higher x material were reported by *Nguyen Duy* et al. [15.27]. *Durand* et al. [15.38] note that THM growth at 600 °C results in p-type behavior but growth at 700 °C gives n-type material. In this author's laboratory current 20 mm diameter ACRT material has reached mid-10^{13} cm^{-3} levels with high mobility after a low-temperature Hg anneal step. This is thought to reflect the improvements made in the purity of starting elements over recent years. The basic p to n conversion process [15.20] has recently been extended by *Capper* et al. [15.10, Chap. 13], [15.39], to produce analytical expressions that can account for the temperature, donor level and composition dependencies of the junction depth.

High minority n-type carrier lifetimes have been reported in all types of material. *Kinch* [15.40] noted values > 1 μs (at 77 K) in $x = 0.2$ SSR crystals, while

Triboulet et al. [15.14] quote 3 μs in THM material. *Pratt* et al. [15.41, 42] found high lifetimes, up to 8 μs in $x = 0.23$ Bridgman and ACRT material with n-type levels of $1-6\times10^{14}\,\mathrm{cm}^{-3}$ and up to 30 μs for equivalent $x = 0.3$ material (192 K). *Higgins* et al. [15.25] quote 3 μs in $x = 0.215$ DME material.

In terms of extrinsic doping, most elements are electrically active in accordance with their position in the periodic table. This is true in SSR material for Group V and VII elements only after a high-temperature treatment, and is linked to the stoichiometry level at the growth temperature, i. e., those elements that substitute on Te lattice sites have to be forced onto the correct sites in Te-rich material. Groups I and III elements are acceptors and donors respectively, on the metal sites. In ACRT crystals, Groups I and III are acceptors and donors respectively, on the metal sites as they are in Bridgman, with the exception of Au [15.22]. Groups V and VII are inactive dopants in those portions of ACRT crystals that are Te-rich as-grown, i. e., $x < 0.3$, but are active dopants for $x > 0.3$ where metal-rich conditions prevail, as found in Bridgman material. In doped material, grown by either standard Bridgman or ACRT, acceptor ionization energies were found to be lower than undoped counterparts [15.22]. Segregation of impurities in SSR is very limited due to the initial fast quench step. By contrast, Bridgman/ACRT, DME and THM benefit from marked segregation of impurities due to their relatively slow growth rate. Impurity segregation behavior was affected by ACRT [15.22]: in general, segregation coefficients decrease in ACRT crystals when compared to standard Bridgman. This segregation leads to very low levels of impurities in both THM [15.26] and Bridgman/ACRT material [15.43]. *Higgins* et al. [15.25] note that they used In and Ga for n-type doping but no p-type doping was reported.

Vere [15.44] has reviewed structural properties and noted grains, subgrains, dislocations, Te precipitates and impurities on dislocations as major problems. Grain boundaries act as recombination centers and generate noise and dark current in devices. Subgrain sizes range from 50–500 μm in both SSR and Bridgman material. *Williams* and *Vere* [15.45] showed how a recrystallization step at high pressure and temperatures > 600 °C coalesces subgrains and eventually eliminates them. Tellurium precipitation was extensively studied [15.45, 46] and was summarized in [15.45]. Precipitates were found to nucleate on dislocations during the quench from the recrystallization step but a 300 °C anneal dissolves the precipitates, generating dislocation loops that climb and lead to dislocation multiplication.

Quenching studies in Bridgman/ACRT crystals grown in flat-based ampoules [15.22] revealed not only a flat growth interface but also that the slow-grown material, produced prior to quenching, was a single crystal. This demonstrated the power of Ekman stirring and the importance of initiating the growth of a single crystal grain. In Bridgman material, which is close to stoichiometric (Te excess < 1%) little Te precipitation occurs. Above this limit, precipitates/inclusions of 0.1–0.5 mm in size are seen. Work in this author's laboratory [15.24] has demonstrated flat solid-liquid interfaces in the $x \approx 0.21$ region of 20 mm diameter ACRT crystals, mirroring the high level of compositional uniformity in these crystals.

Triboulet et al. [15.14] found no Te precipitates in their THM material, despite using a Te zone. Several groups [15.17, 47] noted an increase in dislocation density and in rocking curve widths on moving from the center to the edge of THM crystals. The addition of ACRT to the THM process led to increased growth rates with no detriment to the Te precipitation or grain structure. *Higgins* et al. [15.25] annealed their DME material at 500 °C in crystal form to reduce their Te precipitate content.

We have recently [15.10, Chap. 1], [15.24, 48] extended our ACRT growth process to produce high-x material for various near-IR applications (Figs. 15.4–15.6). This necessitated improving the ampoule sealing procedure to manage the higher pressures caused by the higher growth temperatures. Assessment techniques were extended beyond the normal FTIR wavelength characterization to include imaging of defects using a mid-IR camera system.

To summarize, the two main problems with all bulk material are the limitation on size, only Bridgman/ACRT and THM have reproducibly produced material above 20 mm diameter (up to 70 mm in the author's lab), and the fact that it is rarely completely single crystal in nature. These two limitations led to the drive in the late 1970s and early 1980s to develop various epitaxial growth techniques.

15.2 Liquid Phase Epitaxy (LPE)

The process of LPE emerged in the 1980s as the predominant materials growth technology for the fabrication of second-generation MCT focal plane arrays (FPAs). Material was routinely grown for high-performance photoconductive, photovoltaic, and laser-detector devices in the entire 2–18 μm spectral region. This is manifested by increased size (over bulk material), enabling large-array formats to be made; reduced cycle time; better material characteristics in terms of composition uniformity, crystal quality, and electrical properties, making the material no longer the limiting factor in FPAs; and the realization of some advanced device structures, either to simplify the FPA design (backside-illuminated detectors) or to enhance detector performance (double-layer heterostructures) [15.6].

Two completely different technical approaches were pursued: growth from Te solutions and growth from Hg solutions. One major advantage of the Hg-solution technology is its ability to produce layers of excellent surface morphology due to the ease of melt decanting. Two additional unique characteristics, essential for the fabrication of high-performance double-layer heterojunction (DLHJ) detectors by LPE, are low liquidus temperature (< 400 °C), which makes growth of the cap layer feasible, and ease of incorporating both p-type and n-type temperature-stable impurity dopants, such as As and In, during growth. While layers grown from Hg-rich solutions are easily doped with group VB elements with high solubility [15.49], layers grown from tellurium-rich solutions are not. Group IIIB elements, in particular indium, are readily incorporated from both solutions.

A number of LPE approaches have been used to grow both thin and thick films. The principal LPE techniques used are tipping, dipping and sliding boat techniques. A schematic diagram of these three processes is shown in Fig. 15.7. The tipping and dipping techniques have been implemented by using both tellurium- and mercury-rich solutions, whereas only tellurium-rich solutions have been used with the sliding boat. Both tipping and sliding boat Te-rich techniques are still in use.

Extensive experimental phase diagram and thermodynamic data have been critically reviewed, along with the results calculated by the associated solution model [15.50]. As in bulk growth, full knowledge of the solid–liquid phase relation is essential for proper use of solution-growth processes. In addition, the solid–vapor and liquid–vapor phase relations are of practical importance, especially in view of the high Hg pressure in the growth process and the effect of the vapor of constituent components upon postgrowth annealing and the consequent electrical properties. *Astles* [15.51] reviewed the experimental data of Te-rich LPE growth at 460 to 550 °C; also see *Capper* in [15.6], [15.10, Chap. 5].

15.2.1 Hg-rich Growth

For MCT growth from Hg-rich melts the design and operation of a system is dominated by consideration of the high Hg vapor pressure, which comprises ≥90% of the growth solution. A secondary, but related factor, is the requirement to minimize melt composition variation during and between growths due to solvent or solute loss. These factors led to the evolution of a vertical high-pressure furnace design with a cooled reflux region. The furnace has to provide a controllable, uniform and stable thermal source for the melt vessel, which has to be capable of maintaining at least 550 °C continuously. The cylindrical melt vessel consists of a high-strength stainless steel chamber lined with quartz. Such systems are capable of containing about 10–20 kg of melt at 550 °C for several years with no degradation in melt integrity or purity.

The system must be pressurized and leak-free to keep the Hg-rich melt from boiling or oxidizing at temperatures above 360 °C. Typical pressures range up to 200 psi and the pressurization gas may be high-purity H_2 or a less explosive reducing-gas mixture contain-

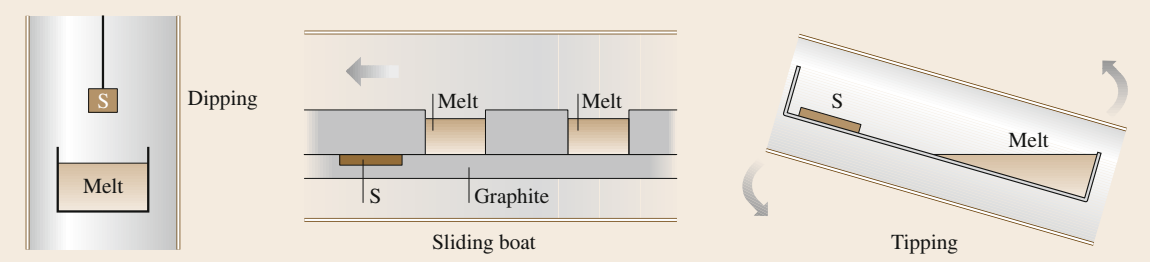


Fig. 15.7 Schematic diagram of the three principal LPE growth techniques. (After [15.2])

ing H_2. The melts are always kept saturated and are maintained near to the growth temperature and pressure between successive runs. The prepared substrates are introduced into the melt through a transfer chamber or airlock. A high-purity graphite paddle with externally actuated shutters holds the substrates.

Growth begins by lowering the paddle plus substrates into the melt and allowing thermal equilibrium to be reached while stirring. A programmed ramp then reduces the melt temperature to the required level at which point the shutters are opened and the substrates are exposed to the melt. The composition of the layer and its variation are determined mainly by the melt composition and its thermal uniformity. Large melts allow the production of layers up to 30 cm^2 with excellent compositional and thickness uniformity and allow dopant impurities to be accurately weighed for incorporation into layers and to maintain stable electrical characteristics over long periods of time. Four layers with a total area of 120 cm^2 can be grown in a single run [15.52]. *Norton* et al. [15.53] also scaled up for the growth of cap layers from Hg-rich solutions, each reactor capable of growth on four 24 cm^2 base layers per run.

15.2.2 Te-rich Growth

A number of problems encountered with bulk crystal growth techniques are solved using MCT growth from tellurium-rich solution. The most important of these is the reduction of the Hg vapor pressure over the liquid by almost three orders of magnitude at the growth temperature. Growth from Te-rich solutions is used in three embodiments: dipping, tipping and sliding boat technologies (Fig. 15.7). While the tipping process may be used for low-cost approaches, it was not as widely used as the sliding boat and the dipping techniques. A comparison of the three techniques is shown in Table 15.1.

Current dipping reactors are capable of growing in excess of 60 cm^2 per growth run and are kept at temperature for long periods, > 6 months. Melts, on the other hand, last a very long time, > 5 years. A sensor-based reactor capable of growing MCT thick layers at relatively high production volumes and excellent reproducibility was described by *Westphal* et al. [15.54]. The reactors utilize resistance temperature devices (RTDs) to control the temperature within a maximum range of $\pm 0.005\,^{\circ}C$ at a temperature of $\approx 480\,^{\circ}C$. The melts weigh about 4 kg with a liquidus temperature of $\approx 480\,^{\circ}C$ for a solid MCT composition $x = 0.225$. Accurate mercury vapor pressure control ensures constant liquidus and thus constant liquid composition.

Layers produced from Te-rich solutions are usually grown by isothermal supersaturation or programmed cooling techniques or a combination of both. The details of the mass transport of Hg-Cd-Te solutions during LPE growth were presented by *Shaw* [15.55].

Sliding-boat LPE normally consists of a graphite boat with a recess in the base for holding the substrate wafer and a movable block with wells that contain the LPE solution and that allows the solution to be brought into contact with the substrate and then wiped off after growth. The main advantages of the technique are the efficient use of solution and the possibility of growing multilayers. The main disadvantages are the need for careful machining of the boat components in order to obtain efficient removal of the solution after growth and the need for precisely sized substrate wafers to fit into the recess in the boat. The sliding-boat growth process has several variants but essentially a polished substrate is placed in the well of a graphite slider and the Te-rich solution is placed in a well in the body of the graphite boat above the substrate and displaced horizontally from it. Normally, a separate well contains the HgTe charge to provide the Hg vapor pressure needed during growth and during cooldown to control the stoichiometry. The boat is then loaded into a silica tube that is flushed with nitrogen/argon prior to the introduction of H_2 for the growth phase. The furnace surrounding the work tube is slid over the boat and the temperature is increased to 10–20 °C above the relevant liquidus.

Table 15.1 Comparison of three LPE growth techniques

LPE growth technique	Advantages	Disadvantages
Sliding boat	Thickness control Large area Double layers	Thick layers difficult to grow Substrate thickness/planarity Control High cost
Tipping	Thickness control Lowest investment for experimental investigation Closed system/good P_{Hg} control	High cost Scale up
Dipping	Thick layer growth High throughput Low cost Large/flexible area	Thin-layer thickness control Double layers

At that point a slow temperature ramp (2–3 °C h^{-1}) is initiated and when the temperature is close to the liquidus of the melt the substrate is slid under the melt and growth commences. After the required thickness of MCT has been deposited (typical growth rates are 9–10 μm h^{-1}) the substrate is withdrawn and the temperature is decreased to an annealing temperature (to fix the p-type level in the as-grown material) before being reduced rapidly to room temperature. Layer thicknesses of 25–30 μm are normally produced for loophole diode applications [15.56].

15.2.3 Material Characteristics

Good composition uniformity, both laterally and in depth, is essential to obtain the required uniform device performance. Growth parameters that need to be optimized in Hg-rich LPE include the degree of supercooling and mixing of the melt, the geometrical configuration of the growth system, the melt size, and the phase diagram. The standard deviation of the cutoff wavelength, for 12-spot measurements by Fourier transform infrared (FTIR) transmission at 80 K across a 30 cm^2 long-wavelength infrared (LWIR) layer, is reported as 0.047.

Composition control and uniformity of layers grown by dipping Te-rich LPE is one of the strengths of this process. The cutoff wavelength reproducibility is typically 10.05 ± 0.18 μm. Dipping Te-rich LPE is used mainly to grow thick films, about 100 μm, hence thickness control is not one of its advantages. Thickness control is about ±15% for layers < 20 μm due to the relatively large amount of solidified material.

For Te-rich sliding boat LPE layers of ≈ 30 μm thickness can show wavelength uniformity at room temperature of 6.5 ± 0.05 μm over 90% of the area of 20×30 mm layers [15.57].

The dislocation density of LPE MCT and its effects on device characteristics has received much attention [15.58, 59]. The dislocation density is dominated mainly by the dislocations of the underlying substrate [15.60]. For layers grown on substrates with ZnTe ≈ 3–4% the dislocations are present only in the interface region and the dislocation density is close to that of the substrate. For sliding boat LPE material, this same variation of dislocation density with depth is seen [15.56] (typical values are $3–7 \times 10^4$ cm^{-2}).

For the production of heterostructure detectors with MCT epitaxial layers, it is essential that proper impurity dopants be incorporated during growth to form well-behaved and stable p-n junctions. An ideal impurity dopant should have low vapor pressure, low diffusivity, and small impurity ionization energy. Group V and group III dopants, As and Sb for p-type, and In for n-type, are the dopants of choice. Hg-rich melts can readily be doped to produce n- and p-type layers, and solubilities of most of the useful dopants are significantly higher than in Te-rich solutions, most notably for group V dopants, which are among the most difficult to be incorporated into MCT.

The accurate determination of dopant concentration in the solid involves the use of Hall effect measurements and secondary ion mass spectrometry (SIMS) concentration profiles, preferably on the same or adjacent samples. The ease of incorporating Group I and Group III dopants into MCT, irrespective of nonstoichiometry, has been confirmed experimentally [15.5, 61].

Excess carrier lifetime is one of the most important material characteristics of MCT since it governs the device performance and frequency response. The objective is to routinely produce material with a lifetime that is limited by Auger processes, or by the radiative process in the case of the medium-wavelength infrared (MWIR) and short-wavelength infrared (SWIR) material [15.62]. It has been reported that intentionally impurity-doped LPE MCT material grown from As-doped Hg-rich melts can be obtained with a relatively high minority-carrier lifetime [15.63]. The 77 K lifetimes of As-doped MWIR ($x = 0.3$) MCT layers are significantly higher than those of undoped bulk MCT and are within a factor of two of the theoretical radiative lifetime. Various annealing schedules were proposed [15.57] that may lead to a reduction in Shockley–Read traps with a consequent increase in lifetime, even in undoped material. Lifetimes in In-doped MWIR MCT were also found to exhibit an inverse linear dependence on the doping concentration [15.63], with $N_d\tau$ products similar to $N_a\tau$ products observed for the As-doped material. The lifetimes of LWIR In-doped LPE material are limited by the Auger process typically at doping levels above 10^{15} cm^{-3}.

The first heterojunction detectors were formed in material grown by Hg-rich LPE [15.64]. For double-layer heterojunctions (DLHJ), a second LPE (cap) layer is grown over the first (base) layer. With dopant types and layer composition controlled by the LPE growth process, this approach offers great flexibility (p-on-n or n-on-p) in junction type and in utilizing heterojunction formation between the cap and absorbing base layers for optimization of detector performance. The key step in the process is to grow the cap layer doped with slow-diffusing impurities, In for an n-type cap layer, and As or Sb for a p-type cap layer.

For future large-area FPAs, Si-based substrates were developed as a replacement for bulk CdZnTe

Table 15.2 Summary of MCT PV detector technologies used for two-dimensional hybrid focal plane arrays. Upper case P and N denote wider-gap layers, e.g., *P-on-n* denotes a wider-gap p-type layer grown onto a narrow-gap n-type absorber layer. (After [15.5])

Photodiode configuration	Junction-formation method	Company
P-on-n mesa	1. Two-layer LPE on CdZnTe: Base: Te-solution slider, indium-doped Cap: Hg-solution dipper, arsenic-doped	BAE North America, USA
P-on-n mesa	1. Two-layer LPE on CdZnTe or Si: Base: Hg-solution dipper, indium doped Cap: Hg-solution dipper, arsenic-doped	Raytheon Infrared Center of Excellence (RIRCoE, formerly SBRC) and Hughes Research Laboratories (HRL), USA
P-on-n planar buried-junction heterostructure	Arsenic implant into indium-doped N-n	Rockwell/Boeing
n^+-n^--p planar homojunction	Boron implant into Hg-vacancy p-type, grown by Hg-solution tipper on 3″ dia. sapphire with MOVPE CdTe buffer; ZnS passivation	
n^+-on-p planar	Ion implant into acceptor-doped p-type LPE film grown by Te-solution slider	Sofradir (Societe Francaise de Detecteurs Infrarouge), France
n-on-p VIP	Ion implantation forms n-on-p diode in p-type MCT layer, grown by Te-solution LPE on CdZnTe and epoxied to silicon ROIC wafer; over the edge contact	DRS Infrared Technologies (formerly Texas Instruments), USA
n-p loophole	Ion beam milling forms n-type islands in p-type Hg-vacancy-doped layer grown by Te-solution LPE on CdZnTe, and epoxied onto silicon ROIC wafer; cylindrical lateral collection diodes	Leonardo MW Ltd (formerly BAE SYSTEMS Infrared Ltd.), UK
n-on-p	Boron implant (n) and Au doping forms the p-type region of diodes grown by Te solution LPE on in-house CdZnTe, flip-chip bonding onto ROIC	AIM, Germany

substrates. This effort is directed at improvements in substrate size, strength, cost, and reliability of hybrid FPAs, particularly during temperature cycling. These alternative substrates, which consist of epitaxial layers of CdZnTe or CdTe on GaAs/Si wafers [15.65] or directly onto Si wafers [15.66], are particularly advantageous for the production of large arrays. High-quality epitaxial MCT has been successfully grown on the Si-based substrates by the Hg-melt LPE technology for the fabrication of p-on-n DLHJ detectors. The first high-performance 128×128 MWIR and LWIR arrays were demonstrated by *Johnson* et al. [15.67]. MWIR arrays as large as 512×512 and 1024×1024 have also been produced [15.68]. A bias-selectable two-color (LWIR/MWIR) detector structure was first fabricated by growing three LPE layers from Hg-rich melts in sequence on a bulk CdZnTe substrate [15.68].

15.2.4 Device Status

The various MCT photodiode designs and processes used by the major FPA manufacturers for both back-illuminated bump-interconnected FPAs and front-illuminated loophole FPAs are summarized in Table 15.2, circa 2010.

Destefanis and coworkers [15.69] described work on large-area and long linear FPAs based on Te-rich sliding boat LPE material. Using a 15 µm pitch they were able to produce 1000×1000 MW arrays of photodiodes and 1500×2 MW and LW long linear arrays. *Gravrand* et al. [15.70] suggest that only by growing on lattice-matched substrates, i. e., CdZnTe plus LPE growth, will the very highest performance FPAs be produced. CdZnTe crystals up to 115 mm diameter are grown in-house to produce substrates up to 47×48 mm

in size. Layers with wavelengths from 1.7–12 μm can be produced.

Wenisch et al. [15.71] recently noted that LPE is still AIM's production process of choice due to its maturity, and they report detectors with 1280 × 1024 pixels on a 15 μm pitch. This was still the case a year later; see *Schirmacher* et al. and *Ziegler* et al. [15.72, 73].

Maxey et al. [15.74] reported that LPE material was also used for avalanche photodiodes in the short-wave (SW) regions of 1.5–1.8 and 2.0–2.4 μm to produce high gain figures.

15.3 Metal-Organic Vapor Phase Epitaxy (MOVPE)

Metal-organic vapor phase epitaxy (MOVPE) of MCT is dominated by the relatively high vapor pressures of mercury that are needed to maintain equilibrium over the growing film. This arises from the instability of HgTe compared with CdTe and requires much lower growth temperatures than are usual for more stable compounds. The MOVPE process was developed as a vapor-phase method that would provide sufficient control over growth parameters at temperatures below 400 °C. The main advantage of this is that the elements (although not mercury) can be transported at room temperature as volatile organometallics and react in the hot gas stream above the substrate or catalytically on the substrate surface. The first mercury chalcogenide growth by MOVPE [15.75, 76] was followed by intense research activity that has brought the technology to its current state of maturity.

Although the MOVPE reaction-cell conditions are far from equilibrium, an appreciation of the vapor-solid equilibrium can determine the minimum conditions needed for growth. At MOVPE growth temperatures, the pressure can vary by three orders of magnitude and remain in equilibrium with a single phase of HgTe. However, the Te_2 partial pressure varies across the phase field in the opposite sense to the Hg partial pressure.

The maintenance of vapor pressure equilibrium within the reactor cell does not automatically lead to the growth of an epilayer but it does enable us to eliminate conditions that would not be suitable. For example, the supply of Hg to the substrate for growth at 400 °C must be greater than the minimum equilibrium pressure of 10^{-2} bar. This pressure is higher than would be conveniently delivered from a mercury organometallic source that would typically be less than 10^{-3} bar. Almost all narrow-gap II-VI epilayers have used a heated elemental mercury source that is capable of delivering the required partial pressure.

Criteria for precursors that are generally applicable to MOVPE are volatility, pyrolysis temperature, stability and volatility of organic products and purity. The latter depends on the synthesis and purification routes for the precursors but techniques such as adduct purification of the group II organometallics have made a major contribution to epilayer purity. A good review of this topic is given by *Jones* [15.77].

The precise control over layer properties that is required for uniform detector arrays and tuning of the response wavelength can only be achieved if the cadmium concentration is controlled to better than 1%. Improvements in reactor design, growth techniques, such as the IMP (interdiffused multilayer process), choice of precursors and operating conditions have all contributed to successful targeting of the composition and thickness. Developments with in situ monitoring have been instrumental in gaining a better understanding of this technology and in reducing variance in epilayer properties. This has been achieved with better system monitoring to measure organometallic concentration and epilayer monitoring with laser reflectometry or ellipsometry [15.78].

A typical example of a horizontal reactor cell with an internal mercury source is shown in Fig. 15.8. The reactor-cell wall temperature is also determined by the thermodynamics, which requires a minimum temperature in order to avoid condensation of mercury and these requirements can be simply determined from the phase diagram. The organometallic supply shows that the DMCd is injected beyond the Hg source, which is the most common practice to avoid reaction in the Hg zone.

Much of the work on precursors was concerned with thermal stability of the tellurium source as this limits the growth temperature. From thermodynamic considerations, it would be desirable to grow at low temperature (below 400 °C) where the requirement on Hg equilibrium pressure will be lower. The initial demonstration of growth using MOVPE was with diethyl telluride (DETe), which requires a temperature above 400 °C for efficient pyrolysis [15.76].

The most successful tellurium source has been diisopropyltelluride (DiPTe), which is used for growth between 350 and 400 °C. In practice, the DiPTe bubbler is held at a temperature above ambient (up to 30 °C) and the lines to the reactor are heated to avoid condensation of the precursor in the feed lines.

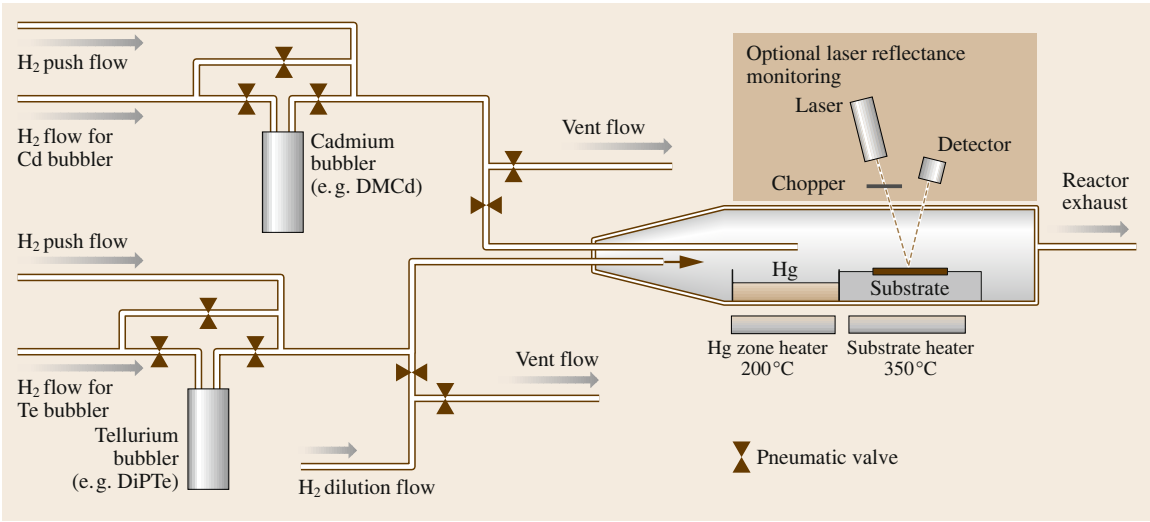


Fig. 15.8 Horizontal reactor cell with internal mercury source and gas supply suitable for the growth of $Hg_{1-x}Cd_xTe$ alloy. (After [15.2])

Unlike MOVPE growth of other II-VI and III-V semiconductors, the growth of an alloy in narrow-gap II-VIs can be achieved by one of two alternative processes: direct alloy growth (DAG) and the interdiffused multilayer process (IMP). The requirement on alloy uniformity is critical because this will determine the cut-on wavelength of a detector. The DAG approach suffers from disadvantages in the thermodynamics and reaction chemistry in that the reaction rate of the tellurium precursor is very different for growing the mercury-containing binary compared with Zn, Cd, or Mn. This leads to a preferential depletion of the non-Hg group II species as the gas flow passes over the substrate, and consequent composition changes in the flow direction.

The interdiffused multilayer process [15.79] separates the growth of HgTe and CdTe so that the growth of the binaries can be independently optimized. The procedure entails the growth of alternate layers of HgTe and CdTe that interdiffuse during the time of growth to give a homogeneous epitaxial layer. The IMP approach relies on the very high interdiffusion coefficients in the MCT pseudobinary and the basic principles are as follows:

1. Interdiffusion must occur at the growth temperature and a nominal time at the end of growth for the last periods to interdiffuse must be allowed.
2. There should be no residual compositional or structural modulation attributable to the IMP oscillations.
3. Interdiffusion at heterointerfaces should be no greater than for direct alloy growth (DAG) at that temperature.
4. Flow velocity is adjusted between the two binary growths to optimize growth uniformity.

The alloy composition is fixed by adjusting the times for growth of the HgTe and CdTe IMP layers. Figure 15.9 shows the uniformity of wavelength over a 75 mm diameter MOVPE layer grown on a GaAs substrate currently in the author's lab [15.80]. By introducing substrate rotation, uniformity was improved dramatically such that, to within a few millimeters of the wafer edge, the composition is uniform to within ± 0.004. The thickness uniformity also improved to within $\pm 2-3\%$. This was compatible with the production of twelve sites of large format 2-D arrays (640×512 diodes on 24 μm pitch) per layer or larger numbers of smaller arrays and was comparable with the uniformity achieved by *Edwall* [15.81]. In the past, other groups have established MOVPE reactor designs capable of large-area uniform growth of MCT on 75 mm wafers [15.81, 82] but these activities have now ceased as attention has focused on MBE growth techniques.

A Polish group has also been very active in MOVPE growth of MCT [15.83]. Growth takes place on 50 mm GaAs substrates, with iodine and arsenic being used as n- and p-type dopants, respectively, to produce fully doped heterostructures.

Piotrowski and Piotrowski [15.10, Chap. 22] have reported that MOVPE-grown material used in uncooled MCT photodetectors can bring their performance close to fundamental limits. These near-room-temperature photodetectors have found increasingly widespread civilian (pyrometry and thermography; gas analysis based on conventional, laser and Fourier transform spectroscopy; free space high transfer rate optical com-

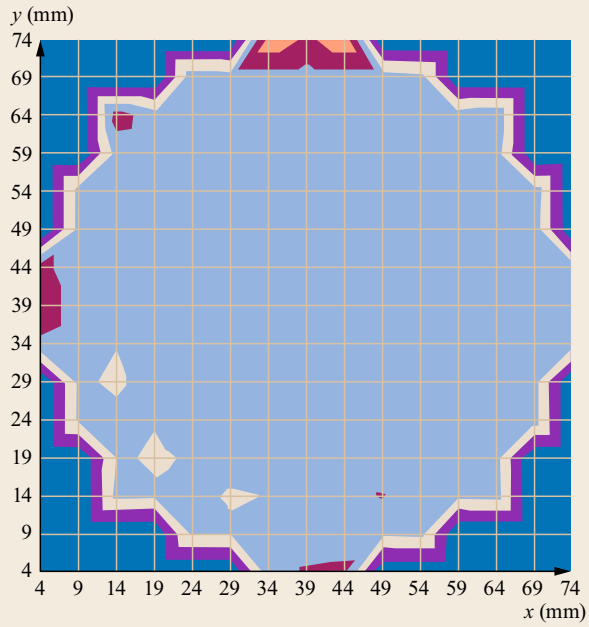


Fig. 15.9 Uniformity of wavelength over a 75 mm diameter MOVPE layer grown on a GaAs substrate. (After [15.10, Chap. 6])

munications; test equipment) and military (night vision, laser range finder and threat warning devices, gun sights, smart munitions) applications.

Maxey et al. [15.74] recently reported MOVPE-grown material being used in various applications at several wavelengths. This included single-color, both MW (MW) and LW (LW), dual-band (MW/LW) and MW higher operating temperature (HOT) devices grown on GaAs 75 mm substrates. Figure 15.10 shows a secondary ion mass spectrometery (SIMS) profile from a dual-band device heterostructure [15.10, Chap. 6]. Device sizes were 1024×768 on a 16 μm pitch (MW) and 640×512 on a 24 μm pitch (LW). The following year [15.84] the same group reported SW avalanche photodiodes made in MOVPE at 1.5−1.8 and 2.0−2.4 μm at a size of 320×256 on a 24 μm pitch, while the MW detectors had increased in size to 1920×1080 on a 12 μm pitch (Fig. 15.11), with a concept device of a 16-Mpixel detector made up of a 4×2 matrix of such arrays. In addition, MW 384×384 on a 20 μm pitch detectors were demonstrated that exhibited fast frame rates of 1000 frames s^{-1} (Fig. 15.12).

The usual methods for determining depth uniformity of an MCT layer are sharpness of the infrared absorption edge and SIMS depth profiles (in particular looking at the Te^{125} secondary ion). Results from

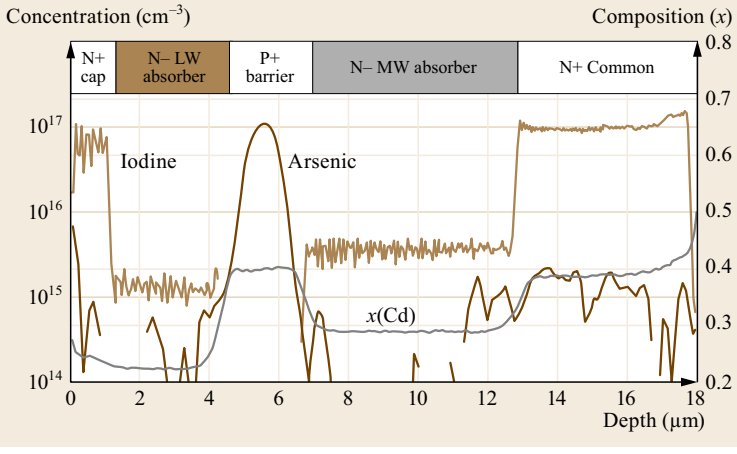


Fig. 15.10 Secondary ion mass spectrometery (SIMS) profile from a dual band device heterostructure. (After [15.10, Chap. 6])

Fig. 15.11 Image taken using FALCON MWIR 1920 × 1080/12 μm pitch. (Courtesy of Leonardo MW Ltd)

these techniques indicate that the IMP structure is fully diffused for IMP periods of the order of 1000 Å and growth temperature in the range of 350 to 400 °C unless the surface becomes faceted during growth when microinhomogeneities may occur [15.85].

15.3.1 Substrate Type and Orientation

The search for the correct substrate material and orientation has been a major area for research in MCT because it is a limiting factor in the quality of the epilayers [15.10, Chap. 6]. Essentially, there are two categories of substrates: (i) lattice-matched II-VI substrates and (ii) nonlattice-matched *foreign substrates*. Examples of the former are CdZnTe and CdSeTe where the alloy compositions are tuned to the lattice parameter of the epilayer. Nonlattice-matched substrates include GaAs, Si and sapphire. The lattice mismatches can be up to 20% but, remarkably, heteroepitaxy is still obtained. The need for a ternary substrate to avoid substantial numbers of misfit dislocations has made the development of the CdTe-based substrate more complex. The small mismatch with CdTe substrates (0.2%) is sufficient to increase the dislocation density to greater than $10^6\,\mathrm{cm}^{-2}$, comparable with some layers on CdTe-buffered GaAs where the mismatch is 14% [15.81]. An additional problem encountered with the lattice-matched substrates is the lamella twins that form on (111) planes in Bridgman-grown crystals. It is possible to cut large (4 × 6 cm) (111) oriented substrates parallel to the twins but (100) substrates have a much lower yield. MOVPE-grown CdTe also twins on the (111), and these twins will propagate through an entire structure.

Hence, the majority of MOVPE growth has concentrated on orientations close to (100), normally with a misorientation to reduce the size of macrodefects, known as hillocks or pyramids. In an analysis of the frequency and shapes of defects on different misorientations, it was concluded [15.86] that the optimum orientation was (100) 3–4° towards the (111)B face. An alternative approach has been to use the (211)B

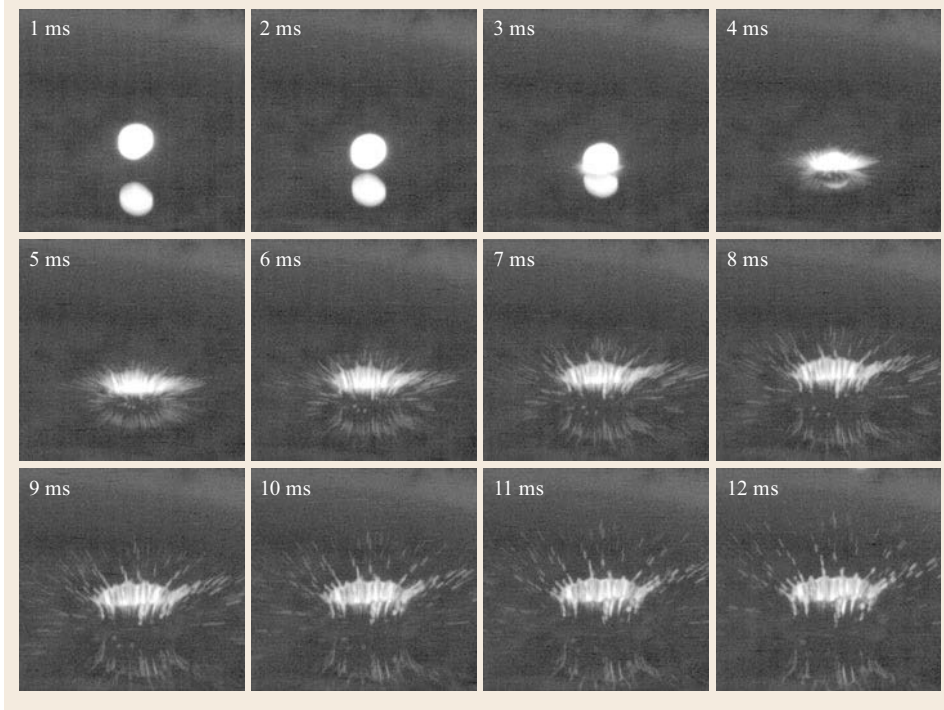


Fig. 15.12 Hot water drop entering cold water 1000 fps, 384 × 384 pixels. (Courtesy of Leonardo MW Ltd)

orientation. In this case the surface appears to be free of macrodefects and is sufficiently misoriented from the (111) to avoid twinning. Dislocation densities of $10^5\,cm^{-2}$ have been measured in MCT grown by IMP onto CdZnTe (211)B substrates [15.81], and diffusion-limited detectors have been fabricated using this orientation [15.87].

The alternative lattice-mismatched substrates were investigated as a more producible alternative to the variable quality of the CdTe family of substrates. As the MCT arrays have to be cooled during operation, there is the risk that the differential thermal contraction between the substrate and multiplexer will break some of the indium contacting columns. The ideal substrate from this point of view is, of course, silicon but, initially, the quality of heteroepitaxy with 20% lattice mismatch was poor. One of the more successful alternative substrate technologies has been the Rockwell PACE-I (producible alternative to CdTe epitaxy), which uses c-plane sapphire with a CdTe buffer layer grown by MOVPE and an MCT detector layer grown by LPE. The sapphire substrates absorb above 6 μm and can only be used for the 3–5 μm waveband. However, even with careful substrate preparation a buffer layer thicker than 5 μm is needed to avoid contamination of the active layer.

GaAs has been the most extensively used alternative substrate, which has been successfully used to reduce the macrodefect density to below $10\,cm^{-2}$ and X-ray rocking curve widths below 100 arcsec have been obtained [15.88]. Due to the large lattice mismatch, the layer nucleates with rafts of misfit dislocations that relieve any strain.

The majority of the MOVPE growth onto silicon substrates has used a GaAs buffer layer to step the change in lattice parameter between silicon and CdTe. By growing a 12 μm thick CdTe buffer layer, the X-ray rocking curve width is reduced to 120 arcsec, larger than the best values obtained on GaAs substrates [15.89]. The advantage of thermal expansion match with the silicon multiplexer has been demonstrated with a midwave 256×256 array where the entire structure was grown by MOVPE [15.90].

15.3.2 Doping

The earliest FPAs were made in as-grown, undoped MOVPE layers where the p-type nature was due to metal vacancies. However, for current and future FPAs, the more complex devices such as p/n heterostructures require extrinsic doping and low-temperature annealing in a mercury vapor to remove metal vacancies. Activation of donors is relatively easy, using either Group III metals or Group VII halogens. Indium has been used for doping concentrations from $< 10^{15}\,cm^{-3}$ up to $10^{18}\,cm^{-3}$. For the p/n structures, the n-type side of the junction is the absorber layer and therefore minority carrier lifetime and diffusion length are also important. An alternative n-type doping approach is to use a halogen such as iodine or chlorine that substitutes on the Group VI sites. The first use of iodine doping [15.91] used a crystalline iodine source, where the vapor was transported via a hydrogen flow and separate injector. The main advantage of iodine over indium is the slower diffusion, which enables more abrupt doped structures to be grown. A liquid source, ethyl iodide (EtI) was used by *Mitra* et al. [15.92] to dope over a range from 10^{15} to $10^{18}\,cm^{-3}$, with no apparent memory. Current work on doped heterostructures [15.80] uses isobutyl iodide (IBI) as with this source donor levels over a more useful range (mid-10^{14} to mid-$10^{17}\,cm^{-3}$) can be achieved.

Acceptor doping proved to be more problematic in MOVPE than donor doping. The first attempts at arsenic doping used AsH_3 that would normally be very stable at growth temperatures of 350 °C. It was proposed by *Capper* et al. [15.93] that the AsH_3 forms an adduct with DMCd in the vapor and causes the deposition of As at a low temperature. It was found that the doping incorporation was much higher in the CdTe layers and that activation of the dopant could be changed by adjusting the VI/II ratio during growth. A second-generation acceptor dopant source (tris(dimethylamino)arsine – DMAAs) is now the preferred route to introduce As into layers. Multilayer, fully doped heterostructures have been grown [15.80, 94], [15.10, Chap. 6] which demonstrate higher operating temperature performance due to Auger suppression techniques. These heterostructures were grown to investigate the device performance in the 3–5 μm and 8–12 μm IR bands. Figure 15.10 shows a SIMS profile of a 3–5 μm device structure grown on a 75 mm diameter GaAs substrate. The excellent control of the As concentration allows the active region to be doped at $4–5 \times 10^{15}\,cm^{-3}$ in the so-called *π-layer* of the heterostructure design. The iodine profile shows well-defined transitions that help define the junction depth. Such layers have been fabricated into mesa 2-D arrays that have been indium-bumped onto Si multiplexers. This material represents the current state of the art in MOVPE.

15.3.3 In situ Monitoring

The use of optical in situ monitoring probes for MOVPE is helping to both understand the kinetic mechanisms and provide monitors suitable for feedback control. Although in situ monitoring is not widely used in MCT growth there have been a number of notable examples where progress has been made as a result

of its application. Laser reflectometry measures the bulk of the film, measuring film thickness, growth rate and some indication of film composition. Reflection difference spectroscopy (RDS) is sensitive to surface composition and will monitor surface reconstruction and adsorbate composition. Spectroscopic ellipsometry is very sensitive to film composition and is useful as an in situ monitor with fast multichannel analysis. A number of in situ characterization techniques have been outlined in a review [15.78], [15.10, Chap. 6].

15.4 Molecular Beam Epitaxy (MBE)

An alternative to MOVPE for the production of advanced multilayer heterostructures is growth by MBE. Figure 15.13 shows a schematic of a simple molecular beam epitaxial system for the growth of semiconductors. Significant progress was made in MCT MBE technology in the 1990s [15.95, 96], [15.10, Chap. 7]. Current MCT MBE technology offers low-temperature growth under an ultrahigh-vacuum environment, in situ n-type and p-type doping, and control of composition, doping and interfacial profiles. All these are essential for the growth of advanced double-layer and multilayer structures for advanced FPAs.

A typical system consists of eight interconnected chambers, including two growth chambers, two preparation chambers with integrated ion etching capabilities and entry chambers, an ESCALAB chamber, two buffer chambers and a transport chamber. Each growth chamber is equipped with dual Hg reservoirs, several effusion cells and dual load locks for effusion cells to enable recharging of the materials without breaking the vacuum. Systems are modular in nature and designed to accommodate sample sizes up to 75 mm in diameter. Garland [15.10, Chap. 7] recently summarized all the issues pertaining to MBE growth and concluded that MBE can now produce LPE-like quality in layers grown on CdZnTe. However, more work still needs to be done to improve growth on CdTe/Si, partly in terms of defect densities and fully active p-type doping.

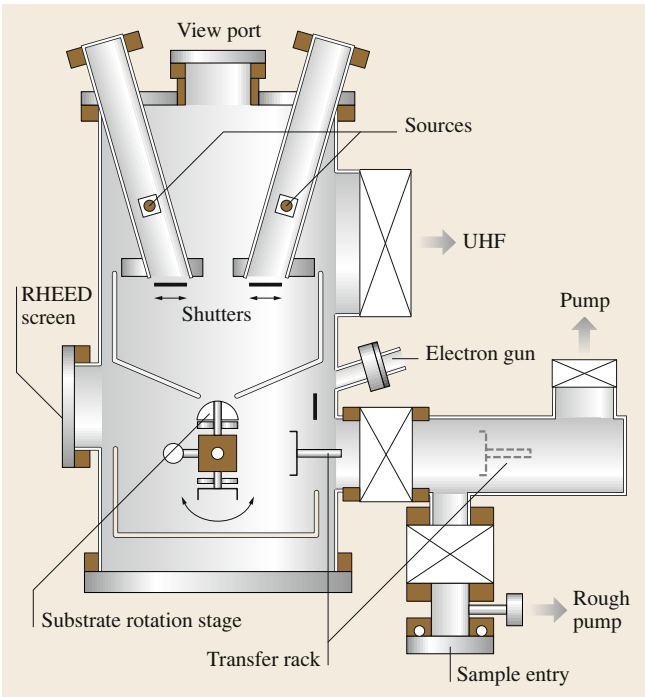


Fig. 15.13 Schematic of a simple molecular beam epitaxial system for the growth of semiconductors

Normally, growth of MCT is carried out at 180–190 °C on (211) CdZnTe substrates using Hg, CdTe and Te sources. However, *Garland* and *Sporken* [15.10, Chap. 4] have discussed the issues around various substrate systems, i. e., CdZnTe, GaAs and Si. They concluded that CdTe/Si is best for SW and MW growth, but CdZnTe is still preferred for LW applications. Typically, the CdZnTe substrates are degreased, etched in Br-methanol solution, rinsed with methanol, then blown dry with nitrogen and mounted on holders, and then thermally cleaned at 350–400 °C under vacuum to remove oxides prior to growth [15.97]. Although MBE is essentially a nonequilibrium growth process, the vapor pressure of the constituent species over an MCT layer is governed by thermodynamics. It has been shown [15.98] that the vapor pressure of Hg over an MCT layer at 170 °C is $\approx 7 \times 10^{-4}$ mbar and 8 mbar for the Te-rich and Hg-rich sides of the phase boundary, respectively. The value of the Hg vapor pressure on the Te-rich side of the phase boundary is similar to the beam equivalent pressure of Hg employed for the growth of MCT by MBE, suggesting that MCT growth occurs on the Te-rich side of the phase boundary, and reflection high-energy electron diffraction studies confirm this.

The MCT alloy composition can be varied readily by choosing the appropriate beam-flux ratios. Over the range from $x = 0.2$–0.50, excellent control of composition can be achieved readily by varying the CdTe source flux with a constant flux of Hg at 3×10^{-4} mbar and Te at 8×10^{-7} mbar during the MBE growth [15.96].

The most widely used n-type dopant for MCT alloys during MBE growth is indium [15.99, 100]. The In concentration can be varied from 2×10^{15} to 5×10^{18} cm^{-3} by adjusting the In cell temperature (450–700 °C) with

no evidence of a memory effect. The doping efficiency of In was almost 100%, evident from the Hall measurement and secondary ion mass spectrometry (SIMS) data, for carrier concentrations $< 2 \times 10^{18}\,\mathrm{cm}^{-3}$.

A critical issue in growing advanced MCT structures is the ability to grow high-quality p-type materials in situ. As, Sb, N, Ag and Li have all been used as acceptors during MBE growth of MCT, with varying degrees of success [15.101]. Most data available is centered on the use of arsenic and two approaches have been investigated. The first is based on photo-assisted MBE to enable high levels of p-type As-doping of CdTe [15.100]. For As-doping during compositionally modulated structure growth only the CdTe layers in a CdTe-MCT combination are doped, as in MOVPE growth. Since the CdTe does not contain Hg vacancies, and is grown under cation-rich conditions, the As is properly incorporated on the Te site and its concentration is proportional to the As flux. The structure then interdiffuses after annealing at high temperature to remove residual Hg vacancies, resulting in p-type, homogeneous MCT. The main disadvantage of this approach is that it requires a high-temperature anneal that results in reduced junction and interface control. An alternative approach is to use cadmium arsenide and correct Hg/Te ratios to minimize Hg vacancies during MCT growth [15.99]. As a result, the As is directed to the group VI sublattice to promote efficient p-type doping. A comparison of the net hole concentration and the SIMS measurement indicates an electrical activity of As acceptors in excess of 60%. *Garland* et al. [15.102] recently discussed the issue of As doping of MCT and conclude that achieving high-lifetime p-type material, i. e., not limited by Shockley–Read–Hall recombination, remains a very important task. Several areas of future work in this area were proposed.

Lateral compositional and thickness uniformity were evaluated on a $2.5 \times 2.5\,\mathrm{cm}^2$ sample and the results showed that the average alloy composition and thickness were $x = 0.219 \pm 0.0006$ and $t = 8.68 \pm 0.064\,\mu\mathrm{m}$, respectively [15.96].

The surface morphology of MCT layers is important from the device fabrication point of view. Scanning electron microscopy (SEM) studies indicate that the surface morphology of MBE-grown MCT alloys is very smooth for device fabrication, except for occasional small undulations ($< 1\,\mu\mathrm{m}$). The crystal quality of MBE MCT layers is illustrated by X-ray rocking curve data for an LWIR double-layer heterojunction structure. The In-doped n-type (about $8\,\mu\mathrm{m}$ thick) base layer peak has a width of < 25 arcsec and is indistinguishable from the CdZnTe substrate. Because the As-doped p-type cap layer is much thinner (about $2\,\mu\mathrm{m}$) and has a different alloy composition, its peak is broader (45 arcsec), but the X-ray FWHM (full width at half-maximum) width still indicates high quality [15.96].

Other material properties such as minority carrier lifetime and etch pit density (EPD) of the material are important for device performance. The lifetime of the photoexcited carriers is among the most important, since it governs diode leakage current and quantum efficiency of detectors. In the case of In-doped n-type layers ($x = 0.2$–0.3) results show that the lifetime ranges from 0.5–$3\,\mu\mathrm{s}$ depending upon the x value and carrier concentration. Measured lifetimes at 77 K approach the Auger recombination limit [15.103].

Extended defects including dislocations, pinholes, particulates, inclusions, microtwins, precipitates and inhomogeneities, occur over lengths of several micrometers and are often observable by selective etching and microscopy. Optical microscopy studies of etched MBE-grown MCT base layers indicate that the EPD of device-quality alloys is of the order of $2 \times 10^5\,\mathrm{cm}^{-2}$, which is in good agreement with the lifetime data and comparable to the best LPE materials for device applications.

15.4.1 Double-Layer Heterojunction Structures

To achieve these structures there have been two approaches pursued in MCT MBE, namely ion implantation and in situ doping [15.101]. The advantage of the in situ doping approach is that it is a simple layer-by-layer growth process and so it is relatively easy to grow multilayer structures. Reproducible p-type doping with low defect density is difficult to achieve, and it requires a very stringent passivation for mesa structures. However, high-performance MCT 64×64 and 128×128 FPAs for MWIR and LWIR infrared detection, using MCT double-layer heterojunction structures (DLHJ) grown by MBE have been demonstrated [15.104]. The performance of detectors fabricated from MWIR and LWIR layers compared very well to the LPE-based production process. An update in this area can be found in [15.10, Chap. 7].

15.4.2 Multilayer Heterojunction Structures

As infrared technology continues to advance, there is a growing demand for multispectral detectors. The original growth process developed for the DLHJ structures was applied and extended for the growth of n-p-p-n multilayer heterojunction (MLHJ) structures for two-color detector applications [15.105]. A SIMS profile of an MLHJ structure confirms that the in situ doping process is able to incorporate $> 2 \times 10^{18}\,\mathrm{cm}^{-3}$ As and control the n-type doping at 1–$2 \times 10^{15}\,\mathrm{cm}^{-3}$ for

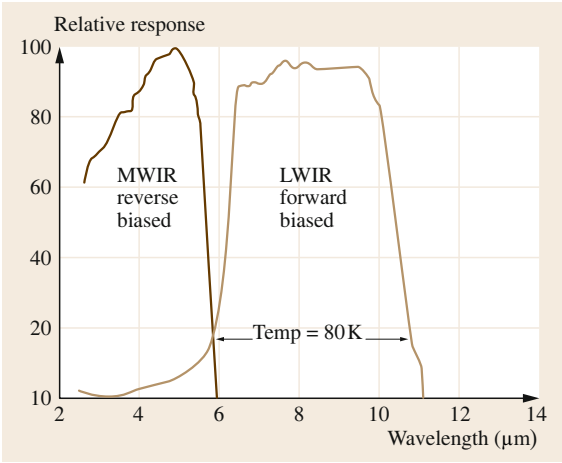


Fig. 15.14 The spectral response data for the MW-LWIR two-color detectors fabricated from an n-p-p-n multilayer heterojunction structure. The data indicate that the MWIR is detected in negative bias mode and LWIR is detected under forward bias conditions, as expected. (After [15.2])

both absorbing layers [15.106]. This dual-band detector is basically a four-layer structure with two p-n junctions, one for each spectral band. The top and bottom layers are the IR absorbers while the middle two layers are transparent. The spectral response of this MW/LW two-color detector is shown in Fig. 15.14. This demonstration indicated that MCT MBE in situ doping technology is capable of growing multilayer structures for advanced IR FPAs. An update in this area can be found in [15.10, Chap. 7].

15.4.3 MCT and CdZnTe Growth on Silicon

The move to the use of Si substrates for the epitaxial growth of MCT FPAs is motivated by several important technological factors [15.96]. Primarily there is the need to establish a thermal expansion match between the Si readout electronics chip and the array of MCT infrared detectors. As the size of such MCT hybrid IR FPAs increases, long-term thermal-cycle reliability can be compromised by the thermal expansion mismatch between Si and bulk CdZnTe substrates. One approach to solving this problem requires the fabrication of the MCT array on an Si substrate. Considerable research activity [15.107, 108] was directed towards achieving epitaxial deposition of high crystalline quality MCT films on Si substrates. Beyond their thermal expansion match to the readout electronics ship, Si substrates for IR FPAs offer additional advantages when compared with CdZnTe substrates, for example they are more readily available in larger sizes with superior mechanical strength and at substantially lower cost. The use of Si substrates also avoids potential problems associated with outdiffusion of fast-diffusing impurities, such as Cu, that has been identified [15.109] as a problem with CdZnTe substrates.

However, the most serious technical challenge faced in fabricating device-quality epitaxial layers of MCT on Si is reduction in the density of threading dislocations resulting from the accommodation of the 19% lattice constant mismatch and the large difference in thermal expansion coefficient between Si and MCT. Dislocation density is known to have a direct effect on IR detector performance [15.58], particularly at low temperature. All efforts to fabricate MCT IR detectors on Si substrates relied upon the prior growth of CdZnTe buffer layers on Si. Growth of $\approx 5\,\mu m$ of CdZnTe is required to allow dislocation annihilation processes to decrease the dislocation density to low $10^6\,cm^{-2}$ [15.108].

As an additional step, initiation layers of ZnTe have been used to facilitate parallel MBE deposition of CdTe(001) on Si(001) [15.104]; ZnTe nucleation layers are also commonly used for this same purpose for growth of CdZnTe on GaAs/Si substrates by other vapor-phase techniques [15.110]. CdTe(001) films with rocking curves as narrow as 78 arcsec and EPD of $1-2\times10^6\,cm^{-2}$ have been demonstrated with this technique. Both (111)- and (001)-oriented MBE CdTe/Si substrates have been used as the basis for demonstration of LPE-grown MCT detectors [15.103, 107]. The (001)-oriented CdTe/Si films have been used in the demonstration of 256×256 MCT hybrid arrays on Si [15.104].

Current state-of-the-art MBE material on 100 mm diameter Si substrates has been discussed in [15.10, Chap. 7]. Dry etching is used to produce array sizes of 128×128 and 1024×1024 with performance equivalent to LPE material. Similar material, grown on CdZnTe this time, by another group [15.111] in the MW and SW regions is used in astronomy applications [15.112].

Other applications of MBE-grown MCT include very long wavelength arrays (onto (211)B CdZnTe substrates) by *Philips* et al. [15.113], two-color (MW $4.5\,\mu m$/SW $2.5\,\mu m$) arrays of 128×128 diodes [15.69, 114], gas detectors in the $2-6\,\mu m$ region [15.115] and $1.55\,\mu m$ avalanche photodiodes using Si substrates [15.66]. All these applications demonstrate the versatility of the MCT MBE growth technique.

Recently, *Vilela* et al. [15.116] reported on the growth of dual-band FPAs in material grown on Si of a size 640×480 on a $20\,\mu m$ pitch, using As and In for p- and n-type doping. *Gavrand* et al. [15.70] describe the use of their material, grown on 100 mm GaAs substrates, for space applications, avalanche photodiodes and dual-band detectors. *Ziegler* et al. [15.117] report on the growth of material, again on 100 mm GaAs substrates, and its use in MW 1280×1040 on a $15\,\mu m$

pitch operating at 80–110 K, and LW 1280 × 1040 on a 15 µm pitch operating at 76 K. Reduced dark current was obtained by grading the composition in single-color devices, an advantage of MBE over LPE. An alternative substrate material has been proposed by *Lei* et al. [15.118]. They used GaSb, which is a better lattice match to MCT and has lower FWHM and EPD than equivalent GaAs substrates, but is three times more expensive, and grew CdTe buffer layers that were at least equivalent in quality to those grown on GaAs substrates. No growth of MCT has been reported on this substrate as yet. Garland [15.10, Chap. 16] recently reviewed the field of superlattices for very-LW detectors. Advantages include bandgap engineering to suppress Auger recombination, a decrease in band-to-band tunneling currents, sharper optical cutoffs whose absorptivities are higher and flatter as a function of wavelength, than their normal MCT counterparts. However, disadvantages include insufficient lateral uniformity across layers, difficulty in p-type doping and interdiffusion issues during p-type dopant activation and device processing.

15.5 Alternatives to MCT

Rogalski and coworkers [15.1, 119, 120] have given details about several Hg-based alternatives to MCT for infrared detection. He concludes that from the range of possibilities only HgZnTe and HgMnTe are serious candidates. Theoretical considerations of *Sher* et al. [15.121] showed that the Hg–Te bond is stabilized by addition of ZnTe, unlike the destabilization that occurs when CdTe is added, i. e., as in MCT.

The pseudobinary phase diagram of HgZnTe shows even more separation of solidus and liquidus than does the equivalent for MCT (Fig. 15.15). This leads to large segregation effects, large composition variations for small temperature changes, as well as the high Hg vapor pressure presenting the usual problems of containment. HgTe and MnTe are not completely miscible over the entire range, the single-phase region is limited to $\approx x < 0.35$. The solidus–liquidus separation in the HgTe-MnTe pseudobinary is ≈ half that in MCT, therefore for equivalent wavelength uniformity requirements crystals of the former must be much more uniform than MCT crystals.

Three methods – Bridgman, SSR and THM – are the most popular ones for the bulk growth of HgZnTe and HgMnTe. The best quality crystals have been produced by THM [15.123], with uniformities of ±0.01 in both axial and radial directions for HgZnTe. For HgMnTe, *Bodnaruk* et al. [15.124] produced crystals of $0.04 < x < 0.2$ with uniformities of ±0.01 and ±0.005 in the axial and radial directions, while *Gille* et al. [15.125] grew $x = 0.10$ crystals with ±0.003 along a 16 mm diameter slice. *Becla* et al. [15.126] achieved a decrease in radial variations by applying a magnetic field.

These bulk crystals are suitable for some device architectures but higher performance is obtained in more

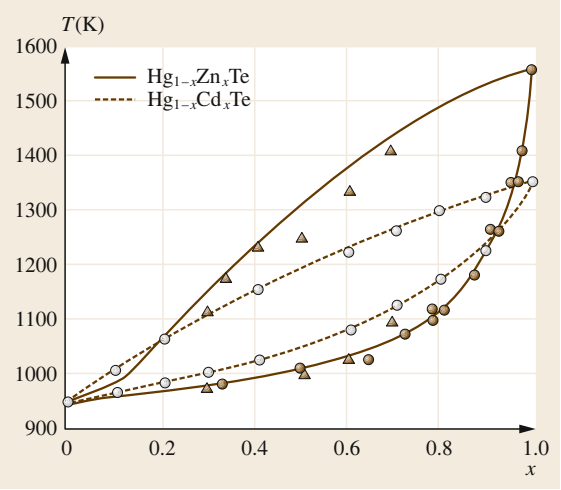


Fig. 15.15 HgTe-ZnTe and HgTe-CdTe pseudobinary phase diagrams. (Reproduced from [15.122])

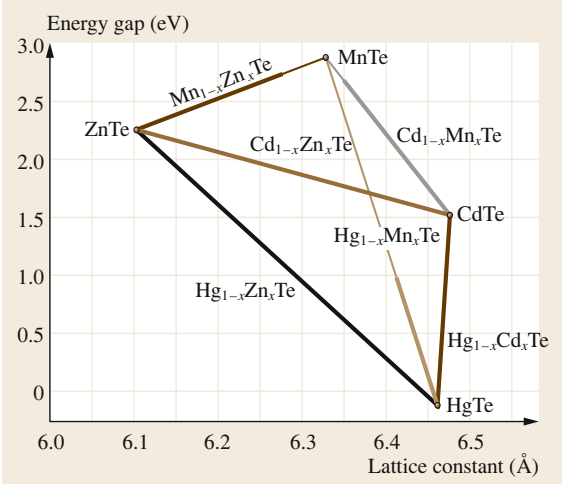


Fig. 15.16 Energy gap versus lattice parameter in Hg-based ternary alloy systems at room temperature. The bold lines indicate ranges of x for which homogeneous crystal phases form. (After [15.3])

sophisticated structures that can only be made by epitaxial techniques. The advantages are the usual ones of lower growth temperature, lower Hg vapor pressure, shorter growth times and less precipitation with improved uniformity of composition. These features have been discussed in detail and descriptions of the various epitaxial processes given by *Rogalski* [15.119, 122, 127]. LPE growth of HgCdZnTe and HgCdMnTe by *Uchino* and *Takita* [15.128] showed that incorporation of Zn or Mn improved the uniformity of composition. HgMnTe films have been grown by MOCVD using the IMP technique, with both n- and p-type layers being produced. As in all epitaxial growth techniques there is a need for suitable (i. e., nearly lattice-matched) large-area single-crystal substrates. Figure 15.16 shows the lattice constants of the Hg-based alloy systems at room temperature. There are larger differences between the lattice parameters of HgTe and ZnTe and HgTe and MnTe than between HgTe and CdTe. At first sight it would appear that CdMnTe would be a suitable substrate system for these ternary alloys, but CdMnTe crystals grown by the Bridgman process are highly twinned and therefore unsuitable for use as substrates.

Acknowledgments. The author wishes to thank the management of Leonardo MW Ltd for permission to publish this chapter.

References

15.1 P. Capper (Ed.): *Properties of Narrow Gap Cadmium-based Compounds*, EMIS Datareview Series, Vol. 10 (IEE, London 1994)

15.2 P. Capper (Ed.): *Narrow-gap II-VI Compounds for Optoelectronic and Electromagnetic Applications* (Chapman & Hall, London 1997)

15.3 P. Capper, C.T. Elliott (Eds.): *Infrared Detectors and Emitters: Materials and Devices* (Springer, New York 2000)

15.4 P. Capper (Ed.): *Bulk Crystal Growth of Electronic, Optical and Optoelectronic Materials* (Wiley, Chichester 2005)

15.5 I. Baker: II–IV Narrow-Bandgap Semiconductors for Optoelectronics. In: *Springer Handbook of Electronic and Photonic Materials*, ed. by S. Kasap, P. Capper (Springer, New York 2006) p. 855

15.6 P. Capper, M. Mauk (Eds.): *Liquid Phase Epitaxy of Electronic, Optical and Optoelectronic Materials* (Wiley, Chichester 2007)

15.7 J. Piotrowski, A. Rogalski: *High-Operating-Temperature Infrared Photodetectors* (SPIE Press, Bellingham 2007)

15.8 M.A. Kinch: *Fundamentals of Infrared Detector Materials* (SPIE Press, Bellingham 2007)

15.9 C.T. Elliott: Proc. SPIE **7298**, 7298-2S (2009)

15.10 P. Capper, J. Garland (Eds.): *Mercury Cadmium Telluride: Growth, Properties and Applications* (Wiley, Chichester 2011)

15.11 A. Rogalski: Hg-based alternatives to MCT. In: *Infrared Detectors and Emitters: Materials and Devices*, ed. by P. Capper, C.T. Elliott (Springer, New York 2000)

15.12 W.F.H. Micklethwaite: Semicond. Semimet., Vol. 18 (1981), Ch. 3

15.13 P.W. Kruse: Semicond. Semimet., Vol. 18 (1981), Ch. 1

15.14 R. Triboulet, T. Nguyen Duy, A. Durand: J. Vac. Sci. Technol. A **3**, 95 (1985)

15.15 P. Capper, J. Harris, D. Nicholson, D. Cole: J. Cryst. Growth **46**, 575 (1979)

15.16 A.W. Vere, B.W. Straughan, D.J. Williams, N. Shaw, A. Royle, J.S. Gough, J.B. Mullin: J. Cryst. Growth **59**, 121 (1982)

15.17 L. Colombo, A.J. Syllaios, R.W. Perlaky, M.J. Brau: J. Vac. Sci. Technol. A **3**, 100 (1985)

15.18 R.K. Sharma, V.K. Singh, N.K. Mayyar, S.R. Gupta, B.B. Sharma: J. Cryst. Growth **131**, 565 (1987)

15.19 J. Ziegler: Method for the production of monocrystalline $Hg_{1-x}Cd_xTe$, US Patent 4 591 410 (1986)

15.20 W.M. Higgins, G.N. Pultz, R.G. Roy, R.A. Lancaster: J. Vac. Sci. Technol. A **7**, 271 (1989)

15.21 J.H. Tregilgas: Prog. Cryst. Growth Charact. **28**, 57 (1994)

15.22 P. Capper: Prog. Cryst. Growth Charact. **28**, 1 (1994)

15.23 P. Capper, J.J.G. Gosney: U.K. Patent Ser., Vol. 8115911 (1981)

15.24 P. Capper, C. Maxey, C. Butler, M. Grist, J. Price: J. Mater. Sci. Mater. Electron. **15**, 721 (2004)

15.25 W.M. Higgins, D.A. Nelson, R.G. Roy, R.P. Murasako, R.A. Lancaster, J. Tower, P. Norton: J. Electron. Mater. **42**, 3320 (2013)

15.26 R. Triboulet: Prog. Cryst. Growth Charact. **28**, 85 (1994)

15.27 Y. Nguyen Duy, A. Durand, J. Lyot: Mater. Res. Soc. Symp. Proc. **90**, 81 (1987)

15.28 A. Durand, J.L. Dessus, T. Nguyen Duy, J. Barbot: Proc. SPIE **659**, 131 (1986)

15.29 P. Gille, F.M. Kiessling, M. Burkert: J. Cryst. Growth **114**, 77 (1991)

15.30 P. Gille, M. Pesia, R. Bloedner, N. Puhlman: J. Cryst. Growth **130**, 188 (1993)

15.31 M. Royer, B. Jean, A. Durand, R. Triboulet: Procédé de préparation d'un lingot cristallin de HgCdTe, French Patent 8 804 370 (1988)

15.32 R.U. Bloedner, P. Gille: J. Cryst. Growth **130**, 181 (1993)

15.33 B. Chen, J. Shen, S. Din: J. Electron Mater. **13**, 47 (1984)

15.34 D.A. Nelson, W.M. Higgins, R.A. Lancaster: Proc. SPIE **225**, 48 (1980)

15.35 C.-H. Su, G. Perry, F. Szofran, S.L. Lehoczky: J. Cryst. Growth **91**, 20 (1988)
15.36 R.R. Galazka, et al.: J. Cryst. Growth **53**, 397 (1981)
15.37 B. Bartlett, P. Capper, J. Harris, M. Quelch: J. Cryst. Growth **47**, 341 (1979)
15.38 A. Durand, J.L. Dessus, T. Nguyen Duy: Proc. SPIE **587**, 68 (1985)
15.39 P. Capper, C.D. Maxey, C.L. Jones, J.E. Gower, E.S. O'Keefe, D. Shaw: J. Electron Mater. **28**, 637 (1999)
15.40 M.A. Kinch: Mater. Res. Soc. Symp. Proc. **90**, 15 (1987)
15.41 R. Pratt, J. Hewett, P. Capper, C. Jones, N. Judd: J. Appl. Phys. **60**, 2377 (1986)
15.42 R. Pratt, J. Hewett, P. Capper, C.L. Jones, M.J.T. Quelch: J. Appl. Phys. **54**, 5152 (1983)
15.43 F. Grainger, I. Gale, P. Capper, C. Maxey, P. Mackett, E. O'Keefe, J. Gosney: Adv. Mater. Opt. Electron. **5**, 71 (1995)
15.44 A.W. Vere: Proc. SPIE **659**, 10 (1986)
15.45 D.J. Williams, A.W. Vere: J. Vac. Sci. Technol. A **4**, 2184 (1986)
15.46 J.H. Tregilgas, J.D. Beck, B.E. Gnade: J. Vac. Sci. Technol. A **3**, 150 (1985)
15.47 C. Genzel, P. Gille, I. Hahnert, F.M. Kiessling, P. Rudolph: J. Cryst. Growth **101**, 232 (1990)
15.48 P. Capper, C. Maxey, C. Butler, M. Grist, J. Price: J. Cryst. Growth **275**, 259 (2005)
15.49 T. Tung: J. Cryst. Growth **86**, 161 (1988)
15.50 T.-C. Yu, R.F. Brebrick: Phase diagrams for HgCdTe. In: *Properties of Narrow Gap Cadmium-based Compounds*, EMIS Datareview Series, Vol. 10, ed. by P. Capper (IEE, London 1994) p. 55
15.51 M.G. Astles: Growth of HgCdTe by liquid phase epitaxy. In: *Properties of Narrow Gap Cadmium-based Compounds*, EMIS Datareview Series No. 10, ed. by P. Capper (IEE, London 1994) p.13
15.52 T. Tung, L.V. DeArmond, R.F. Herald, P.E. Herning, M.H. Kalisher, D.A. Olson, R.F. Risser, A.P. Stevens, S.J. Tighe: Proc. SPIE **1735**, 109 (1992)
15.53 P.W. Norton, P. LoVecchio, G.N. Pultz, J. Hughes, T. Robertson, V. Lukach, K. Wong: Proc. SPIE **2228**, 73 (1994)
15.54 G.H. Westphal, L. Colombo, J. Anderson: Proc. SPIE **2228**, 342 (1994)
15.55 D.W. Shaw: J. Cryst. Growth **62**, 247 (1983)
15.56 I.B. Baker, G.J. Crimes, J. Parsons, E. O'Keefe: Proc. SPIE **2269**, 636 (1994)
15.57 P. Capper, E.S. O'Keefe, C.D. Maxey, D. Dutton, P. Mackett, C. Butler, I. Gale: J. Cryst. Growth **161**, 104 (1996)
15.58 R.S. List: J. Electron. Mater. **22**, 1017 (1993)
15.59 S. Johnson, D. Rhiger, J. Rosbeck, J.M. Peterson, R.E. Taylor, M.E. Bond: J. Vac. Sci. Technol. B **10**, 1499 (1992)
15.60 M. Yoshikawa: J. Appl. Phys. **63**, 1533 (1988)
15.61 P. Capper: J. Vac. Sci. Technol. B **9**, 1667 (1991)
15.62 C.A. Cockrum: Proc. SPIE. **2685**, 2 (1996)
15.63 W.A. Radford, R.E. Kvaas, S.M. Johnson: *Proc. IRIS Specialty Group on Infrared Materials* (Infrared Information Analysis Center, Ann Arbor 1986)
15.64 K.J. Riley, A.H. Lockwood: Proc. SPIE **217**, 206 (1980)
15.65 S.M. Johnson, J.A. Vigil, J.B. James, C.A. Cockrum, W.H. Konkel, M.H. Kalisher, R.F. Risser, T. Tung, W.J. Hamilton, W.L. Ahlgren, J.M. Myrosznyk: J. Electron. Mater. **22**, 835 (1993)
15.66 T. DeLyon, A. Hunter, J. Jensen, M. Jack, V. Randall, G. Chapman, S. Bailey, K. Kosai: paper given at 2002 US Workshop on Physics and Chemistry of II-VI Materials, San Diego, USA (2002)
15.67 S. Johnson, J. James, W. Ahlgren, et al.: Mater. Res. Soc. Symp. Proc. **216**, 141 (1991)
15.68 P.R. Norton: Proc. SPIE **2274**, 82 (1994)
15.69 G. Destefanis, A. Astier, J. Baylet, P. Castelein, J.P. Chamonal, E. De Borniol, O. Gravand, F. Marion, J.L. Martin, A. Million, P. Rambaud, F. Rothan, J.P. Zanatta: J. Electron. Mater. **32**, 592 (2003)
15.70 O. Gravand, G. Destefanis, S. Bisotto, N. Baier, J. Rothman, L. Mollard, D. Brellier, L. Rubaldo, A. Kerlain, V. Destefanis, M. Vuillermet: J. Electron. Mater. **42**, 3349 (2013)
15.71 J. Wenisch, H. Bitterlich, M. Bruder, P. Fries, R. Wollrab, J. Wendler, R. Breiter, J. Ziegler: J. Electron. Mater. **42**, 3186 (2013)
15.72 W. Schirmacher, R. Wollrab, H. Lutz, T. Shallenberg, J. Wendler, J. Ziegler: J. Electron. Mater. **42**, 2778 (2013)
15.73 J. Ziegler, J. Wenisch, R. Breiter, D. Eich, H. Figgemeier, P. Fries, H. Lutz, R. Wollrab: J. Electron. Mater. **42**, 3186 (2013)
15.74 C.D. Maxey, I. Baker, J. Fitzmaurice, J. Price, J. Wilson, L. Hipwood, L. Pillans, P. Knowles: paper presented at US II-IV Workshop, Seattle, 2012 (unpublished)
15.75 T.F. Kuech, J.O. McCaldin: J. Electrochem. Soc. **128**, 1142 (1981)
15.76 S.J.C. Irvine, J.B. Mullin: J. Cryst. Growth **55**, 107 (1981)
15.77 A.C. Jones: J. Cryst. Growth **129**, 728 (1993)
15.78 S.J.C. Irvine, J. Bajaj: Semicond. Sci. Technol. **8**, 860 (1993)
15.79 J. Tunnicliffe, S. Irvine, O. Dosser, J. Mullin: J. Cryst. Growth **68**, 245 (1984)
15.80 C.D. Maxey, J. Camplin, I.T. Guilfoy, J. Gardner, R.A. Lockett, C.L. Jones, P. Capper: J. Electron. Mater. **32**, 656 (2003)
15.81 D.D. Edwall: J. Electron. Mater. **22**, 847 (1993)
15.82 S. Murakami: J. Vac. Sci. Technol B **10**, 1380 (1992)
15.83 P. Madejczyk, A. Piotrowski, W. Gawron, K. Klos, J. Pawluczyk, J. Rutkowski, J. Piotrowski, A. Rogalski: Opto-Electron. Rev. **13**(3), 239 (2005)
15.84 C.D. Maxey, I. Baker, L. Hipwood, P. Thorne, A. Bradford, A. Ashcroft, G. Finger: paper presented at US II-IV Workshop, Chicago, 2013 (unpublished)
15.85 S.J.C. Irvine, D. Edwall, L. Bubulac, R.V. Gil, E.R. Gertner: J. Vac. Sci. Technol. B **10**, 1392 (1992)
15.86 D.W. Snyder, S. Mahajan, E.I. Ko, P.J. Sides: Appl. Phys. Lett. **58**, 848 (1991)
15.87 P. Mitra, Y.L. Tyan, F.C. Case, R. Starr, M.B. Reine: J. Electron. Mater. **25**, 1328 (1996)
15.88 A.M. Kier, A. Graham, S.J. Barnett, J. Giess, M.G. Astles, S.J.C. Irvine: J. Cryst. Growth **101**, 572 (1990)

15.89 S.J.C. Irvine, J. Bajaj, R.V. Gil, H. Glass: J. Electron. Mater. **24**, 457 (1995)
15.90 S.J.C. Irvine, E. Gertner, L. Bubulac, R.V. Gil, D.D. Edwall: Semicond. Sci. Technol. **6**, C15 (1991)
15.91 C.D. Maxey, P. Whiffin, B.C. Easton: Semicond. Sci. Technol. **6**, C26 (1991)
15.92 P. Mitra, Y.L. Tyan, T.R. Schimert, F.C. Case: Appl. Phys. Lett. **65**, 195 (1994)
15.93 P. Capper, C. Maxey, P. Whiffin, B. Easton: J. Cryst. Growth **97**, 833 (1989)
15.94 C.D. Maxey, C.J. Jones, N. Metcalfe, R.A. Catchpole, N.T. Gordon, A.M. White, C.T. Elliot: Proc. SPIE **3122**, 453 (1996)
15.95 R.D. Rajavel, D. Jamba, O.K. Wu, J.A. Roth, P.D. Brewer, J.E. Jensen, C.A. Cockrum, G.M. Venzor, S.M. Johnson: J. Electron. Mater. **25**, 1411 (1996)
15.96 O.K. Wu: T.J. deLyon, R.D. Rajavel, J.E. Jensen: Molecular beam epitaxy of HgCdTe. In: *Narrow-gap II-VI Compounds for Optoelectronic and Electromagnetic Applications*, ed. by P. Capper (Chapman & Hall, London 1997) p. 97
15.97 O.K. Wu, D.R. Rhiger: Other compound semiconductor films. In: *Characterization in Compound Semiconductor Processing*, ed. by Y. Strausser, G.E. McGuire (Butterworth-Heinemann, Oxford 1995) p. 83
15.98 T. Tung, L. Golonka, R.F. Brebrick: J. Electrochem. Soc. **128**, 451 (1981)
15.99 O. Wu, D. Jamba, G. Kamath: J. Cryst. Growth **127**, 365 (1993)
15.100 J. Arias, S. Shin, D. Copper, M. Zandian, J.G. Pasko, E.R. Gernter, R.E. De Wames, J. Singh: J. Vac. Sci. Technol. A **8**, 1025 (1990)
15.101 O.K. Wu: Mater. Res. Soc. Symp. Proc. **340**, 565 (1994)
15.102 J.W. Garland, C. Green, S. Sivananthan: J. Electron. Mater. **42**, 3331 (2013)
15.103 V. Lopes, A.J. Syllaios, M.C. Chen: Semicond. Sci. Technol. **8**, 824 (1993)
15.104 S.M. Johnson, T.J. de Lyon, C. Cockrum, W.J. Hamilton, T. Tung, F.I. Gesswein, B.A. Baumgratz, L.M. Ruzicka, O.K. Wu, J.A. Roth: J. Electron. Mater. **24**, 467 (1995)
15.105 G. Kamath, O. Wu: Method of growing p-type group II-VI material, US Patent 5 028 561 (1991)
15.106 O.K. Wu, R.D. Rajavel, T.J. de Lyon, J.E. Jensen, C.A. Cocrum, S.M. Johnson, G.M. Venzor, G.R. Chapman, J.A. Wilson, E.A. Patten, W.A. Radford: Proc. SPIE **2685**, 16 (1996)
15.107 F.T. Smith, P.W. Norton, P. Lo Vecchio, N. Hartle, M. Weiler: J. Electron. Mater. **24**, 1287 (1995)
15.108 T.J. DeLyon, D. Rajavel, O.K. Wu: Proc. SPIE. **2554**, 25 (1995)
15.109 J.P. Tower, S.P. Tobin, M. Kestigian, P.W. Norton, A.B. Bollong, H.F. Schaake, C.K. Ard: J. Electron. Mater. **24**, 497 (1995)
15.110 N. Karam, R. Sudharsanan, A. Mastrovito, M.M. Sanfacon, F.T.J. Smith, M. Leonard, N.A. El-Masry: J. Electron. Mater. **24**, 483 (1995)
15.111 M. Zandian, J.D. Garnett, R.E. DeWames, M. Carmody, J.G. Pasko, M. Farris, C.A. Cabelli, D.E. Cooper, G. Hildebrandt, J. Chow, J.M. Arias, K. Vural, D.N.B. Hall: J. Electron. Mater. **32**, 803 (2003)
15.112 I.S. McLean: paper given at 2002 US Workshop on Physics and Chemistry of II-VI Materials, San Diego, USA (2002)
15.113 J.D. Philips, D.D. Edwall, D.L. Lee: J. Electron. Mater. **31**, 664 (2002)
15.114 L.A. Almeida, M. Thomas, W. Larsen, K. Spariosu, D.D. Edwall, J.D. Benson, W. Mason, A.J. Stolz, J.H. Dinan: J. Electron. Mater. **31**, 669 (2002)
15.115 J.P. Zanatta, F. Noel, P. Ballet, N. Hdadach, A. Million, G. Destefanis, E. Mottin, E. Picard, E. Hadji: J. Electron. Mater. **32**, 602 (2003)
15.116 M.F. Vilela, K.R. Olsson, E.M. Norton, J.M. Peterson, K. Rybnicek, D.H. Rhiger, C.W. Fulk, J.W. Bangs, D.D. Lofgreen, S.M. Johnson: J. Electron. Mater. **42**, 3231 (2013)
15.117 J. Ziegler, J. Wenisch, R. Breiter, D. Eich, H. Figgemeier, P. Fries, H. Lutz, R. Wollrab: J. Electron. Mater. **43**, 2955 (2014)
15.118 W. Lei, R.J. Gu, J. Antoszewski, J. Dell, L. Faraone: J. Electron. Mater. **43**, 2788 (2014)
15.119 A. Rogalski: *New Ternary Alloy Systems for Infrared Detectors* (SPIE Optical Engineering Press, Bellingham 1994)
15.120 J. Piotrowski, A. Rogalski: *High-Operating-Temperature Infrared Photodetectors* (SPIE, Bellingham 2007)
15.121 A. Sher, A.B. Chen, W.E. Spicer, C.K. Shih: J. Vac. Sci. Technol. A **3**, 105 (1985)
15.122 A. Rogalski: Prog. Quantum Electron. **13**, 299 (1989)
15.123 R. Triboulet: J. Cryst. Growth **86**, 79 (1988)
15.124 O.A. Bodnaruk, I.N. Gorbatiuk, V.I. Kalenik, et al.: Neorg. Mater. **28**, 335 (1992), in Russian)
15.125 P. Gille, U. Rössner, N. Puhlmann, et al.: Semicond. Sci. Technol. **10**, 353 (1995)
15.126 P. Becla, J.-C. Han, S. Matakef: J. Cryst. Growth **121**, 394 (1992)
15.127 A. Rogalski: Infrared Phys. **31**, 117 (1991)
15.128 T. Uchino, K. Takita: J. Vac. Sci. Technol. A **14**, 2871 (1996)

16. Wide-Bandgap II–VI Semiconductors: Growth and Properties

Minoru Isshiki, Jifeng Wang

Wide-bandgap II–VI compounds are been applied to optoelectronic devices, especially light-emitting devices in the short-wavelength region of visible light, because of their direct gap and suitable bandgap energies. Many methods have been extensively applied to grow high-quality films and bulk single crystals from the vapor and liquid phases.

This chapter firstly discusses the basic properties and phase diagrams of wide-bandgap II–VI compounds such as ZnS, ZnO, ZnSe, ZnTe, CdSe and CdTe. Then the growth methods and recent progress in films and bulk crystal growth are reviewed. In the epitaxial growth methods, the focus is on liquid-phase epitaxy (LPE), vapor-phase epitaxy (VPE) containing conventional VPE, hot-wall epitaxy (HWE), metalorganic chemical vapor deposition (MOCVD) or metalorganic phase epitaxy (MOVPE), molecular-beam epitaxy (MBE) and atomic-layer epitaxy (ALE). In bulk crystal growth, two typical growth methods, chemical/physical vapor transport (CVT/PVT) and Bridgman techniques, are introduced.

16.1 **Crystal Properties** 367
16.1.1 Basic Properties 367
16.1.2 Phase Diagram 368

16.2 **Epitaxial Growth** 371
16.2.1 The LPE Technique 371
16.2.2 Vapor-Phase Epitaxy Techniques 372

16.3 **Bulk Crystal Growth** 375
16.3.1 The CVT and PVT Techniques 375
16.3.2 Hydrothermal Growth 377
16.3.3 Bridgman and Gradient Freezing (GF) Method 379
16.3.4 The Traveling Heater Method (THM) 380
16.3.5 Other Methods 380

16.4 **Conclusions** 381

References 381

Wide-bandgap II–VI compounds are expected to be one of the most vital materials for high-performance optoelectronics devices such as light-emitting diodes (LEDs) and laser diodes (LDs) operating in the blue or ultraviolet spectral range. Additionally, the high ionicity of these compounds makes them good candidates for high electro-optical and electromechanical coupling. The basic promises of wide-bandgap materials can be found in Fig. 16.1.

Thin films were commonly grown using the conventional vapor-phase epitaxy (VPE) method for 60 s. With the development of science and technology, new and higher requirements arose for material preparation. For this reason, novel epitaxial growth techniques were developed, including hot-wall epitaxy (HWE) [16.1], metalorganic chemical vapor deposition (MOCVD) [16.2], molecular-beam epitaxy (MBE) [16.3], metalorganic molecular-beam epitaxy (MOMBE) [16.4] and atomic-layer epitaxy (ALE) [16.5]. Using these growth methods, film thickness can be controlled, and quality can be improved.

On the other hand, basic research work into growing bulk crystals of wide-bandgap II–VI compounds has been carried out. Focus was put on high-purity, high-quality, large single crystals [16.6–10]. Since the electrical and optical properties of semiconductor compounds are drastically affected by impurities and native defects, purity and quality are very important for fundamental research and engineering application where they are used as substrates. Bulk single crystals of these wide-bandgap II–VI compounds have been grown from the vapor, liquid and solid phases. Vapor-phase growth includes chemical vapor transport (CVT) and physical vapor transport (PVT) methods; liquid-phase methods includes growth from the melt or solvent. Among these growth methods, melt growth is most suitable to pro-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_16

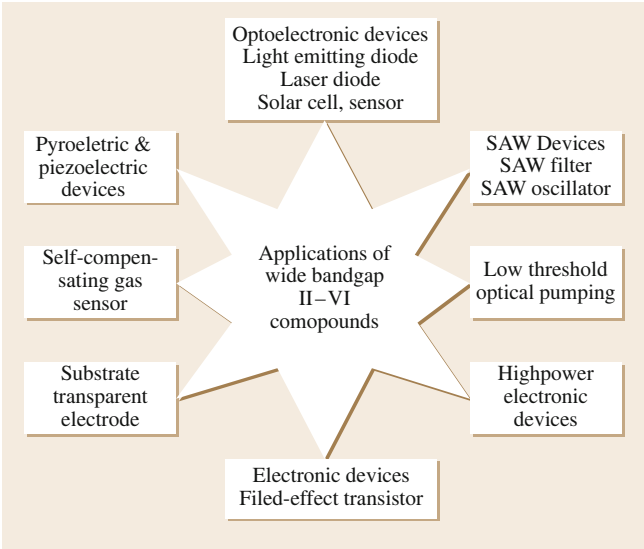


Fig. 16.1 Application of II–VI wide-bandgap compounds

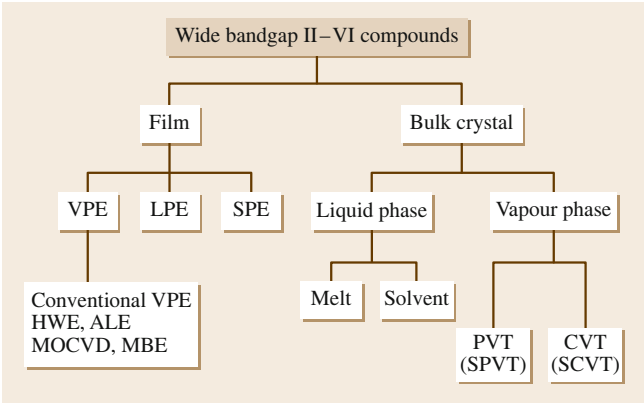


Fig. 16.2 Film and bulk-crystal growth techniques for II–VI wide-bandgap compounds

duce sizable bulk crystals for relatively short growth duration. Growth methods for films and bulk crystals of wide-bandgap II–VI compounds are summarized in Fig. 16.2 and the details can be found in Sects. 16.2 and 16.3.

This chapter firstly describes the physical and chemical properties of these wide-bandgap II–VI compounds, then reviews the growth techniques and introduces the main results in preparing film and bulk single crystals of ZnS, ZnO, ZnSe, ZnTe, CdTe, and so on.

16.1 Crystal Properties

16.1.1 Basic Properties

Wide-bandgap compound semiconductors have higher melting points. Due to their high ionicity, the overheating phenomenon occurs when they are heated to their melting point. Owing to the higher vapor pressures at their melting points, it is difficult to grow bulk crystals from melt. On the other hand, it is easy to grow bulk crystals as well as their films from the vapor phase. Therefore, before introducing film and bulk crystal growth, it is necessary to review their physical and chemical properties. Table 16.1 shows the properties of some of the main II–VI compound semiconductor materials [16.11–24].

Table 16.1 Properties of some wide-bandgap II–VI compound semiconductors

Material Property	ZnS	ZnO	ZnSe	ZnTe	CdS	CdSe	CdTe
Melting point (K)	2038 (WZ, 150 atm)	2248	1797	1513	2023 (WZ, 100 atm)	1623	1370 (ZB)
Energy gap E_g at 300 K (eV) (ZB*/WZ*)	3.68/3.911	–/3.4	2.71/–	2.394	2.50/2.50	–/1.751	1.475
$dE_g/dT(\times 10^{-4}$ eV/K) ZB/WZ	4.6/8.5	–/9.5	4.0/–	5.5/–	–/5.2	–/4.6	5.4/–
Structure	ZB/WZ	WZ	ZB/WZ	ZB	WZ	WZ	ZB
Bond length (μm)	2.342 (WZ)	1.977 (WZ)	2.454 (ZB)	2.636 (ZB)	2.530 (ZB)	2.630 (ZB)	2.806 (ZB)
Lattice constant (ZB) a_0 at 300 K (nm)	0.541	–	0.567	0.610	0.582	0.608	0.648
ZB nearest-neighbor dist. at 300 K (nm)	0.234	–	0.246	0.264	0.252	0.263	0.281
ZB density at 300 K (g/cm^3)	4.11	–	5.26	5.65	4.87	5.655	5.86
Lattice constant (WZ) at 300 K (nm)							
$a_0 = b_0$	0.3811	0.32495	0.398	0.427	0.4135	0.430	–
c_0	0.6234	0.52069	0.653	0.699	0.6749	0.702	–
c_0/a_0	1.636	1.602	1.641	1.637	1.632	1.633	–
WZ density at 300 K (g/cm^3)	3.98	5.606	–	–	4.82	5.81	–
Symmetry ZB/WZ	C6me/F43m	–/C6me	–/F43m	–/F43m	C6me/F43m	C6me/F43m	–/–
Electron affinity χ (eV)			4.09	3.53	4.79	4.95	4.28
Stable phase(s) at 300 K	ZB & WZ	WZ	ZB	ZB	ZB & WZ	ZB & WZ	ZB
Solid–solid phase transition temperature (K)	1293	–	1698	–	–	403	1273(?)
Heat of crystallization ΔH_{LS} (kJ/mol)	44	62	52	56	58	45	57
Heat capacity C_P (cal/mol K)	11.0	9.6	12.4	11.9	13.2	11.8	–
Ionicity (%)	62	62	63	61	69	70	72
Equilibrium pressure at c.m.p. (atm)	3.7	–	1.0	1.9	3.8	1.0	0.7
Minimum pressure at m.p. (atm)	2.8	7.82	0.53	0.64	2.2	0.4–0.5	0.23
Specific heat capacity (J/(g K))	0.469	–	0.339	0.16	0.47	0.49	0.21
Thermal conductivity (W/(cm K))	0.27	0.6	0.19	0.18	0.2	0.09	0.01

Table 16.1 (continued)

Material Property	ZnS	ZnO	ZnSe	ZnTe	CdS	CdSe	CdTe
Thermo-optical coefficient $(dn/dT)(\lambda = 10.6\,\mu m)$	4.7	–	6.1	–	–	–	11.0
Electrooptical coefficient r_{41} (m/V) $(\lambda = 10.6\,\mu m)$	2×10^{-12}	–	2.2×10^{-12}	4.0×10^{-12} $(r_{41} = r_{52} = r_{63})$	–	–	6.8×10^{-12}
Linear expansion coefficient $(10^{-6}\,K^{-1})$ ZB/WZ	–/6.9	2.9/7.2	7.6/–	8.0/–	3.0/4.5	3.0/7.3	5.1/–
Poisson ratio	0.27		0.28				0.41
Dielectric constant $\varepsilon_0/\varepsilon_\infty$	8.6/5.2	8.65/4.0	9.2/5.8	9.3/6.9	8.6/5.3	9.5/6.2	2.27/–
Refractive index ZB/WZ	2.368/2.378	–/2.029	2.5/–	2.72/–	–/2.529	2.5/–	2.72/–
Absorption coeff. (including two surfaces) $(\lambda = 10.6\,\mu m)(cm^{-1})$	≤ 0.15	–	$1-2\times10^{-3}$	–	≤ 0.007	≤ 0.0015	≤ 0.003
Electron effective mass (m^*/m_0)	–0.40	–0.27	0.21	0.2	0.21	0.13	0.11
Hole effective mass m^*_{dos}/m_0	–	–	0.6	circa 0.2	0.8	0.45	0.35
Electron Hall mobility 300 K for $n =$ lowish $(cm^2/(V\,s))$	165	125	500	340	340	650	1050
Hole Hall mobility at 300 K for $p =$ lowish $(cm^2/(V\,s))$	5	–	30	100	340	–	100
Exciton binding energy (meV)	36	60	21	10	30.5	15	12
Average phonon energy (meV) ZB/WZ	16.1/17.1	–	15.1/–	10.8/–	–/13.9	18.9/25.4	5.8/–
Elastic constant $(10^{10}\,N/m^2)$							
C_{11}	1.01±0.05	–	8.10±0.52	0.72±0.01	–	–	5.57
C_{12}	0.64±0.05	–	4.88±0.49	0.48±0.002	–	–	3.84
C_{44}	0.42±0.04	–	4.41±0.13	0.31±0.002	–	–	2.095
Knoop hardness (N/cm^2)	0.18	0.5	0.15	0.13	–	–	0.10
Young's modulus	10.8 Mpsi	–	10.2 Mpsi	–	45 GPa	5×10^{11} dyne/cm^2	3.7×10^{11} dyne/cm^2

m.p. – melting point; c.m.p. – congruent melting point; ZB – zinc blende; WZ – wurtzite

16.1.2 Phase Diagram

It is necessary to understand the phase diagram to grow high-quality film and bulk single crystals. Figure 16.3 shows the phase diagrams reported for ZnS [16.25], ZnSe [16.25, 26], ZnTe [16.25], CdSe [16.27] and CdTe [16.28, 29]. Although much work has been done, there some exact thermodynamic data are still lacking, especially details close to the congruent point. Unfortunately, the phase diagram of ZnO is not available in spite of its growing importance in applications.

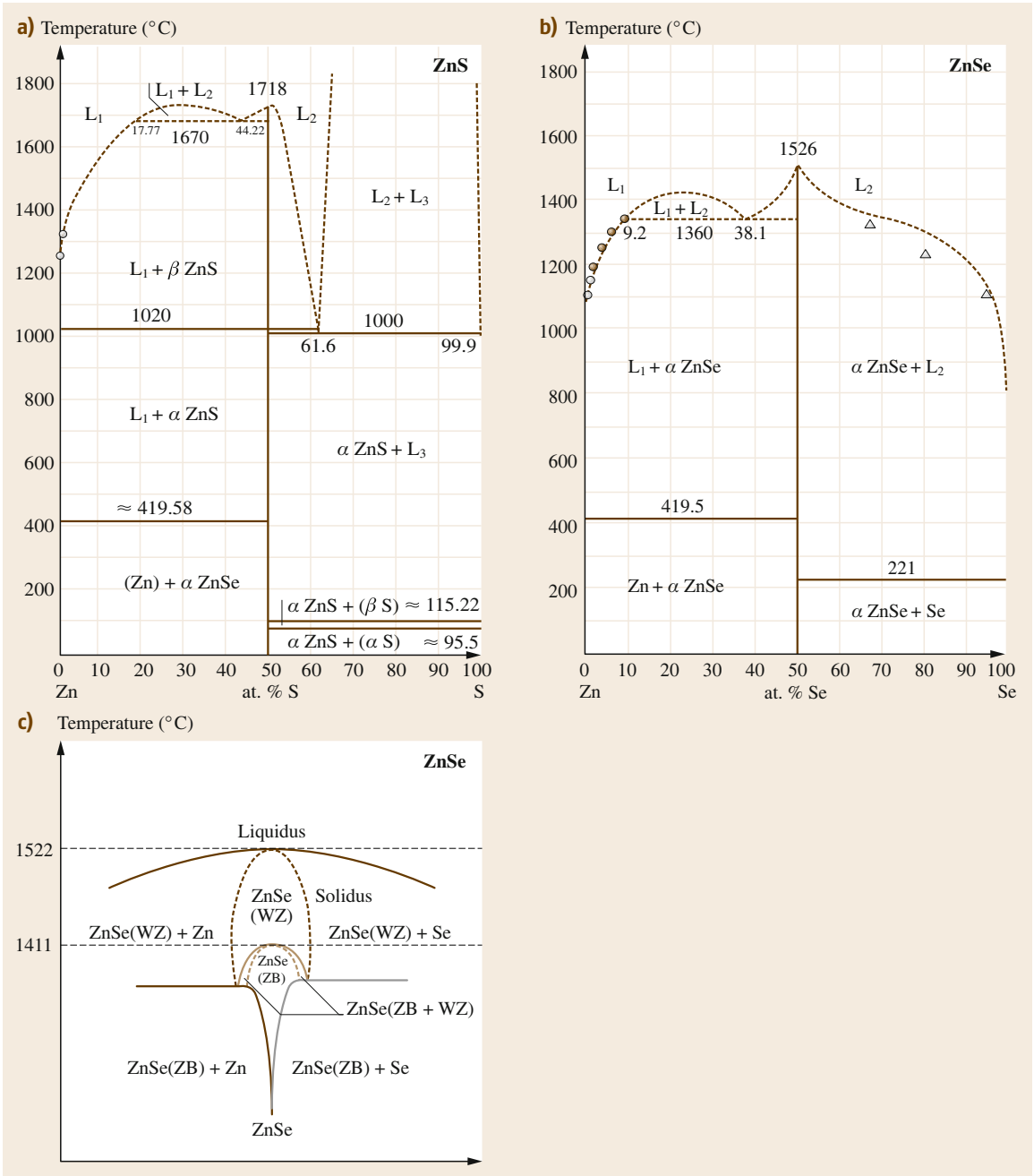


Fig. 16.3a–g Phase diagrams of some main wide-bandgap II–VI compounds. **(a)** ZnS, **(b)** ZnSe, **(c)** ZnSe phase diagram near the congruent point; **(d)** ZnTe; **(e)** CdSe; **(f)** CdTe; **(g)** CdTe phase diagram near the congruent point

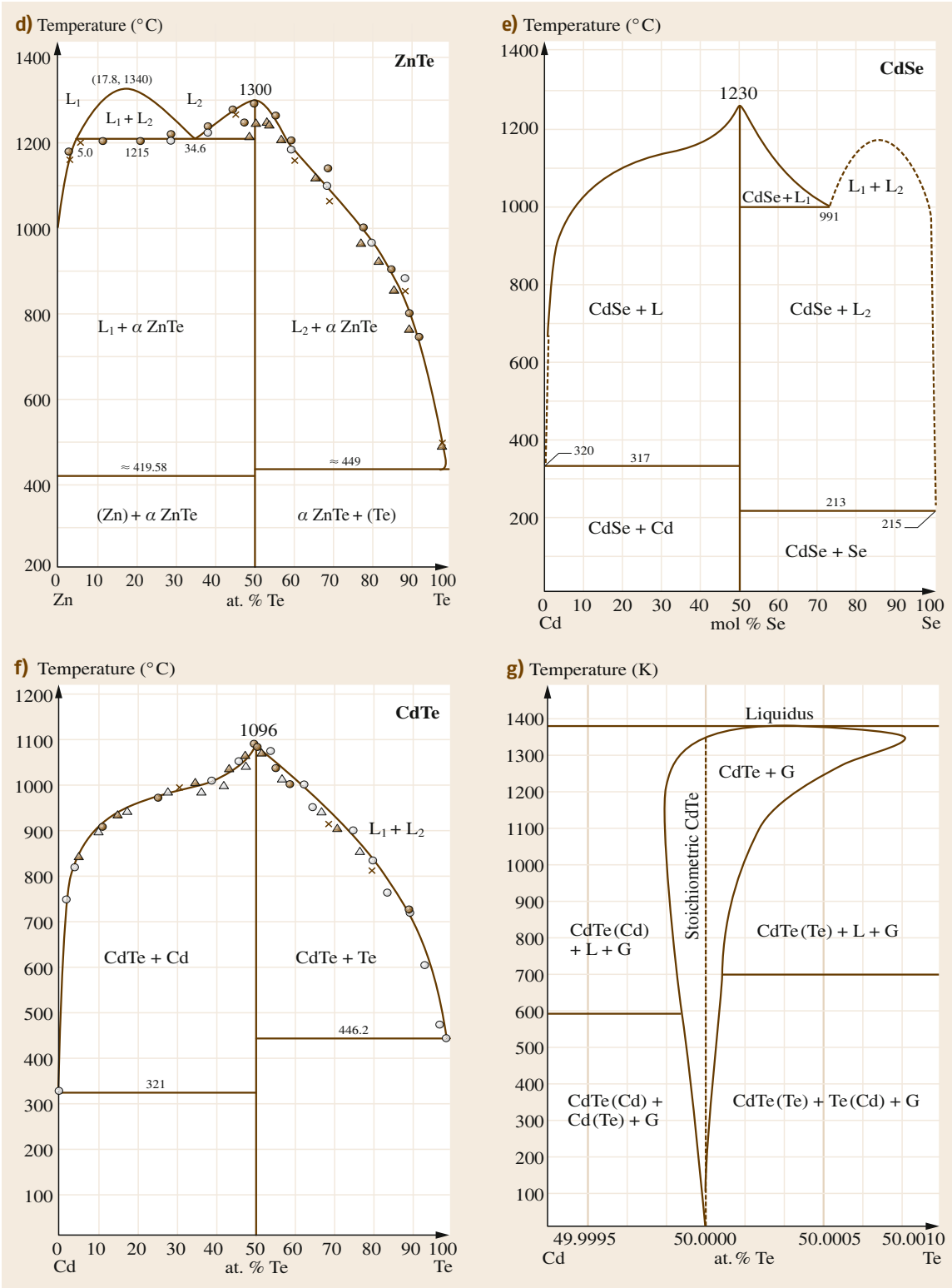


Fig. 16.3 (continued)

16.2 Epitaxial Growth

Epitaxial growth of wide-bandgap II–VI compounds was mainly carried out using liquid-phase epitaxy (LPE), or VPE. VPE includes several techniques, such as conventional VPE, hot-wall epitaxy (HWE), metalorganic chemical vapor deposition (MOCVD) or metalorganic phase epitaxy (MOVPE), molecular-beam epitaxy (MBE), metalorganic molecular-beam epitaxy (MOMBE) and atomic-layer epitaxy (ALE), etc. Each of these methods has its advantages and disadvantages. They are summarized in Table 16.2.

In the case of hetero-epitaxy, the mismatch between substrate material and epitaxial layer affects the growing structure and quality of the epitaxial layer. The mismatch should be made as small as possible when choosing the pair of materials (substrate and epitaxial material). Furthermore, the difference between the thermal expansion coefficients of the pair of materials has to be considered to obtain high-quality epitaxial layer [16.30].

16.2.1 The LPE Technique

LPE growth occurs at near-thermodynamic-equilibrium conditions. There are two growth methods. The first is called equilibrium cooling, in which the saturated solution is in contact with the substrate and the temperature is lowered slowly, the solution becomes supersaturated; meanwhile a slow epitaxial growth on the substrate is initiated. The second is the step-cooling process, in which the saturated solution is cooled down a few degrees (5−20 K) to obtain a supersaturated solution. The substrate is inserted into the solution, which is kept at this cooled temperature. Growth occurs first due to the supersaturation, and will slow down and stop finally. For both techniques, if the substrate is dipped in sequence into several different melt sources, multiple layer structures can be grown. LPE can successfully and inexpensively grow homo- and heterostructures. As the growth is carried out under thermal equilibrium, an epilayer with a very low native defect density can be obtained.

The LPE method can be used to grow high-quality epilayers, such as ZnS [16.31], ZnSe [16.31, 32], ZnSSe [16.33], ZnTe [16.34], etc. *Werkhoven* et al. [16.32] grew ZnSe epilayers by LPE on ZnSe substrates in a low-contamination-level environment. In their study, the width of bound exciton lines in low-temperature photoluminescence spectra was used to define the quality of the material, and the energy of the lines was used to identify trace impurities. The photoluminescence (PL) results showed that the ZnSe epitaxial layer has the high quality. The sharpest spectra occurred in layers grown rapidly on a previously grown buffer layer, indicating the importance of impurity out-diffusion from the substrate into the growing layer. The sharpness of these bound exciton lines indicates that the total concentration of electrically active impurities ($N_A + N_D$) was below $10^{17}\,\mathrm{cm}^{-3}$.

Table 16.2 Strengths and weaknesses of several epitaxial growth techniques

LPE	Thermodynamic equilibrium growth Easy-to-use materials Low-temperature growth High purity Multiple layers Thickness control not very precise Poor surface/interface morphology
HWE	Easy-to-use materials Low cost Thermodynamic equilibrium Hard to grow thick layers Thickness control not very precise
VPE	Easy to operate Economic Thinner layers High growth rates Easier composition control High temperature (800−1000 °C)
ALE	Gaseous reaction for deposition Low-temperature growth Precise composition Low growth rate Safety precautions needed
MOCVD	Gaseous reaction for deposition Precise composition Patterned/localized growth Potentially easier large-area multiple-wafer scale-up Low-temperature growth High-vapor-pressure materials growth allowed About 1 ML/s deposition rate Expensive equipment Safety precautions needed
MBE and MOMBE	Physical vapor deposition Ultra-high-vacuum environment About 1 ML/s deposition rate In situ growth-front monitoring Precise composition Low growth rate Sophisticated equipment Limit for high-vapor-pressure materials growth (MBE)

ML – monolayer

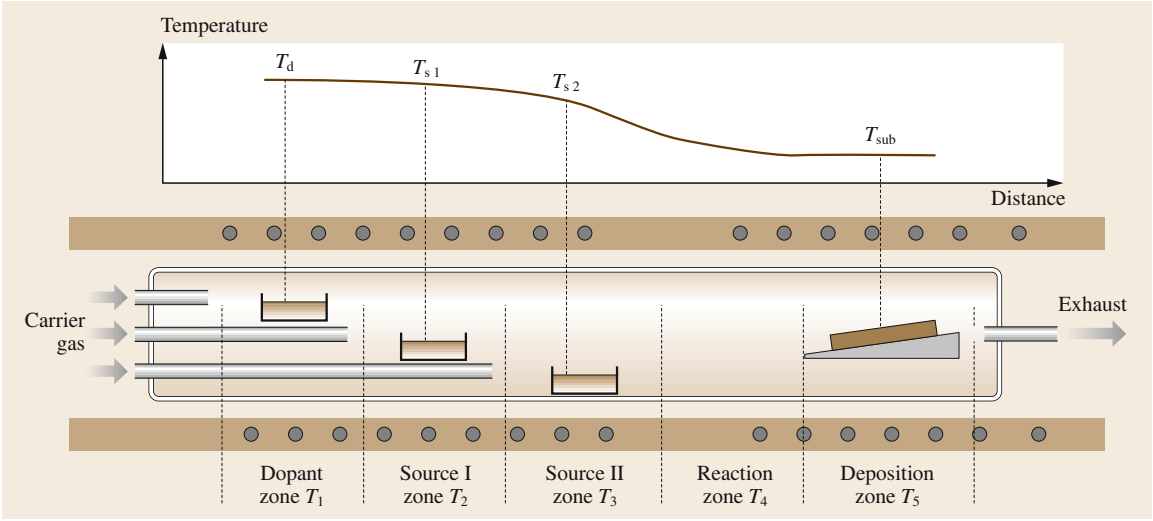


Fig. 16.4 A typical VPE growth system

16.2.2 Vapor-Phase Epitaxy Techniques

Conventional VPE

A typical VPE growth system is shown in Fig. 16.4. In VPE growth, thin films are formed by the deposition of atoms from the vapor phase. There are two types of transport mechanisms for the source materials, physical vapor deposition (PVD) without any chemical reaction, and chemical vapor deposition (CVD), where the formation of the deposited film is the result of a chemical reaction of the precursors on the substrate. In VPE growth, there are several important parameters, such as the source temperature, the substrate temperature, the flow rate of the carrier gas, the growth pressure, and so on. These determine the growth rate, composition, and crystallinity of the epitaxial layers.

The VPE technique is the most popular in semiconductor epitaxial growth. Since the vapor pressures of all wide-bandgap II–VI materials are high, their epitaxial layers can be grown by this method. As an example, high-quality ZnS single-crystal films have been grown on a Si substrate using hydrogen as a carrier gas [16.35, 36]. Furthermore, the Iida group [16.37] doped N and P into a ZnS epilayer and studied their behavior in details. The N and P were expected to compensate the native donor state and to result in an insulating material. The results showed that the doped acceptors N and P reduced the donor density and an insulating material was obtained. Later, this group was successful in preparing a p-type ZnS epilayer using NH_3 as an acceptor dopant [16.38]. Many efforts have been made to grow ZnSe epilayers on different substrates, especially on GaAs in the past two decades [16.39, 40]. p-type ZnSe was also obtained by this technique [16.41]. In addition, other compounds, ZnTe [16.42], CdS [16.43], CdSe [16.44], and CdTe [16.45], were also studied using this technique.

The HWE Technique

Hot-wall epitaxy [16.1] has proved to be a very successful growth method for II–VI compound epitaxial layers. Its principal characteristic is the growth of thin films under conditions near thermodynamic equilibrium. Compared with other VPE methods, the HWE technique has the advantages of low cost, simplicity, convenience, and relatively high growth rate. In particular, it can control deviation from stoichiometry during growth of an epilayer.

A schematic diagram of the improved HWE system is showed in Fig. 16.5; it consists of four independent furnaces. The source material placed at the middle is transported to the substrate. The region of the growth reactor between the source and substrate, called the hot wall, guarantees a nearly uniform and isotropic flux of molecules onto the substrate surface. To control the deviation from stoichiometry, the reservoir part is placed at the bottom with a constituent element.

HWE has been applied to growing II–VI compound epilayers, such as CdTe [16.46], CdS [16.47], CdSe [16.48], ZnTe [16.49], and also to producing heterostructures for laser and photovoltaic detector fabrication [16.50]. Most research using HWE technique has focused on CdTe growth. *Wang* et al. [16.46] optimized growth conditions and grew high-quality CdTe epitaxial films using the HWE apparatus shown in Fig. 16.5. All the CdTe epilayers show mirror-like surfaces. Results from PL and x-ray diffraction (XRD) show that CdTe epilayers on GaAs suffer from a bi-

axial compressive stress, that this stress is rapidly relaxed within a thickness of about 5 μm, and that it remains in the epilayer up to a film thickness of 15 μm. Although this heterosystem has a 14.6% lattice mismatch and −26% thermal expansion mismatch at 300 K, high-quality CdTe epilayers, with a full-width half-maximum (FWHM) of 0.26 meV for bound-exciton emission lines in 4.2-K PL, about 90 arcs for (400) diffraction in four-crystal XRD spectrum, were prepared by selecting suitable growth conditions and epilayer thickness.

Recently, significant results have been achieved for CdTe/Si (111) epilayer growth by HWE. *Lalev* et al. [16.51] reported that high-quality CdTe (111) epilayers with *A* polarity were directly grown on hydrogen-terminated Si (111) without any preheating treatment. Through the originally designed two-step growth regime, the crystal quality of CdTe film was significantly improved, and the best FWHM value of 118 arcs from four-crystal rocking curves was obtained for a 5 μm-thick epilayer.

The MOCVD Technique

MOCVD or MOVPE is an improvement over conventional VPE. Since its introduction in 1968 [16.52], this technique has been established as one of the techniques for epitaxial growth of compound semiconductors both for research and production. The factors that have allowed MOCVD to reach this popularity are the purity and abruptness of the grown layers together with the flexibility of the technique, which makes the growth of almost all compound semiconductors possible. This abrupt transition in the composition of the epitaxial structure is necessary for the fabrication of digital or analog alloy system.

The development history of MOCVD technique is equivalent to that of source precursors. Since ZnSe epilayers were grown by MOCVD [16.52], many source precursors of II–VI elements have been developed. $Zn(CH_3)_2$ dimethylzinc (DMZn) and $Zn(C_2H_5)_2$ diethylzinc (DEZ) were used at the beginning of MOCVD growth [16.53]. ZnSe and ZnS films were grown using these metalorganic sources and inorganic H_2Se or H_2S. Unfortunately, the quality of these films was very poor. For this reason, $Se(CH_3)_2$ dimethylselenide (DMSe) and $Se(C_2H_5)_2$ diethylselenium (DESe) were developed [16.54]. The quality of epilayers was greatly improved. From then, II and VI elemental gas precursors were proposed one after another. *Wright* et al. grew a ZnSe film using (DMZn-$(NEt_3)_2$) triethylamine adduct of dimethylzinc [16.55]. *Hirata* et al. [16.56] and *Nishimura* et al. [16.57] proposed methylselenol (MSeH) and tertiarybutylselenol (*t*-BuSeH) as Se sources, respectively. Methylallylselenide

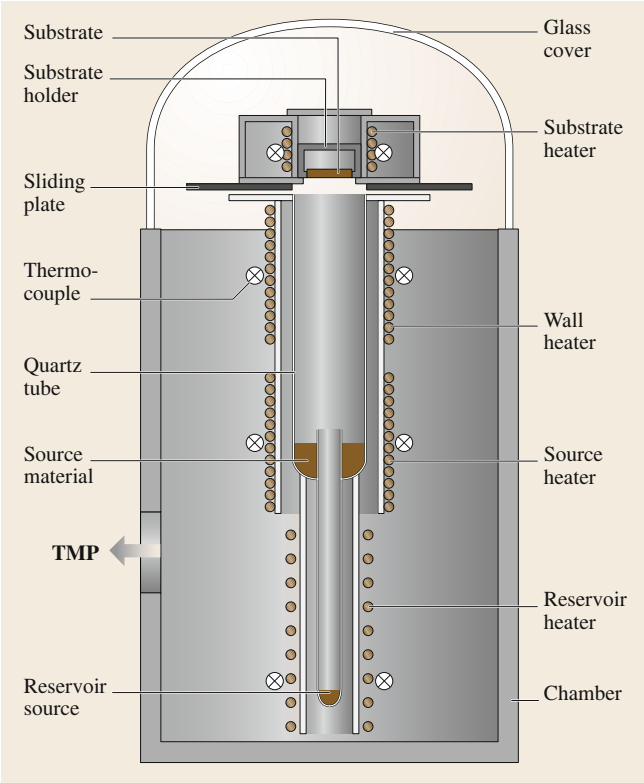


Fig. 16.5 Diagram of a typical HWE system growth chamber

(MASe) [16.58], diallyl-selenide (DASe) [16.59], t-butylallylselenide (*t*-BuASe) [16.60], tertiarybutyl-selenide (*Dt*-BuSe) [16.61], were also used as Se sources. *Fujita* et al. found methylmercaptan (MSH) as an S source [16.53]. Besides these, many other source precursors of II–VI elements, such as $Cd(CH_3)_2$ dimethylcadmium (DMCd) [16.62], $Te(CH_3)_2$ dimethyltelluride (DMTe) [16.62], $Te(C_2H_5)_2$ diethyltelluride (DETe), $Te(C_3H_7)_2$ diisopropyltelluride (DIPTe) [16.63], $S(C_2H_5)_2$ diethylsulfide (DES), $S(C_4H_9)_2$ ditertiarybutylsulfide (DTBS), $(C_4H_9)SH$ tertiarybutylthiol (tBuSH), have been used.

The great advantage of using metalorganics is that they are volatile at moderately low temperatures. Since all constituents are in the vapor phase, precise electronic control of gas flow rates and partial pressures is possible. This, combined with pyrolysis reactions that are relatively insensitive to temperature, allows efficient and reproducible deposition.

The substrate wafer is placed on a graphite susceptor inside a reaction vessel and heated by a radio-frequency (RF) induction heater. The growth temperature depends on the type of compounds grown. Growth is carried out in a hydrogen atmosphere at a pressure of 100–700 torr. The growth precursors decompose on

contact with the hot substrate to form epitaxial layers. Each layer is formed by switching the source gases to yield the desired structure.

The films of almost all wide-bandgap II–VI compounds have been grown by MOCVD technique. Most work has been done on p-ZnSe epilayers in the past two decades [16.64–66]. The highest hole concentration of $8.8 \times 10^{17}\,\mathrm{cm}^{-3}$ was reported with a NH_3 doping source [16.67]. Recently, quantum wells (QWs) and quantum dots (QDs) of these wide-bandgap compounds have become the focus. Successful pulsed laser operation at 77 K in ZnCdSe/ZnSe/ZnMgSSe QW-structure separated-confinement heterostructures has been realized [16.68].

MBE and MOMBE

MBE was developed at the beginning of the 1970s to grow high-purity high-quality compound semiconductor epitaxial layers on some substrates [16.69, 70]. To date, it has become a very important technique for growing almost all semiconductor epilayers. An MBE system is basically a vacuum evaporation apparatus. The pressure in the chamber is commonly kept below $\approx 10^{-11}$ torr. Any MBE process is dependent on the relation between the equilibrium vapor pressure of the constituent elements and that of the compound [16.71]. There are a number of features of MBE that are generally considered advantageous for growing semiconducting films: the growth temperature is relatively low, which minimizes any undesirable thermally activated processes such as diffusion; the epilayer thickness can be controlled precisely; and the introduction of different vapor species to modify the alloy composition and to control the dopant concentration can be conveniently achieved by adding different beam cells with proper shutters. These features become particularly important in making structures involving junctions.

Metalorganic molecular-beam epitaxy growth (MOMBE) is one of the variations of the MBE system [16.72, 73]. The difference is that metalorganic gaseous sources are used as the source materials. Therefore, this growth technique has the merits of MOCVD and MBE.

MBE or MOMBE techniques have been used to grow epilayers of almost all wide-bandgap II–VI semiconductors [16.74, 75]. Due to its features, it is very successful in growing super-thin layers, such as single quantum wells (SQWs), multiple quantum wells (MQWs) [16.76, 77] and nanostructures [16.78].

In nanostructures, quantum dot (QD) structures have attracted a lot of attention in recent years. This field represents one of the most rapidly developing areas of current semiconductor. They present the utmost challenge to semiconductor technology, rendering possible fascinating novel devices. QD are nanometer-size semiconductor structures where charge carriers are confined in all three spatial dimensions. They are neither atomic nor bulk semiconductor, but may best be described as artificial atoms.

In the case of heteroepitaxial growth there are three different growth modes [16.79]: (a) Frank–van der Merwe (FM) or layer-by-layer growth, (b) Volmer–Weber (VW) or island growth, and (c) Stranski–Krastanov (SK) or layer-plus-island growth. Which growth mode will be adopted in a given system depends on the surface free energy of the substrate, (σ_s), that of the film, (σ_f), and the interfacial energy (σ_i). Layer-by-layer growth mode occurs when $\Delta\sigma = \sigma_f + \sigma_i - \sigma_s = 0$. The condition for FM-mode growth is rigorously fulfilled only for homoepitaxy, where $\sigma_s = \sigma_f$ and $\sigma_i = 0$. If the FM-mode growth condition is not fulfilled, then three-dimensional crystals form immediately on the substrate (VW mode). For a system with $\Delta\sigma = 0$ but with a large lattice mismatch between the substrate and the film, initial growth is layer-by-layer. However, the film is strained. As the film grows, the stored strain energy increases. This strained epilayer system can lower its total energy by forming isolated thick islands in which the strain is relaxed by interfacial misfit dislocations, which leads to SK growth in these strained systems. The SK growth mode occurs when there is a lattice mismatch between the substrate and the epilayer, causing the epilayer to be strained, which results in the growth of dot-like self-assembled islands. Wire-like islands can grow from dot-like islands via a shape transition which helps strain relaxation.

For nanostructure fabrication, a thin epilayer is usually grown on a substrate. This two-dimensional (2-D) layer is used to fabricate lower-dimensional structures such as wires (1-D) or dots (0-D) by lithographic techniques. However, structures smaller than the limits of conventional lithography techniques can only be obtained by self-assembled growth utilizing the principles of SK or VW growth. For appropriate growth conditions, self-assembled epitaxial islands can be grown in reasonably well-controlled sizes [16.80].

Because wide-bandgap II–VI materials typically have stronger exciton–phonon interactions than III–V materials, their nanostructures are expected to be very useful in fabricating optoelectronics devices and in exploring the exciton nature in low-dimensional structure. Self-assembled semiconductor nanostructures of different system, such as CdSe/ZnSe [16.81], ZnSe/ZnS [16.82], CdTe/ZnTe [16.83], CdS/ZnSe [16.84], are thought to be advantageous for future application. MBE/MOMBE [16.81, 84], MOCVD [16.82], HWE [16.85] are the main growth techniques used

to obtain such structures. MBE is the most advanced technique for the growth of controlled epitaxial layers. With the advancement of nanoscience and nanotechnology, lower-dimensional nanostructures are being fabricated by lithographic techniques from two-dimensional epitaxial layers. Alternately, self-assembled, lower-dimensional nanostructures can be fabricated directly by self-assembly during MBE growth.

Atomic-Layer Epitaxy

ALE is a chemical vapor deposition technique [16.5] where the precise control of the system parameters (pressure and temperature) causes the reaction of adsorption of the precursors to be self-limiting and to stop with the completion of a single atomic layer. The precursors are usually metalorganic molecules. The special feature of ALE is that the layer thickness per cycle is independent of subtle variations of the growth parameters. The growth rate is only dependent on the number of growth cycles and the lattice constant of the deposited material. The conditions for thickness uniformity are fulfilled when material flux on each surface unit is sufficient for monolayer saturation. In an ALE reactor, this means freedom in designing the precursor transport and its interaction with the substrates.

The advantages obtainable with ALE depend on the material to be processed and the type of application. In single-crystal epitaxy, ALE may be a way to obtain a lower epitaxial crystal-growth temperature. It is also a method for making precise interfaces and material layers needed in superlattice structures and superalloys. In thin-film applications, ALE allows excellent thickness uniformity over large areas. The process has primarily been developed for processing of compound materials. ALE is not only used to grow conventional thin films of II–VI wide-bandgap compounds [16.5, 86, 87], but is also a powerful method for the preparation of monolayers (MLs) [16.88].

16.3 Bulk Crystal Growth

Bulk crystal is the most important subject studied in recent decades. The quality of bulk crystals is the most important aspect of electronic device design. To date, many growth methods have been developed to grow high-quality crystals. Significant improvements have been made in bulk crystal growth with regard to uniformity, reproducibility, thermal stability, diameter control, and impurity and dopant control. According to the phase balance, crystals can be grown from vapor phase, liquid (melt) phase, and solid phase.

16.3.1 The CVT and PVT Techniques

Crystal growth from the the vapor phase is the most basic method. It has advantages that growth can be performed at lower temperatures. This can prevent from phase transition and undesirable contamination. Therefore, this method has commonly been used to grow II–VI compound semiconductors. Crystal growth techniques from the vapor phase can be divided into chemical vapor transport (CVT) and physical vapor transport (PVT). CVT is based on chemical transport reactions that occur in a closed ampoule having two different temperature zones. Figure 16.6 shows a typical schematic diagram of the CVT technique. In the high-temperature region, the source AB reacts with the transport agent X

$$2\mathrm{AB} + 2\mathrm{X_2} \rightleftharpoons 2\mathrm{AX_2(gas)} + \mathrm{B_2(gas)}\ . \qquad (16.1)$$

In the low-temperature region, the reverse reaction takes place. The whole process continues by back-diffusion of the X_2 generated in the lower-temperature region. The transport agent X usually employed is hydrogen (H_2), a halogen (I_2, Br_2, Cl_2), a halide (HCl, HBr), and so on. For example, I_2 has been used as a transport agent for ZnS, ZnSe, ZnTe and CdS [16.89]; HCl, H_2, Cl_2, NH_3 [16.90], and C and CH_4 [16.91] have been used as the transport agents for ZnO. According to [16.89]: the typical growth temperature for ZnS is 1073−1173 K, for ZnSe 1023−1073 K, for ZnTe

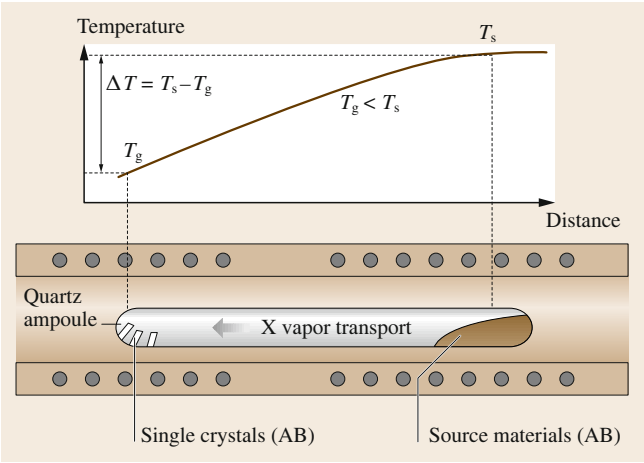


Fig. 16.6 Diagram of a conventional chemical vapor transport system

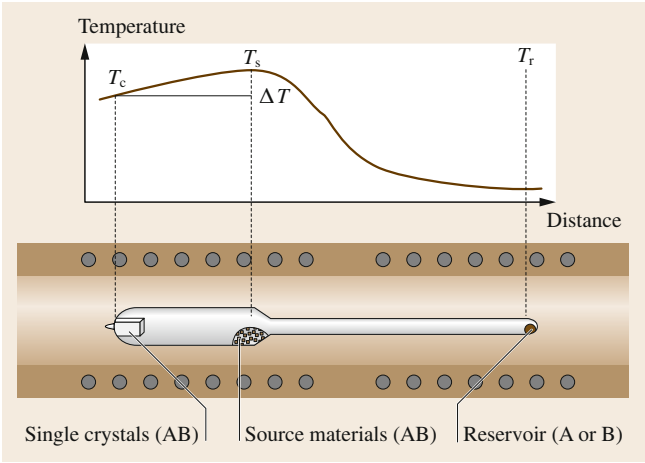


Fig. 16.7 Diagram of the physical vapor transport Piper–Polich method

973−1073 K; ΔT is 5−50 K; the concentration of the transport agent is 0.5−5 mg/cm^3 of the ampoules vapor space; the aspect ratios are 5−17 at ampoule diameters of 10−20 mm. According to [16.91]: the growth temperature for ZnO is 1228−1273 K, and ΔT is 5−10 K. The transport rate does not strongly depend on the initial amount of carbon when the concentration of the transport agent is over 0.3 mg/cm^3.

The PVT method is similar to CVT, but the transport agent is not used. This technique is based on the dissociative sublimation of compounds. Initially, the Piper–Polich method was developed, in 1961 [16.92]. *Prior* [16.93] improved the PVT method using a reservoir to control the deviation from stoichiometry; the experimental arrangement is shown in Fig. 16.7. The constituent element is placed in the reservoir. The reservoir temperature can be calculated according to the solid–vapor equilibrium

$$2\mathrm{AB} \rightleftharpoons 2\mathrm{A(gas)} + \mathrm{B_2(gas)}\,. \tag{16.2}$$

The total pressure (p) in ampoule is given by

$$\begin{aligned} p &= p_\mathrm{A} + p_{\mathrm{B}_2} \\ &= p_\mathrm{A} + K p_\mathrm{A}^{-2} \\ &= \left(\frac{K}{p_{\mathrm{B}_2}}\right)^{1/2} + p_{\mathrm{B}_2}\,, \end{aligned} \tag{16.3}$$

where p_A and p_{B_2} are the partial pressures of the group II and VI elements respectively and $K = p_\mathrm{A}^{-2} p_{\mathrm{B}_2}$ is the equilibrium constant of (16.2). At any temperature, there is minimum total pressure (p_min), which corresponds to the condition,

$$p_\mathrm{A} = 2p_{\mathrm{B}_2} = 2^{1/3} K^{2/3}\,. \tag{16.4}$$

Under this condition, the vapor-phase composition is stoichiometric and growth rate is maximum [16.94]. After modifying this method to use a closed ampoule, it was applied to grow high-purity and high-quality crystals of II–VI compounds.

The PVT of II–VI compounds takes advantage of the volatility of both components of the compound semiconductor. This same volatility, coupled with typically high melting points, makes melt growth of these materials difficult. In the PVT process, an ampoule containing a polycrystalline source of the desired II–VI compound is heated to a temperature that causes the compound to sublime at a rate conducive to crystal growth. The ampoule is typically placed in a furnace having a temperature gradient over the length of the ampoule, so that the polycrystalline source materials sublime at the end with the higher temperature. The end of the ampoule where the crystal is to be grown is then maintained at a lower temperature. This temperature difference causes supersaturation, and vaporized molecules from source materials eventually deposit at the cooler end. In order to control the deviation from stoichiometry, a reservoir is often used (Fig. 16.7). One of the constituent elements is placed in it. By selecting the proper growth conditions, the rate of deposition can be set to a value leading to growth of high-quality crystals. Typically, PVT growth of II–VI compounds is carried out at temperatures much lower than their melting points; this gives benefits in terms of reduced defects, which are related to the melt growth of II–VI compounds such as voids and/or inclusions of excess components of the compound, and also helps to reduce the contamination of the growing crystal from the ampoule. Other effects, such as the reduction of point defects, are also typically found when crystals grown by PVT are compared to crystals grown by melt techniques. Although claims have been made that the lower temperatures of physical vapor transport crystal growth should also reduce the twinning found in most of the cubic II–VI compound crystals, the reduction is not usually realized in practice. The assumption that the twinning is a result of cubic/hexagonal phase transitions is not found to be the determining factor in twin formation.

Ohno et al. [16.95] grew cubic ZnS single crystals by the iodine transport method without a seed. By means of Zn-dip treatment, this low-resistivity crystal was used for homoepitaxial MOCVD growth, and a metal–insulator–semiconductor (MIS)-structured blue LED, which yielded an external quantum efficiency as high as 0.05%. They found that crystal quality was significantly improved by prebaking the ZnS powder in H_2S gas prior to growth. The growth rate also increased by three times.

Isshiki et al. [16.96] purified zinc by a process consisting of vacuum distillation and overlap zone melting in pure argon. Using refined zinc and commercial high-purity Se, high-quality ZnSe single crystals were grown by the same method, as reported by *Huang* and *Igaki* [16.97]. The emission intensities of donor-bound exciton (I_2) are remarkably small. The emission intensities of the radiative recombinations of free excitons (E_X) are very strong [16.98]. These intensities indicated the crystal had a very high purity and a very low donor concentration, and they suggest that the purity of the grown crystal strongly depends on the purity of the starting materials. This method is suitable for preparing high-purity crystals, since a purification effect is expected during growth. Impurities with a higher vapor pressure will condense at the reservoir portion and those with a lower vapor pressure will remain in the source crystal. This effect was confirmed by the PL results [16.99]. As for these crystals, photoexcited cyclotron resonance measurements have been attempted and cyclotron resonance signals due to electrons [16.100] and heavy holes [16.101] have been detected for the first time. The cyclotron mobility of electrons under $B = 7\,\mathrm{T}$ is $2.3\times10^5\,\mathrm{cm^2/(V\,s)}$. This indicates that the quality of the grown crystals is very high. Furthermore, the donor concentration in the crystal is estimated to be $4\times10^{14}\,\mathrm{cm^3}$ by analyzing the temperature dependence of the cyclotron mobility [16.99].

The crystals are grown in a self-seeded approach by the CVT or PVT techniques introduced above. This limits single-crystal volume to several $\mathrm{cm^3}$. Meanwhile, grain boundaries and twins are easy to form during growth. In order to solve these problems, seeded chemical vapor transport (SCVT) and seeded physical vapor transport (SPVT), the so-called modified Lely method, have been developed [16.102]. The difference between SCVT/SPVT and CVT/PVT is that a seed is set in the crystal growth space before growth starts. The most successful method of eliminating twin formation has usually been by using a polycrystal or single-crystal seed. Even this seeding cannot assure complete elimination of twinning unless seeding is done carefully. The usual method of using small seeds and increasing the diameter of the growing crystal are dependent on the preparation and condition of the walls of the ampoule and the furnace profiles required to eliminate spurious nucleation from the walls. Since the use of a seed crystal provides better control over the nucleation process, high-quality single crystals can be grown [16.103, 104]. Using this technique, sizable single crystals of II–VI wide-bandgap compounds has been commercialized.

Fujita et al. [16.105] grew ZnS single crystals as large as $24\,\mathrm{mm}\times14\,\mathrm{mm}\times14\,\mathrm{mm}$ by the SCVT method using iodine as a transport agent. The average linear growth rate was about $10^{-6}\,\mathrm{cm/s}$. The crystal size depended strongly on the ampoule geometry and the temperature difference between the seed and solvent. The study of the electrical properties showed that the annealed crystal was n-type.

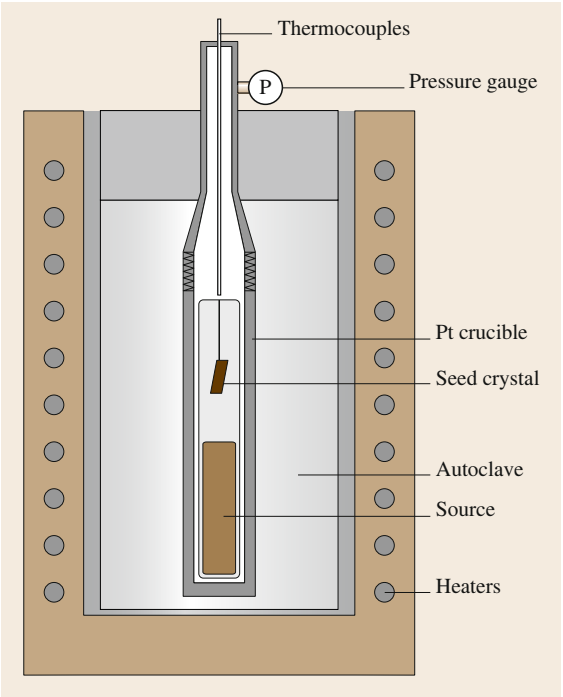


Fig. 16.8 Diagram of a typical hydrothermal technique for growing ZnO single crystals

16.3.2 Hydrothermal Growth

The hydrothermal technique is a method for growing crystal from aqueous solvent [16.106]. Figure 16.8 shows a diagram of hydrothermal techniques. The hydrothermal method of crystal growth has several advantages:

1. Due to the use of a closed system, it is easier to control oxidization or maintain conditions that allow the synthesis of phases that are difficult to attain by other methods, such as compounds of elements in oxidation states, especially for transition-metal compounds.
2. Crystal grow occurs under lower thermal strain, and thus may contain a lower dislocation density than when the crystal is grown from a melt, where large thermal gradients exist.
3. The method has proven to be very useful for the synthesis of the so-called low-temperature phases.

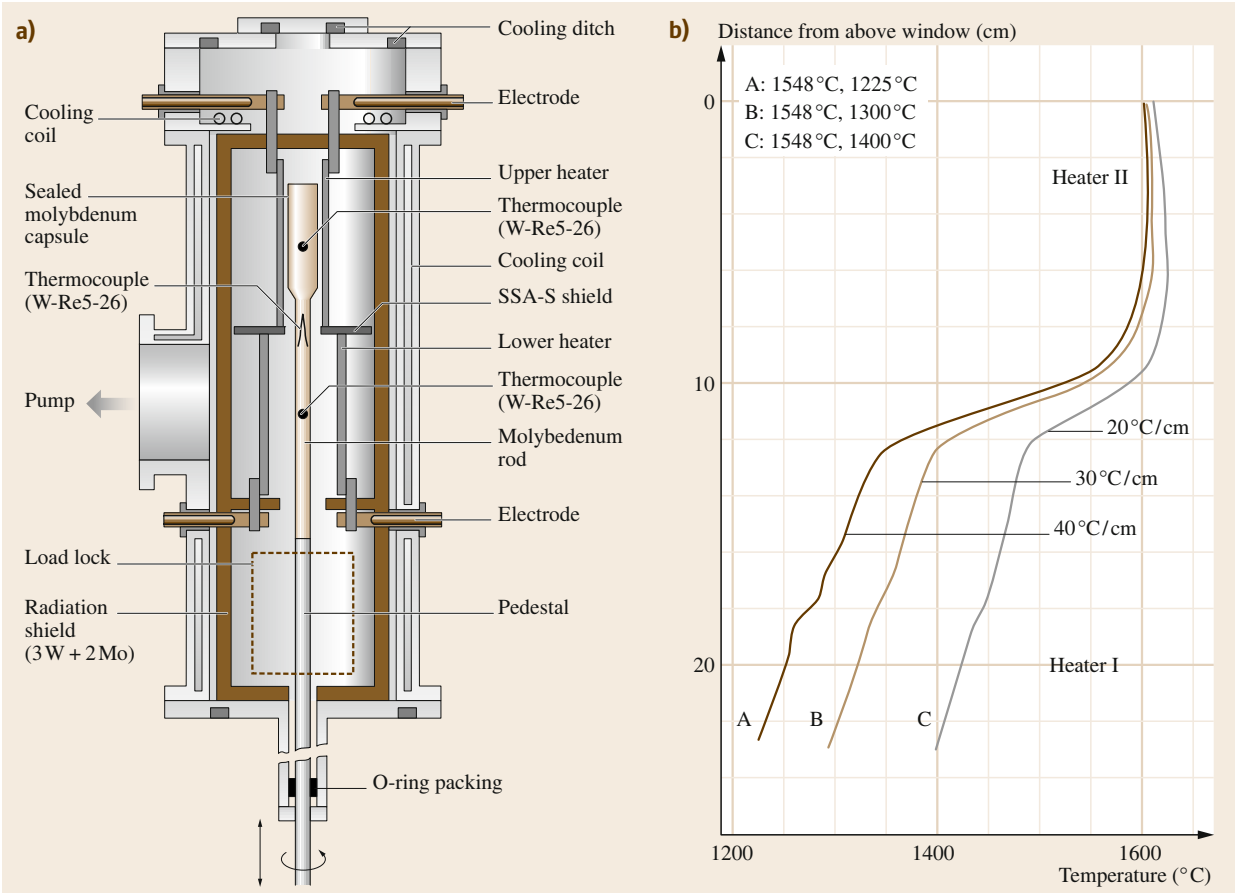


Fig. 16.9 (a) Scheme of the vertical Bridgman growth system, and (b) its temperature profiles

4. It can be employed for large-scale synthesis of piezoelectric, magnetic, optic, ceramic, and many other special materials.
5. Hydrothermal synthesis results in rapid convection and very efficient solute transfer, which results in comparatively rapid growth of larger, purer, and dislocation-free crystals.

The most successful example of obtaining II–VI compounds is growing single ZnO crystals.

ZnO crystals are considered extrinsic n-type piezoelectric semiconductors. Undoped crystals have a typical resistivity of 0.1−100 Ω cm and a drift mobility of 10−125 $cm^2/(V\,s)$. Low carrier concentrations can be approached by special growth and annealing methods. ZnO crystal is quite transparent in the range 0.4−6.0 μm. Slight absorption is sometimes found around 2.2−2.3 μm and an additional slight absorption is found at 3.42 μm. Since 1953, *Walker* [16.107] and many other researchers [16.108–111] have grown large ZnO single crystals with hydrothermal techniques and other methods.

The seed, suspended by a Pt wire, and sintered ZnO strings as a source material (nutrient), together with a KOH (3M) and LiOH (1M) aqueous solution, were put into a Pt crucible [16.112] (the hydrothermal conditions are different in different papers). The seed crystals and the source material were separated by a Pt baffle. The crucible was sealed by welding and put into an autoclave. This hydrothermal autoclave is made of high-strength steel. Then, the autoclave was put into a vertical furnace. The temperature of the autoclave was raised to about 673 K, which produced 0.1 GPa of pressure. the growth temperature was monitored by a thermocouple inserted in the autoclave. Seed crystals grew to about 10 mm or bigger after two weeks. The crystal habit of hydrothermal ZnO crystals grown on basal plane seeds shows that growth direction in [0001] is faster 3 times than $[000\bar{1}]$ [16.113].

16.3.3 Bridgman and Gradient Freezing (GF) Method

From the viewpoint of industrial production, melt growth is the most useful for obtaining large single crystals. VPE growth has limitations with regard to crystal size and productivity. The Bridgman technique is a typical crystal growth method from melt. Bridgman growth can be simply understood in terms of a molten charge that passes through a temperature gradient at a slow speed and solidifies when the temperature is below the melting point of this material. If the ampoule and furnace are stationary and the temperature is gradually reduced by keeping the temperature gradient at the interface constant, this growth process is called the gradient freezing (GF) method [16.114]. In Bridgman or GF growth, single crystals can be grown using either seeded or unseeded ampoules or crucibles.

The Bridgman method has been most extensively used to grow II–VI wide-bandgap compounds such as CdTe, ZnTe and ZnSe crystals, because of the simplicity of the growth apparatus, the high growth rate and the availability of crystals of appropriate size and quality. There are two Bridgman method techniques: the high-pressure technique [16.115] and the closed technique [16.116]. In the former, it is inevitable that a compositional deviation from stoichiometry occurs during melting. Since the properties and structural perfection of these compound crystals are correlated very strongly with this nonstoichiometry [16.117], compositional deviation must be controlled during melt growth. *Omino* et al. [16.116] and *Wang* et al. [16.118, 119] have adopted a closed double crucible to prevent deviation of the melt from stoichiometric composition during Bridgman growth. Figure 16.9 shows a closed vertical Bridgman growth furnace and its temperature profiles.

In Bridgman growth, the temperature gradient (G) and growth rate (R) are very important parameters since they determine the shape of the solid–liquid interface. For this reason, the relationship between temperature gradient and growth rate is investigated. The experimental results are summarized in Fig. 16.10 [16.120, 121]. The experimental results show that, to grow a ZnSe single crystal, it is necessary that the G/R ((temperature) gradient/(growth) rate) value should be limited to 57−175 K h/cm^2. The most suitable value of G/R, assessed from the determined optimum temperature gradient and growth rate, is 83 K h/cm^2. *Wang* et al. [16.118, 119] found the optimum conditions to include: a special temperature program for removing the gas bubbles generated in melt, an overheating temperature of 76 K from the melting point of 1797 K, a temperature gradient of 30 K/cm and a growth rate of 3.6 mm/h as marked by the open squares in Fig. 16.10. Under these growth conditions, twin-free high-quality ZnSe single crystals (Fig. 16.11) were grown using a polycrystalline seed. Chemical etching on the cleaved (110) plane revealed that the average value of the etch pit density (EPD) is about 2×10^5 cm^{-2}. The rocking curves of four-crystal XRD showed a full-width at half-maximum (FWHM) value of 19 arcs. The resolved intensive free-exciton, bound-exciton emission lines and the weak donor-acceptor-pair (DAP) emission bands are observed in the PL spectra at 4.2 K. The FWHM of the I_1^d emission was smaller than 0.5 meV. On the other hand, the deep-level emission bands were almost not observed. All these results suggest that the ZnSe single crystals grown by this method are of very high quality.

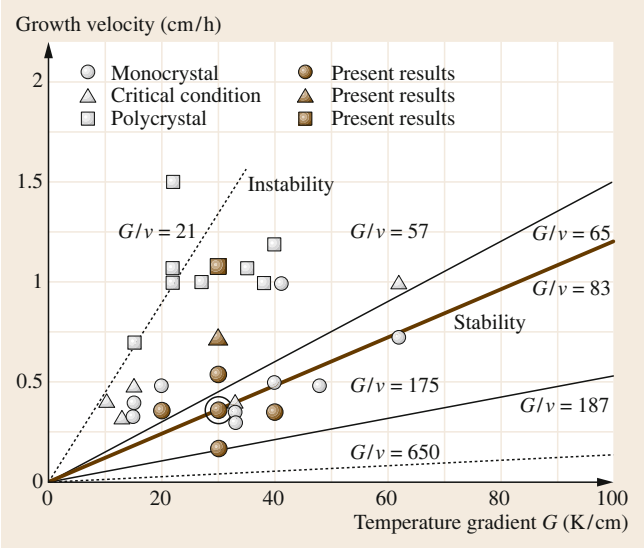


Fig. 16.10 Relationship between growth velocity and temperature gradient at the growth interfaces

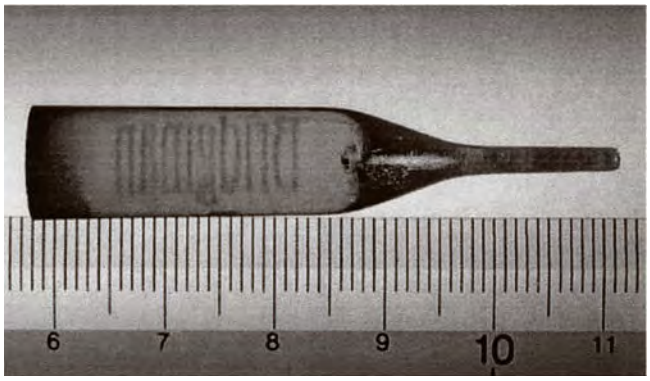


Fig. 16.11 Photograph of twin-free ZnSe single crystal grown by the vertical Bridgman technique

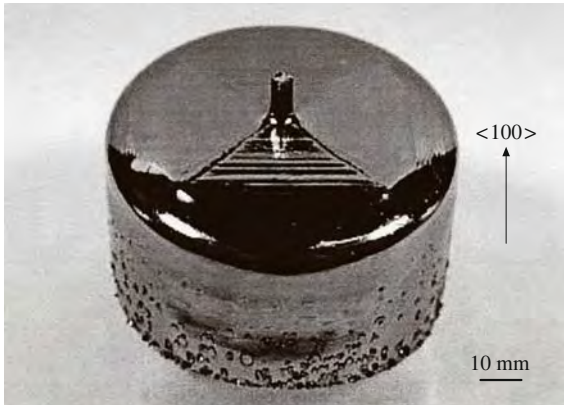


Fig. 16.12 ZnTe single crystal grown by a combination of the GF and Kyropoulos methods

Asahi et al. [16.122] successfully grew ZnTe single crystals with a diameter of 80 mm and a length of 50 mm by the vertical gradient freezing (VGF) method. In this method, a high-pressure furnace was used and the melt was encapsulated by B_2O_3 during crystal growth. The growth direction was nearly ⟨111⟩ or ⟨110⟩. When long ZnTe crystals were grown, polycrystals were found at the tail. It seems to be difficult to grow an ingot longer than 50 mm. The researchers believed that this is because the shape of the solid–liquid interface easily becomes concave against the liquid at the tail owing to the low thermal conductivity of ZnTe. Evaluation of the crystals showed that the FWHMs of the rocking curve measured by XRD were about 20 arcs, and the EPDs were 5×10^3–1×10^4 cm^{-2}.

16.3.4 The Traveling Heater Method (THM)

The traveling heater method (THM) [16.123] is a solution growth process whereby polycrystalline feed material with an average constant composition is progressively dissolved under the influence of a temperature gradient, followed by deposition in single-crystal form onto a seed with the same composition. The growth proceeds by the relative translation of the heater and charge. THM is particularly useful for the growth of binary and ternary semiconductor alloys, such as CdTe and CdZnTe [16.124]. In such materials, the wide separation between the solidus and liquidus in the pseudo-binary (CdTe–ZnTe) phase diagram imposes a monotonic variation in composition of the solid in the melt growth processes. The THM process ensures a constant macroscale composition in the crystal grown. Since the process takes place at a temperature below the melting point, contamination from the container is also reduced. The reduced operating temperature also leads to a lower ambient pressure within the growth environment, and a reduced risk of ampoule fracture.

16.3.5 Other Methods

Besides the growth methods described, there are many other techniques for growing single crystals of II–VI wide-bandgap compounds. Zone melting [16.125] and solid-state recrystallization (SSR) [16.126] are often used to grow bulk crystals of ZnSe, ZnS and CdTe.

Recently, *Asahi* et al. [16.127] proved that B_2O_3 is a suitable encapsulant for ZnTe melt growth. Furthermore, B_2O_3 and a total weight of 6N Zn and Te was charged into a pyrolitic boron nitride (pBN) crucible. Then this crucible was put into a high-pressure furnace with five heaters. A ZnTe seed was used to pull the ZnTe crystal. Before growth started, the starting materials were heated to 1573 K and kept at this temperature for several hours. The pressure in the growth furnace was kept at 1.5–2 MPa using Ar gas during growth. The temperature gradient on the solid–liquid surface was about 10–20 K/cm. The growth rate was 2–4 mm/h. Under these growth conditions, ZnTe single crystals with a diameter of 80 mm and a height of 40 mm were successfully grown using a combined GF/Kyropoulos method [16.127] (Fig. 16.12).

Some growth methods for film and bulk crystals are summarized in Table 16.3.

Table 16.3 Growth methods for films and bulk crystals of wide-bandgap II–VI compounds

	Epitaxial growth						Bulk growth			
	LPE	VPE	HWE	ALE	MOCVD	MBE	CVT	PVT	Hydrothermal	Bridgman
ZnS	✓	✓		✓	✓	✓	✓	✓		
ZnO					✓	✓	✓		✓	
ZnSe	✓	✓	✓	✓	✓	✓	✓	✓		✓
ZnTe	✓	✓	✓	✓	✓	✓	✓	✓		✓
CdS		✓	✓	✓	✓	✓	✓	✓		✓
CdSe		✓	✓	✓	✓	✓	✓	✓		✓
CdTe		✓	✓	✓	✓	✓	✓	✓		✓

16.4 Conclusions

It was expected that II–VI wide-bandgap compounds would become applicable to optoelectronic devices, especially LEDs and LDs in the short-wavelength visible-light region. However, as they are very strongly bonded with a nearly equal balance of covalent and ionic bonding, it is challenging to grow high-purity high-quality single crystals. Another problem with their application to devices is the difficulty of controlling the conductive type. This is because native defects commonly occur in these semiconductors. These native defects can have either a donor or acceptor character, or even be amphoteric, and they act as compensating centers. Furthermore, these defects react with dopant impurities to form complexes. This makes it difficult to reverse the conductive type. For example, in the case of ZnSe, it is still difficult to obtain low-resistivity p-type crystals or epilayers that can be used to fabricate devices. Therefore, a significant improvement in the understanding of the fundamental physical and chemical properties needs to be achieved. In particular, the specifications for many applications are very demanding, and considerable progress needs to be made in growth, particularly in the areas of reproducibility, convenient shape, conductivity, and structural perfection.

References

16.1 A. Lopez-Otero: Thin Solid Films **49**, 1 (1978)
16.2 H.M. Manasevit, W.I. Simpson: J. Electrochem. Soc. **118**, 644 (1971)
16.3 L.L. Chang, R. Ludeke: In: *Epitaxial Growth, Part A*, ed. by J.W. Matthews (Academic, New York 1975) p. 37
16.4 E. Veuhoff, W. Pletschen, P. Balk, H. Luth: J. Cryst. Growth **55**, 30 (1981)
16.5 T. Suntola: Mater. Sci. Rep. **4**, 261 (1989)
16.6 M.M. Faktor, R. Heckingbottom, I. Garrett: J. Cryst. Growth **9**, 3 (1971)
16.7 I. Kikuma, M. Furukoshi: J. Cryst. Growth **41**, 103 (1977)
16.8 Y.V. Korostelin, V.J. Kozlovskij, A.S. Nasibov, P.V. Shapkin: J. Cryst. Growth **159**, 181 (1996)
16.9 J.F. Wang, A. Omino, M. Isshiki: Mater. Sci. Eng. **83**, 185 (2001)
16.10 S.H. Song, J.F. Wang, G.M. Lalev, L. He, M. Isshiki: J. Cryst. Growth **252**, 102 (2003)
16.11 H. Harmann, R. Mach, B. Sell: In: *Current Topics Materials Science*, Vol. 9, ed. by E. Kaldis (North-Holland, Amsterdam 1982) p. 1
16.12 P. Rudolph, N. Schäfer, T. Fukuda: Mater. Sci. Eng. **15**, 85 (1995)
16.13 R. Shetty, R. Balasubramanian, W.R. Wilcox: J. Cryst. Growth **100**, 51 (1990)
16.14 K.W. Böer: *Survey of Semiconductor Physics, Vol. 1: Electrons and Other Particles in Bulk Semiconductors* (Van Nostrand, New York 1990)
16.15 C.M. Wolf, N. Holonyak, G.E. Stillman: *Physical Properties of Semiconductors* (Prentice Hall, New York 1989)
16.16 L. Smart, E. Moore: *Solid State Chemistry*, 2nd edn. (Chapman Hall, New York 1995)
16.17 E. Lide (Ed.): *Handbook of Chemistry and Physics*, 2nd edn. (CRC, Boca Raton 1973)
16.18 J. Singh: *Physics of Semiconductors and Their Heterostructures* (McGraw–Hill, New York 1993)
16.19 N. Yamamoto, H. Horinaka, T. Miyauchi: Jpn. J. Appl. Phys. **18**, 225 (1997)
16.20 H. Neumann: Kristall Technik **15**, 849 (1980)
16.21 J. Camassel, D. Auvergne, H. Mathieu: J. Phys. Colloq. **35**, C3–67 (1974)
16.22 W. Shan, J.J. Song, H. Luo, J.K. Furdyna: Phys. Rev. **50**, 8012 (1994)
16.23 K.A. Dmitrenko, S.G. Shevel, L.V. Taranenko, A.V. Marintchenko: Phys. Status Solidi B **134**, 605 (1986)
16.24 S. Logothetidis, M. Cardona, P. Lautenschlager, M. Garriga: Phys. Rev. B **34**, 2458 (1986)
16.25 R.C. Sharma, Y.A. Chang: J. Cryst. Growth **88**, 192 (1988)
16.26 H. Okada, T. Kawanaka, S. Ohmoto: J. Cryst. Growth **165**, 31 (1996)
16.27 N.K. Abrikosov, V.F. Bankina, L.B. Poretzkaya, E.V. Skudnova, S.N. Chichevskaya: *Poluprovodnikovye chalkogenidy i splavy na ikh osnovje* (Nauka, Moscow 1975) (in Russian)
16.28 R.F. Brebrick: J. Cryst. Growth **86**, 39 (1988)
16.29 M.R. Lorenz: In: *Physics and Chemistry of II–VI Compounds*, ed. by M. Aven, J.S. Prener (North Holland, Amsterdam 1967) p. 210
16.30 T. Yao: Optoelectron. Dev. Technol. **6**, 37 (1991)
16.31 H. Nakamura, M. Aoki: Jpn. J. Appl. Phys. **20**, 11 (1981)
16.32 C. Werkhoven, B.J. Fitzpatrik, S.P. Herko, R.N. Bhargave, P.J. Dean: Appl. Phys. Lett. **38**, 540 (1981)
16.33 H. Nakamura, S. Kojima, M. Wasgiyama, M. Aoki: Jpn. J. Appl. Phys. **23**, L617 (1984)
16.34 V.M. Skobeeva, V.V. Serdyuk, L.N. Semenyuk, N.V. Malishin: J. Appl. Spectrosc. **44**, 164 (1986)
16.35 P. Lilley, P.L. Jones, C.N.W. Litting: J. Mater. Sci. **5**, 891 (1970)
16.36 T. Matsumoto, T. Morita, T. Ishida: J. Cryst. Growth **53**, 225 (1987)
16.37 S. Zhang, H. Kinto, T. Yatabe, S. Iida: J. Cryst. Growth **86**, 372 (1988)
16.38 S. Iida, T. Yatabe, H. Kinto: Jpn. J. Appl. Phys. **28**, L535 (1989)
16.39 P. Besomi, B.W. Wessels: J. Cryst. Growth **55**, 477 (1981)

16.40 T. Kyotani, M. Isshiki, K. Masumoto: J. Electrochem. Soc. **136**, 2376 (1989)
16.41 N. Stucheli, E. Bucher: J. Electron. Mater. **18**, 105 (1989)
16.42 M. Nishio, Y. Nakamura, H. Ogawa: Jpn. J. Appl. Phys. **22**, 1101 (1983)
16.43 N. Lovergine, R. Cingolani, A.M. Mancini, M. Ferrara: J. Cryst. Growth **118**, 304 (1992)
16.44 O. De Melo, E. Sánchez, S. De Roux, F. Rábago-Bernal: Mater. Chem. Phys. **59**, 120 (1999)
16.45 M. Kasuga, H. Futami, Y. Iba: J. Cryst. Growth **115**, 711 (1991)
16.46 J.F. Wang, K. Kikuchi, B.H. Koo, Y. Ishikawa, W. Uchida, M. Isshiki: J. Cryst. Growth **187**, 373 (1998)
16.47 J. Humenberger, G. Linnet, K. Lischka: Thin Solid Films **121**, 75 (1984)
16.48 F. Sasaki, T. Mishina, Y. Masumoto: J. Cryst. Growth **117**, 768 (1992)
16.49 B.J. Kim, J.F. Wang, Y. Ishikawa, S. Sato, M. Isshiki: Phys. Stat. Sol. (a) **191**, 161 (2002)
16.50 A. Rogalski, J. Piotrowski: Prog. Quantum Electron. **12**, 87 (1988)
16.51 G.M. Lalev, J. Wang, S. Abe, K. Masumoto, M. Isshiki: J. Crystal Growth **256**, 20 (2003)
16.52 H.M. Manasevit: Appl. Phys. Lett. **12**, 1530 (1968)
16.53 S. Fujita, M. Isemura, T. Sakamoto, N. Yoshimura: J. Cryst. Growth **86**, 263 (1988)
16.54 H. Mitsuhashi, I. Mitsuishi, H. Kukimoto: J. Cryst. Growth **77**, 219 (1986)
16.55 P.J. Wright, P.J. Parbrook, B. Cockayne, A.C. Jones, E.D. Orrell, K.P. O'Donnell, B. Henderson: J. Cryst. Growth **94**, 441 (1989)
16.56 S. Hirata, M. Isemura, S. Fujita, S. Fujita: J. Cryst. Growth **104**, 521 (1990)
16.57 S. Nishimura, N. Iwasa, M. Senoh, T. Mukai: Jpn. J. Appl. Phys. **32**, L425 (1993)
16.58 K.P. Giapis, K.F. Jensen, J.E. Potts, S.J. Pachuta: Appl. Phys. Lett. **55**, 463 (1989)
16.59 S.J. Pachuta, K.F. Jensen, S.P. Giapis: J. Cryst. Growth **107**, 390 (1991)
16.60 M. Danek, J.S. Huh, L. Foley, K.F. Jenson: J. Cryst. Growth **145**, 530 (1994)
16.61 W. Kuhn, A. Naumov, H. Stanzl, S. Bauer, K. Wolf, H.P. Wagner, W. Gebhardt, U.W. Pohl, A. Krost, W. Richter, U. Dümichen, K.H. Thiele: J. Cryst. Growth **123**, 605 (1992)
16.62 J.K. Menno, J.W. Kerri, F.H. Robert: J. Phys. Chem. B **101**, 4882 (1997)
16.63 H.P. Wagner, W. Kuhn, W. Gebhardt: J. Cryst. Growth **101**, 199 (1990)
16.64 Taskar, B.A. Khan, D.R. Dorman, K. Shahzad: Appl. Phys. Lett. **62**, 270 (1993)
16.65 Y. Fujita, T. Terada, T. Suzuki: Jpn. J. Appl. Phys. **34**, L1034 (1995)
16.66 J. Wang, T. Miki, A. Omino, K.S. Park, M. Isshiki: J. Cryst. Growth **221**, 393 (2000)
16.67 M.K. Lee, M.Y. Yeh, S.J. Guo, H.D. Huang: J. Appl. Phys. **75**, 7821 (1994)
16.68 A. Toda, T. Margalith, D. Imanishi, K. Yanashima, A. Ishibashi: Electron. Lett. **31**, 1921 (1995)
16.69 A. Cho: J. Vac. Sci. Tech. **8**, S31 (1971)
16.70 C.T. Foxon: J. Cryst. Growth **251**, 130 (2003)
16.71 T. Yao: In: *The Technology and Physics of Molecular Beam Epitaxy*, ed. by E.H.C. Parker (Plenum, New York 1985) p. 313, Chap. 10
16.72 E. Veuhoff, W. Pletschen, P. Balk, H. Luth: J. Cryst. Growth **55**, 30 (1981)
16.73 M.B. Panish, S. Sumski: J. Appl. Phys. **55**, 3571 (1984)
16.74 Y.P. Chen, G. Brill, N.K. Dhar: J. Cryst. Growth **252**, 270 (2003)
16.75 H. Kato, M. Sano, K. Miyamoto, T. Yao: J. Cryst. Growth **237–239**, 538 (2002)
16.76 M. Imaizumi, M. Adachi, Y. Fujii, Y. Hayashi, T. Soga, T. Jimbo, M. Umeno: J. Cryst. Growth **221**, 688 (2000)
16.77 W. Xie, D.C. Grillo, M. Kobayashi, R.L. Gunshor, G.C. Hua, N. Otsuka, H. Jeon, J. Ding, A.V. Nurmikko: Appl. Phys. Lett. **60**, 463 (1992)
16.78 S. Guha, A. Madhukar, K.C. Rajkumar: Appl. Phys. Lett. **57**, 2110 (1990)
16.79 E. Bauer, J.H. van der Merwe: Phys. Rev. B **33**, 3657 (1986)
16.80 J. Drucker, S. Chapparro: Appl. Phys. Lett. **71**, 614 (1997)
16.81 S.H. Xin, P.D. Wang, A. Yin, C. Kim, M. Dobrowolska, J.L. Merz, J.K. Furdyna: Appl. Phys. Lett. **69**, 3884 (1996)
16.82 M.C. Harris Liao, Y.H. Chang, Y.H. Chen, J.W. Hsu, J.M. Lin, W.C. Chou: Appl. Phys. Lett. **70**, 2256 (1997)
16.83 Y. Terai, S. Kuroda, K. Takita, T. Okuno, Y. Masumoto: Appl. Phys. Lett. **73**, 3757 (1998)
16.84 M. Kobayashi, S. Nakamura, K. Wakao, A. Yoshikawa, K. Takahashi: J. Vac. Sci. Technol. B **16**, 1316 (1998)
16.85 S.O. Ferreira, E.C. Paiva, G.N. Fontes, B.R.A. Neves: J. Appl. Phys. **93**, 1195 (2003)
16.86 M.A. Herman, J.T. Sadowski: Cryst. Res. Technol. **34**, 153 (1999)
16.87 M. Ahonen, M. Pessa, T. Suntola: Thin Solid Films **65**, 301 (1980)
16.88 M. Ritala, M. Leskelä: Nanotechnology **10**, 19 (1999)
16.89 H. Hartmann: J. Cryst. Growth **42**, 144 (1977)
16.90 M. Shiloh, J. Gutman: J. Cryst. Growth **11**, 105 (1971)
16.91 S. Hassani, A. Tromson-Carli, A. Lusson, G. Didier, R. Triboulet: Phys. Stat. Sol. (b) **229**, 835 (2002)
16.92 W.W. Piper, S.J. Polich: J. Appl. Phys. **32**, 1278 (1961)
16.93 A.C. Prior: J. Electrochem. Soc. **108**, 106 (1961)
16.94 T. Kiyosawa, K. Igaki, N. Ohashi: Trans. Jpn. Inst. Metala **13**, 248 (1972)
16.95 T. Ohno, K. Kurisu, T. Taguchi: J. Cryst. Growth **99**, 737 (1990)
16.96 M. Isshiki, T. Tomizono, T. Yoshita, T. Ohkawa, K. Igaki: J. Jpn. Inst. Metals **48**, 1176 (1984)
16.97 X.M. Huang, K. Igaki: J. Cryst. Growth **78**, 24 (1986)
16.98 M. Isshiki, T. Yoshita, K. Igaki, W. Uchida, S. Suto: J. Cryst. Growth **72**, 162 (1985)
16.99 M. Isshiki: J. Cryst. Growth **86**, 615 (1988)
16.100 T. Ohyama, E. Otsuka, T. Yoshita, M. Isshiki, K. Igaki: Jpn. J Appl. Phys. **23**, L382 (1984)

16.101 T. Ohyama, K. Sakakibara, E. Otsuka, M. Isshiki, K. Igaki: Phys. Rev. B **37**, 6153 (1988)
16.102 Y.M. Tairov, V.F. Tsvetkov: J. Cryst. Growth **43**, 209 (1978)
16.103 G. Cantwell, W.C. Harsch, H.L. Cotal, B.G. Markey, S.W.S. McKeever, J.E. Thomas: J. Appl. Phys. **71**, 2931 (1992)
16.104 V. Yu, Korostelin, V.I. Kozlovsky, A.S. Nasibov, P.V. Shapkin: J. Cryst. Growth **161**, 51 (1996)
16.105 S. Fujita, H. Mimoto, H. Takebe, T. Noguchi: J. Cryst. Growth **47**, 326 (1979)
16.106 K. Byrappa: In: *Hydrothermal Growth of Crystal*, ed. by K. Byrappa (Pergamon, Oxford 1991)
16.107 A.C. Walker: J. Am. Ceram. Soc. **36**, 250 (1953)
16.108 R.A. Laudice, E.D. Kolg, A.J. Caporaso: J. Am. Ceram. Soc. **47**, 9 (1964)
16.109 M. Suscavage, M. Harris, D. Bliss, P. Yip, S.-Q. Wang, D. Schwall, L. Bouthillette, J. Bailey, M. Callahan, D.C. Look, D.C. Reynolds, R.L. Jones, C.W. Litton: MRS Internet J. Nitride Semicond. Res. **4S1**, G3.40 (1999)
16.110 L.N. Demianets, D.V. Kostomarov: Ann. Chim. Sci. Mater. **26**, 193 (2001)
16.111 N. Ohashi, T. Ohgaki, T. Nakata, T. Tsurumi, T. Sekiguchi, H. Haneda, J. Tanaka: J. Kor. Phys. Soc. **35**, S287 (1999)
16.112 D.C. Look, D.C. Reynolds, J.R. Sizelove, R.L. Jones, C.W. Litton, G. Gantwell, W.C. Harsch: Solid State Commun. **105**, 399 (1988)
16.113 T. Sekiguchi, S. Miyashita, K. Obara, T. Shishido, N. Sakagami: J. Cryst. Growth **214/215**, 72 (2000)
16.114 P. Höschl, M. Yu, Ivanov, E. Belas, J. Franc, R. Grill, D. Hlidek, P. Moravec, M. Zvara, H. Sitter, A. Toth: J. Cryst. Growth **184/185**, 1039 (1998)
16.115 T. Fukuda, K. Umetsu, P. Rudolph, H.J. Koh, S. Iida, H. Uchiki, N. Tsuboi: J. Cryst. Growth **161**, 45 (1996)
16.116 A. Omino, T. Suzuki: J. Cryst. Growth **117**, 80 (1992)
16.117 I. Kikuma, M. Furukoshi: J. Cryst. Growth **71**, 136 (1985)
16.118 J.F. Wang, A. Omino, M. Isshiki: J. Cryst. Growth **214/215**, 875 (2000)
16.119 J. Wang, A. Omino, M. Isshiki: J. Cryst. Growth **229**, 69 (2001)
16.120 J.F. Wang, A. Omino, M. Isshiki: Mater. Sci. Eng. B **83**, 185 (2001)
16.121 P. Rudolph, N. Schäfer, T. Fukuda: Mater. Sci. Eng. R **15**, 85 (1995)
16.122 T. Asahi, A. Arakawa, K. Sato: J. Cryst. Growth **229**, 74 (2001)
16.123 M. Ohmori, Y. Iwase, R. Ohno: Mater. Sci. Eng. B **16**, 283 (1999)
16.124 R. Triboulet: Prog. Cryst. Growth Char. Mater. **128**, 85 (1994)
16.125 H.H. Woodbury, R.S. Lewandowski: J. Cryst. Growth **10**, 6 (1971)
16.126 R. Triboulet: Cryst. Res. Technol. **38**, 215 (2003)
16.127 T. Asahi, T. Yabe, K. Sato: The Japan Society of Applied Physics and Related Societies, Extended Abstracts, The 50th Spring Meeting, (2003) p. 332

17. Structural Characterization

Paul D. Brown

The aim of this chapter is to convey the basic principles of x-ray and electron diffraction, as used in the structural characterization of semiconductor heterostructures. A number of key concepts associated with radiation–material and particle–material interactions are introduced, with emphasis placed on the nature of the signal used for sample interrogation. Various modes of imaging and electron diffraction are then described, followed by a brief appraisal of the main techniques used to prepare electron-transparent membranes for TEM analysis. A number of case studies on electronic and photonic material systems are then presented in the context of a growth or device development program; these emphasize the need to use complementary techniques when characterizing a given heterostructure.

17.1 **Radiation–Material Interactions** 386
17.2 **Particle–Material Interactions** 387
17.3 **X-ray Diffraction** 390
17.4 **Optics, Imaging and Electron Diffraction** 393
17.4.1 Electron Diffraction and Image Contrast Analysis 397
17.4.2 Microdiffraction and Polarity 399
17.4.3 Reflection High-Energy Electron Diffraction 401
17.5 **Characterizing Functional Activity** 403
17.6 **Sample Preparation** 404
17.7 **Case Studies – Complementary Characterization of Electronic and Optoelectronic Materials** 406
17.7.1 Identifying Defect Sources Within Homoepitaxial GaN 407
17.7.2 Cathodoluminescence/Correlated TEM Investigation of Epitaxial GaN 408
17.7.3 Scanning Transmission Electron Beam Induced Conductivity of $Si/Si_{1-x}Ge_x/Si(001)$ 409
17.8 **Concluding Remarks** 411
References 412

The functional properties of semiconductors emanate from their atomic structures; and, the interrelationship between materials processing, microstructure and functional property lies at the heart of semiconductor science and technology. Hence, if we are to elucidate how the functional properties of a semiconductor depend on processing history (the growth or device fabrication procedures used), then we must study the development of the microstructure of the semiconductor by applying an appropriate combination of analytical techniques to the given bulk crystal, heterostructure or integrated device structure.

The main aim of this chapter is to provide a general introduction to the techniques used to characterize the structures of semiconductors. Thus, we consider techniques such as x-ray diffraction (XRD) and electron diffraction, combined with diffraction contrast imaging, alongside related techniques used for chemical microanalysis, since modern instruments such as analytical electron microscopes (AEMs) provide a variety of operational modes that allow both structure and chemistry to be investigated, in addition to functional activity. For example, chemical microanalyses of the fine-scale structures of materials can be performed within a scanning electron microscope (SEM) and/or a transmission electron microscope (TEM), using the techniques of energy dispersive x-ray (EDX) analysis, wavelength dispersive x-ray (WDX) analysis or electron energy loss spectrometry (EELS). In addition, the electrical and optical properties of semiconductors can be investigated in situ using the techniques of electron beam induced conductivity (EBIC) or cathodo-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_17

luminescence (CL), respectively. Techniques such as x-ray photoelectron spectrometry (XPS; also known as electron spectroscopy for chemical analysis, ESCA), secondary ion mass spectrometry (SIMS) or Rutherford backscattering spectrometry (RBS), can also be used to study semiconductor chemistry. We should also mention reflection high-energy electron diffraction (RHEED), which can be used for the rapid structural assessment of the near surface of bulk or thin film samples.

A far from exhaustive list of acronyms one might come across when assessing a given sample is incorporated into the general list of abbreviations at the start of this book. We can organize these techniques used to characterize materials into groups based on a number of viewpoints: for example, with respect to the material property being investigated; whether they are destructive or nondestructive; bulk or near-surface assessment techniques; based on radiation-material or particle-material interactions, or based on elastic or inelastic scattering processes, or whether they are diffraction-, imaging- or spectroscopic-based techniques.

In this broad introduction to the structural characterization of semiconductors, we focus on the interaction of a material with radiation and/or particles. Thus, we start by considering the interactions of photons, electrons or ions with a sample and the nature of the signals used for structural or chemical microanalysis. We then briefly focus on the techniques of x-ray and electron diffraction, and issues regarding the formation of TEM images. The aim is simply to convey an appreciation of the underlying principles, the applicability of these characterization techniques and the information content provided by them. We also briefly consider aspects of sample preparation, and the chapter closes with a variety of TEM-based case studies of semiconductor heterostructures, which are included to illustrate some of the approaches used to characterize fine-scale microstructure, as well as to emphasise the need for complementary analysis when assessing the interrelationships between processing, structure and functional property. Much literature already exists in this area, and a selection of references is provided at the end of the chapter that tackle many of the topics we cover here in more detail [17.1–22].

17.1 Radiation–Material Interactions

Each part of the electromagnetic spectrum has quantum energies that can be used to elicit certain forms of excitation at the atomic or molecular level. Different parts of the electromagnetic spectrum will interact with matter in different ways, according to the energy states within the material, allowing absorption or ionization effects to occur. The salient features of these various radiation–material interaction processes are summarized in the schematics shown in Fig. 17.1.

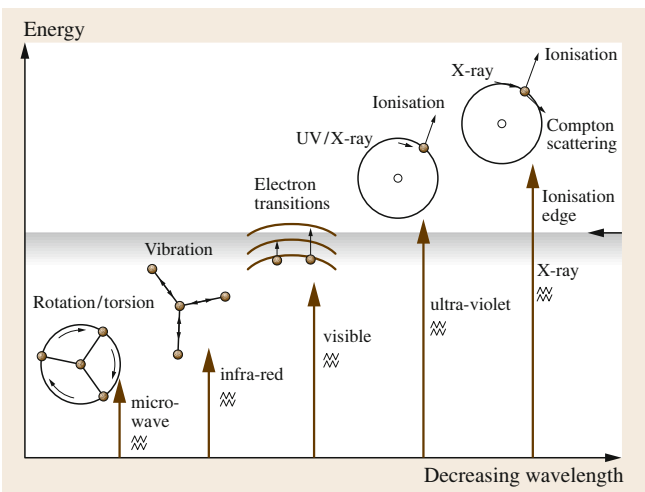


Fig. 17.1 Schematics illustrating the various interactions of energetic radiation with a molecule or atom

As the quantum energy increases from radio waves, through microwaves, to infrared and visible light, absorption increases, whilst specific quantized ionization effects come into play upon moving further into the ultraviolet and x-ray parts of the spectrum. Microwave and infrared radiation, for example, can interact with the quantum states of molecular rotation and torsion, leading to the generation of heat. Strong absorption also occurs within metallic conductors, leading to the induction of electric currents. Visible and ultraviolet light can elevate electrons to higher energy levels in what is known as the photoelectric effect, which is essentially the liberation of electrons from matter by short-wavelength electromagnetic radiation when all of the incident radiation energy is transferred to an electron. This process can be explained in terms of the absorption of discrete photon energies, with electrons being emitted when the photon energy exceeds the material's work function for the case of weakly bound electrons or the binding energy for more strongly bound inner shell electrons (Fig. 17.2a).

X-ray and γ-ray quantum energies are generally too high to be completely absorbed in direct

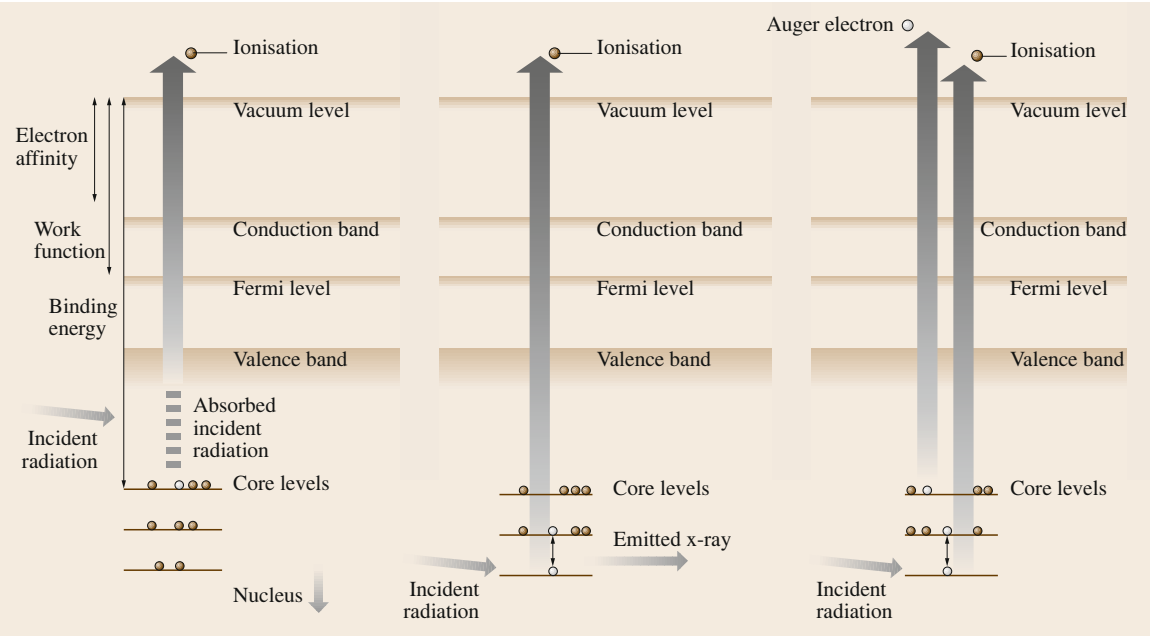


Fig. 17.2a–c Schematics illustrating **(a)** the photoelectric effect; **(b)** characteristic x-ray emission and **(c)** Auger electron emission

electron transitions, but they can induce ionization, with the displacement of electrons from atoms to form ions. The relaxation of a high-energy electron to the vacated state can lead to the emission of a characteristic x-ray (Fig. 17.2b) or the emission of an Auger electron (Fig. 17.2c). During the process of ionization, some of the incident photon energy is transferred to the ejected electron in the form of kinetic energy, and the scattering of a lower energy photon (longer wavelength x-ray) occurs, termed Compton scattering. x-rays can also be scattered elastically by shell electrons, without the loss of energy, through a process of absorption and re-emission.

17.2 Particle–Material Interactions

The interaction of an energetic particle with the surface of a material is most commonly associated with the process of sputtering – the non-thermal removal of atoms from a surface under ion bombardment. Transfer of momentum to the surface atoms is followed by a chain of collision events leading to the ejection of matrix atoms. Figure 17.3 illustrates the various signals produced when a particle interacts with a material, depending on the energy available. In broad terms, processes in the range 10^4–10^5 eV are associated with ion implantation, the 10–10^3 eV energy range is associated with sputtering, whilst the creation of activated point defects occurs in the 1–10^3 eV range. At lower 0.1–100 eV energies, desorption of surface impurity atoms occurs, whilst energies in the range 0.01–1 eV are associated with the enhanced mobility of surface condensing particles, such as those required during growth. In practical terms, the process of sputtering is most efficient when the masses of the incident and ejected particles are similar, whilst it is also dependent on the sputtering gas pressure, the energy spread of the particles, the bias conditions and the sample geometry.

In this context, we should briefly mention the technique of SIMS, which uses heavy ions, typically in the range 2–30 keV, with an ion current density of $\approx$ 1 mA/cm^2. The sputtered atoms consist of a mixture of neutrals and ions, and the latter may be mass-spectrally analyzed in order to perform elemental depth profiling. The sensitivity of the technique is very high – for example on the scale of dopant concentrations within semiconductors – but standards are required for quantitative analysis. One variant of SIMS makes use of lower energy primary ions (0.5–2 keV, $\approx$ 1 nA/cm^2) with an almost negligible sputter rate, thus enabling

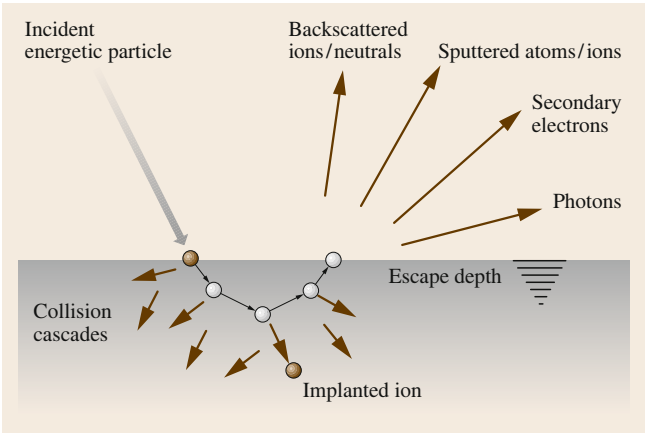


Fig. 17.3 Schematic illustrating the signals produced by the interaction of an energetic ion beam with a sample

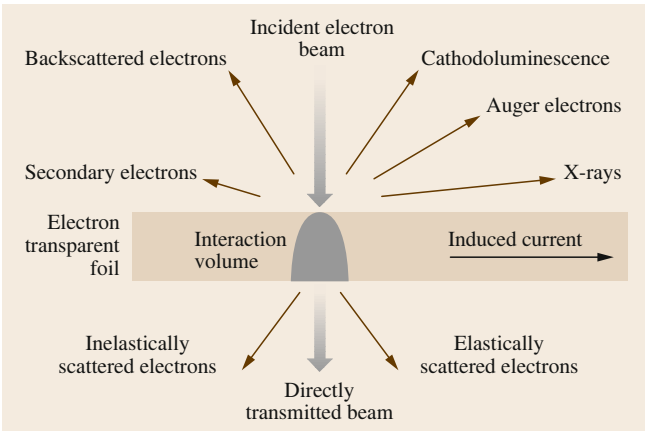


Fig. 17.4 Schematic illustrating the signals produced by the interaction of an energetic electron beam with a thin sample foil

surface chemical analysis. Conversely, the technique of RBS makes use of a very high energy (2−3 MeV) beam of light ions bombarding a sample normal to its surface. An ion such as helium is chosen to avoid the effects of sputtering, whilst high energy is required to overcome the problem of ion neutralization and the screening interaction potential between the ion and the nucleus associated with techniques such as low-energy ion scattering (LEIS). As an energetic positive ion penetrates the sample it loses energy, mainly due to collisions with electrons, but occasionally (and more significantly) with nuclei. The energy of the backscattered ion depends on the depth of the collision and the mass of the target atom. By measuring the energy spectrum of the recoiling ions, information on elemental composition and depth within a sample can be obtained.

By way of comparison, Fig. 17.4 illustrates the variety of signals produced from the interaction of an energetic beam of electrons with a thin ($< 1\,\mu$m) semiconductor sample. In order to make sense of the origins of the many different signals, we must consider the phenomenon of electron scattering, which underpins them all. The process of elastic (Rutherford) scattering arises from an electrostatic (Coulomb) interaction with the nucleus and surrounding electrons of an atom, leading to a change in the incident electron direction without loss of energy. Elastically scattered electrons contribute to the formation of diffraction patterns and diffraction contrast images in TEM. Conversely, inelastically scattered electrons have, by definition, lost a certain amount of energy. In this context, core–shell interaction processes produce scattered electrons whose energies depend on the atomic number of the scatterer, and an analysis of the loss of electron energy (up to ≈ 1000 eV) is the basis of the technique of electron energy loss spectroscopy (EELS). In addition, plasmon scattering (5−30 eV loss) can occur, due to the interactions of incident electrons with waves in the conduction band of a metal, or the bonding electrons of non-metals. The signatures from plasmon-scattered electrons can dominate the low-energy regimes of EEL spectra, providing information on sample thickness. Phonon scattering (≈ 1 eV loss) can also occur, which is the interaction of incident electrons with quantized atomic vibrations within a sample, leading to the production of heat. In an interaction of an electron beam with a bulk sample, nearly all of the incident energy ends up being dissipated through such phonon interactions. The probability of each type of scattering interaction is commonly expressed either as a cross-section, representing the apparent area the scattering process presents to the electron, or as a mean free path, which is the average distance the electron travels before being scattered.

A variety of secondary events also occur as a direct consequence of these primary electron scattering processes. For example, following interactions with a high-energy beam of electrons, excited atoms may subsequently relax in a number of different ways (in a similar fashion to the processes induced by incident high-energy x-ray photons). If core-level (inner shell) electrons are displaced, the relaxation of electrons from higher energy shells to the lower energy core states can lead to the discrete emission of x-ray photons characteristic of the atomic number of the element concerned. This is referred to as the *K, L, M, N*, series, and Moseley's law states that the square root of the frequency of the characteristic x-rays of this series, for certain elements, is linearly related to the atomic number. Discrimination of these characteristic x-rays, as a function of energy or wavelength, provides the basis for the EDX or the WDX techniques, respectively. Alterna-

Table 17.1 Overview of characterization techniques. EDS = energy dispersive x-ray spectroscopy, BSE = backscattered electron

Technique	Primary beam	Energy	Signals detected	Assessment	Spatial resolution	Elements detected	Detection limit
AES	Electron	0.5–10 keV	Auger electron	Surface composition	Lateral ≈ 200 nm (LaB_6 source) lateral ≈ 20 nm (FE source) depth ≈ 2–20 nm	Li–U	≈ 0.1–1 at % (sub-monolayer) accuracy ≈ 30%
RBS	Ion (He^{2+})	> 1 MeV	Ion (He^{2+})	Depth composition and thickness	Lateral ≈ 1 mm depth ≈ 5–20 nm	B–U	≈ 0.001–10 at %
SEM/EDS	Electron	0.3–30 keV	Electron (SE, BSE) x-ray (characteristic)	Surface morphology and composition	≈ 1–3 nm (SE) ≈ 2–5 nm (BSE) lateral > 0.3 µm (EDS) depth ≈ 0.5–3 µm (EDS)	B–U	≈ 0.1–1 at % accuracy ≈ 20% (depends on matrix)
SIMS	Ion (Cs^+, O_2^+, Ga^+)	1–30 keV	Ion (secondary)	Depth trace composition	Lateral ≈ 60 µm (Dynamic SIMS) lateral ≈ 1 µm (Static SIMS) depth ≈ 2–20 nm	H–U	≈ 10^{-10}–10^{-5} at %
TEM/EDS/ EELS	Electron	100–400 keV	Electron (elastic, inelastic) x-ray (characteristic)	Structure and chemistry of thin sections (high resolution)	≈ 0.1–0.3 nm lateral > 2 nm (EDS) lateral ≈ 1 nm (EELS) energy resolution ≈ 1 eV (EELS)	Up to U	≈ 0.1–1 at % accuracy ≈ 20% (depends on matrix)
XPS	X-ray	1–15 keV	Photo electron	Surface composition (chemical bonding)	Lateral ≈ 10 µm – 2 mm depth ≈ 1–10 nm	Li–U	≈ 0.1–1 at % (sub-monolayer) accuracy ≈ 30%
XRD	X-ray	1–20 keV	X-ray	Structure	Lateral ≈ 10 µm depth ≈ 0.1–10 µm	Low Z may be difficult to detect	≈ 3 at % in a two-phase mixture (≈ 0.1 at % for synchrotron) accuracy ≈ 10%
XRF	X-ray	30 kV/20 mA	X-ray (fluorescent)	Composition	Lateral ≈ 0.1–10 mm depth ≈ 10 nm	Na–U	≈ ppb–ppm, accuracy ≈ 10%

tively, a secondary process of Auger electron emission may occur, particularly for low atomic number materials, whereby outer shell electrons are ejected with a characteristic kinetic energy. It should be noted that characteristic x-rays may also induce the emission of lower energy x-rays within a sample, and this is the basis of x-ray fluorescence (XRF) and the origin of background scintillation during EDX analysis.

Further, if an outer (valence) electron state is vacant, relaxation across the band-gap of a semiconductor may occur with the emission of light, and this constitutes the basis of the CL technique. Alternatively, a current may be induced within a sample and nonradiative recombination pathways in the presence of structural defects (and a collection junction) provide the contrast mechanism for the EBIC technique that profiles the electrical activity within a crystalline semiconductor. Also, incident electrons that interact with atomic nuclei may become backscattered with energies comparable to the incident energy, and used to image a sample surface with contrast that is dependent on the average local composition. Secondary (low-energy, $< 50\,\text{eV}$) electrons (SEs), produced by a variety of mechanisms, may also be emitted and escape from the near-surface of a sample. SEs can be used to obtain topographic images of irregular surfaces since they are easily absorbed.

Table 17.1 lists the most commonly used characterization techniques based on x-ray, ion or electron beam interactions with a semiconductor, and broadly indicates their limits of applicability.

17.3 X-ray Diffraction

The basic concepts behind radiation-material interactions and scattering link into the concept of diffraction, whereby the spatial distribution and intensity of scattered x-rays or electrons provide information on the arrangement of atoms in a periodic sample. The theory of wave-particle duality proposes that an electron may be considered to be a wave rather than a particle when discussing diffraction. Electrons are scattered by electric fields within a crystal whilst x-rays are scattered by shell electrons. Nevertheless, the geometry of diffraction is very similar in both cases, being governed by Bragg's law.

The basic principles of diffraction (in reflection, transmission or glancing angle geometries) may be introduced with reference to x-ray scattering and interference. X-rays are a form of energetic electromagnetic radiation of wavelength $\approx 10^{-10}$–$10^{-11}\,\text{m}$, of comparable size to the spacing of atoms within a solid. A crystal lattice comprises a regular array of atoms; and the electron clouds around these atoms act as point sources for spherical x-ray wavelets, through a process of absorption and re-emission, when interaction with an incident x-ray beam occurs. Constructive interference of the scattered wavelets, related to the Huygens's principle for wave propagation (Fig. 17.5), occurs in preferred directions, depending on the Bravais lattice of the crystal and the x-ray wavelength. The positions of the resultant maxima in scattering intensity may be used to deduce crystal plane spacings and hence the structure of an unknown sample when correlated with database values. Geometrical considerations show that the scattering angles corresponding to diffracted intensity maxima can be described by the Bragg equation $n\lambda = 2d \sin\theta$ (sometimes expressed with Miller indices notation as $\lambda = 2d_{hkl} \sin\theta_{hkl}$). Formally, this equation describes the minimum condition for the coherent diffraction of a monochromatic x-ray beam from a set of planes of a primitive lattice. Figure 17.6 illustrates the geometrical conditions associated with Bragg diffraction from a set of $\{hkl\}$ planes spaced d_{hkl} apart, with x-rays incident at a Bragg angle θ being diffracted through an angle 2θ. The path difference between the x-rays *reflected* from successive planes must be equivalent to an integer number of wavelengths n for constructive interference to occur.

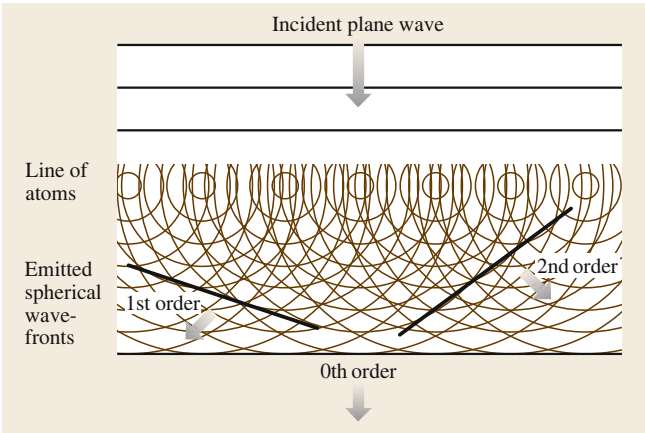


Fig. 17.5 Schematic illustrating Huygens's principle, with a reconstruction of spherically emitted x-ray wavefronts providing diffracted intensity in specific directions

In order to interpret the information contained in the *intensity* of the diffracted beam, which is measured in practice, it is necessary to consider the amplitude of

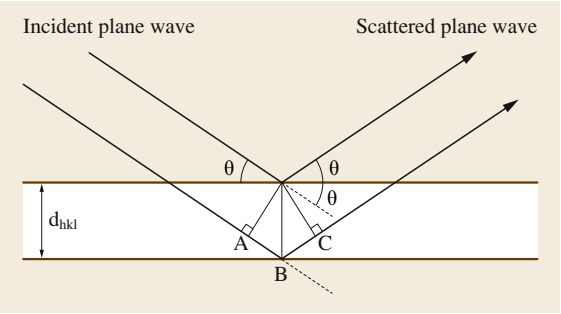


Fig. 17.6 Geometric illustration of Bragg diffraction. For constructive interference, $n\lambda = AB + BC = 2d \sin\theta$

the elastically scattered waves. (The phase of the scattered beam is much more difficult to measure.) A useful concept to introduce at this stage is that of the atomic scattering factor f, which is a measure of the amplitude of the wave scattered by an atom, which depends on the number of electrons in the atom. Formally, the atomic scattering factor is the ratio of the amplitude of the wave scattered by an atom to the amplitude scattered in the same direction by a free classical electron. (For the purpose of such discussions, an electron orbiting an atom is considered to be a free classical electron.) The scattered amplitude (and hence intensity = |amplitude|^2) varies with direction, with higher angles θ being associated with lower amplitudes.

When x-rays interact with a periodic crystal lattice, it is considered that each atom scatters with an amplitude f into an *hkl* reflection, while there is a summation of all of the scattering amplitudes from different atoms, with phase differences that depend on *hkl* and the relative positions of the atoms. This leads to the concept of the structure factor, F_{hkl}, which is the total scattering amplitude from all of the atoms in one unit cell of a lattice. Formally, the structure factor is the ratio of the amplitude scattered by a unit cell into an *hkl* reflection to that scattered in the same direction by a free classical electron. The phase difference between waves scattered from two different atoms depends on the Miller indices of the reflection being considered and the fractional coordinates of the atoms within the unit cell. In general terms, for a reflection from a set of $\{hkl\}$ planes, the phase difference ϕ between the wave scattered by an atom at the origin and that scattered by an atom with fractional coordinates x, y, z is given by $\phi = 2\pi(hx + ky + lz)$. In theoretical terms, the resultant intensity of the scattered beam is denoted $I_{hkl} \propto |F_{hkl}|^2$, where $F_{hkl} = \sum f_j(\theta) \exp(i\phi_j)$, which is a summation of the individual scattered sinusoidal waves, performed over both phase and amplitude. In a practical diffraction experiment, however, the combination of photoionization and Compton scattering can act to diminish the scattered beam intensity. Consideration also needs to be given to the effects of absorption, along with the effect of multiplicity, which arises from the number of symmetrical variants of a unit cell, and the geometrical and polarization factors specific to a given experimental arrangement.

The crystallographic structure of an unknown material can, nevertheless, be analyzed via the diffraction of x-rays of known wavelength. For example, for a crystal system with orthogonal axes, the general formula which relates plane spacing, d_{hkl}, to the plane index $\{hkl\}$ and the lattice parameters a, b, c is $1/(d_{hkl})^2 = h^2/a^2 + k^2/b^2 + l^2/c^2$. For a cubic system, this simplifies to

$$d_{hkl} = \frac{a}{\sqrt{(h^2 + k^2 + l^2)}} = \frac{a}{\sqrt{N}} .$$

Thus, $\lambda = 2a \sin\theta/\sqrt{N}$, from which $N = (4a^2/\lambda^2) \times \sin^2\theta$ and hence $N \propto \sin^2\theta$. Accordingly, the 2θ angles of scattering arising from the process of x-ray diffraction may be used to identify values for N, against which *hkl* indices can be assigned and the lattice identified. Information is contained within the intensities and widths of the diffracted x-ray peaks, in addition to their positions. The process of formal identification can be automated by electronically referencing crystallographic databases containing the positions and relative magnitudes of the strongest diffraction peaks from known compounds. Systematic or partial absences may also occur depending on the lattice symmetry, and the intensities of some peaks may be weak due to the motif of the crystal lattice – the number of atoms associated with each lattice point. For example, sphalerite or wurtzite lattices may show weak reflections that are absent for face-centered cubic or hexagonal close-packed structures, respectively.

Partial differentiation of the Bragg equation gives $2(\sin\theta)\delta d + 2d(\cos\theta)\delta\theta = 0$, from which $\delta d/\delta\theta = -d\cot\theta$. Thus, for a fixed error in θ, the error in d_{hkl} will minimize as $\cot\theta$ tends to zero (as θ tends to 90°). Hence, diffraction techniques are more accurate when measurements are made at high 2θ angles. Similarly, it can be shown that the sensitivities and resolutions of diffraction measurements improve at high scattering angles. The sensitivity of the x-ray diffraction technique, for example, is sufficiently high to enable subtle stress measurements to be made, with elastic changes in plane spacings leading to small shifts in diffracted peak positions. Accordingly, diffraction techniques may be used to investigate temperature-dependent order–disorder transitions within alloys and preferred orientation effects. The high accuracy of lat-

tice parameter measurement similarly enables good compositional analysis of an alloy when Vegard's law applies, i. e., the alloy exhibits a linear dependence of lattice parameter on composition, between the extremes of composition. Brief reference should be made here to the related technique of neutron diffraction, which is able to sample over a much larger range of d_{hkl} spacings, which makes it particularly useful for the ab initio determination of very complex structures. The very rigorous approach of Rietveld refinement can also be used to model a complete diffraction profile, deducing the structure from first principles using atom positions as fitted parameters.

The principle behind the generation of x-rays from characteristic transitions, used when performing chemical microanalyses of unknown samples by EDX or WDX, also provides the route to defining an x-ray source from a known target sample for the purpose of XRD. In practical terms, the x-ray tube of a diffractometer is an evacuated vessel in which electrons from a hot filament are focused onto a cooled metal target, such as Cu or Ni. The x-ray spectrum generated consists of x-rays characteristic of the target material, along with a background emission of x-rays of continuous wavelengths, termed Bremsstrahlung (braking radiation), which are caused by the acceleration of electrons in the vicinity of nuclei. The x-rays emerging from the tube, through a window made from a material of low atomic number, can be filtered and collimated to define a beam of specific wavelength.

The precise geometry used in a diffraction experiment will depend on the form of the sample and the information content required. For example, crystalline samples such as semiconductor heterostructures can be rotated within a cylindrical geometrical framework that enables successive sets of crystal planes to be brought into play for detection. Figure 17.7 illustrates the basic geometrical arrangements for a four-circle diffractometer that allows diffraction data to be acquired. Computer-controlled rotation of the sample around ω, χ and ϕ axes, as the detector is rotated about the 2θ axis, enables the positions and intensities of *hkl* reflections to be recorded. By way of example, Fig. 17.8 shows a $2\theta/\omega$ plot recorded from a heteroepitaxial GaN/GaAs(001) sample, with reflections attributable to both epilayer and substrate, whilst the full width at half maximum (FWHM) values for the diffraction peaks provide a measure of the mosaic spread of the subgrains within the sample.

Since amorphous materials do not exhibit long-range order, their diffraction profiles show diffuse intensities rather than well-defined maxima. Partially crystalline materials may show broad diffraction peaks, from which it is possible to approximate the crystal

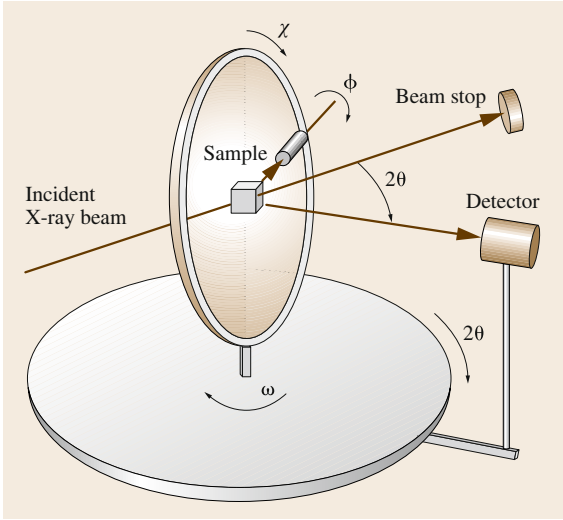


Fig. 17.7 X-ray four-circle diffractometer

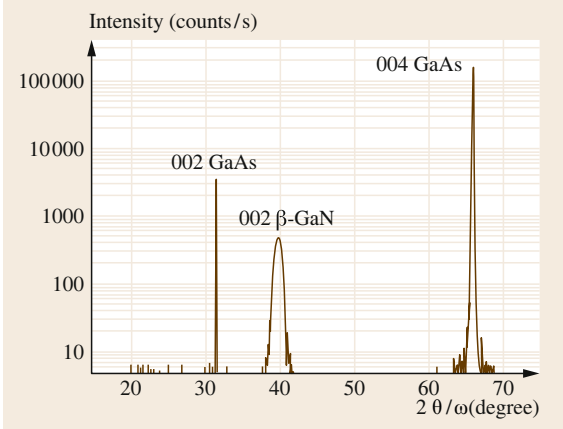


Fig. 17.8 Example of a $2\theta/\omega$ plot for cubic GaN/GaAs(001)

size from the λ–peak-width relationship. Other x-ray diffraction techniques include Laue back reflection, that can be used conveniently to orient bulk single crystals, for example for sectioning prior to use as substrates for heteroepitaxial growth. Alternatively, the Debye–Scherrer method can be used for powder samples, since a significant number of crystal grains will always be in an orientation that satisfies the Bragg equation for each set of {*hkl*} planes. In this scattering arrangement, the diffracted rays form cones coaxial with the incident x-ray beam, with each cone of diffracted rays corresponding to a Bragg reflection from a specific set of lattice planes in the sample.

To reiterate, it is the combination of Bragg's law and the structure factor equation that enables the directions and intensities of beams scattered from a crystal to

be predicted. In this context, it is instructive to briefly compare XRD with electron diffraction. Electrons are scattered by the periodic potential – the electric field – within a crystal lattice, whilst x-rays are scattered by outer shell electrons. Since x-rays and electrons exhibit comparable and comparatively small wavelengths, respectively, on the scale of the plane spacings of a crystal lattice, this equates to large and small angles of scattering, respectively, for the diffracted beams. Accordingly, XRD offers greater accuracy than electron diffraction for the measurement of lattice parameters. It should also be noted that XRD is essentially a kinematic process based on single scattering events, whilst electron diffraction is potentially more complex due to the possibility of dynamic (or plural) scattering processes which can affect the generated intensities. Also, electrons are more strongly absorbed than x-rays, so there is need for very thin sample foils, typically $< 1\,\mu m$, for the purposes of transmission electron diffraction (TED) experiments. However, electrons are more easily scattered by a crystal lattice than x-rays, albeit through small angles, so an electron-transparent sample foil is capable of producing intense diffracted beams. X-rays require a much greater interaction volume to achieve a considerable diffraction intensity. The effectiveness of the technique of electron diffraction becomes most apparent when combined with TEM-based chemical microanalysis and imaging techniques. This enables features such as small grains and embedded phases, or linear or planar defect structures such as dislocations and domain boundaries, to be investigated in detail.

Before describing some variants of the electron diffraction technique, a few concepts related to imaging and modes of operation of the TEM need to be introduced.

17.4 Optics, Imaging and Electron Diffraction

The aim of a microscope-based system is to image an object at high magnification, with optimum resolution and without distortion. The concepts of magnification and resolution associated with imaging in electron microscopy are usually introduced via light ray diagrams for optical microscopy. The constraints on achieving optimum resolution in TEM are generally considered to be lens aberration and astigmatism. The concepts of depth of field and depth of focus must also be considered.

If we consider the objective lens shown in Fig. 17.9, a single lens is characterized by a focal length f and a magnification M. The expression $1/f = 1/u + 1/v$ relates the focal length to the object distance u and the image distance v for a thin convex lens. The magnification of this lens is then given by $M = v/u = f/(u-f) = (v-f)/f$, from which it is apparent that $u-f$ must be small and positive for a large magnification to be obtained. In practice, a series of lenses are used to achieve a high magnification overall whilst minimizing distortion effects. For the combined projection optical microscope system shown in Fig. 17.9, magnification scales as $M' = (v-f)(v'-f')/ff'$.

Resolution may be defined as the smallest separation of two points on an object that can be reproduced distinctly within an image. The resolution of an optical lens system is diffraction-limited since light must pass through a series of apertures, and so a point source is imaged as a set of Airy rings. Formally, the minimum resolvable separation of two point sources, imaged as two overlapping sets of Airy rings, is given by the Rayleigh criterion, whereby the center of one set of Airy rings overlaps the first minimum of the second set of Airy rings. The defining equation for resolution is

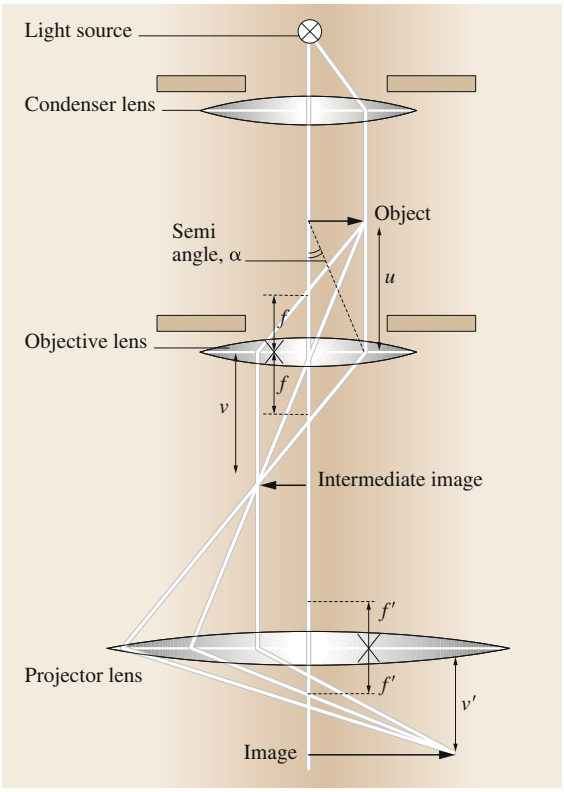


Fig. 17.9 Schematic ray diagram for a transmission projection optical microscope

given by $r = 0.61\lambda/n\sin\alpha$, where λ is the wavelength of the imaging radiation, n is the refractive index of the lens, and α is the semiangle subtended at the lens. The combined term $n\sin\alpha$ is the *numerical aperture* of the lens. Thus, resolution can be improved by decreasing λ or by increasing n or α. It is noted that magnification beyond the resolution limit of the system is possible, but there is no benefit to achieving this, and so it is termed *empty magnification*.

The performance of a TEM can be described in analogous terms, comprising an electron gun and a series of electromagnetic lenses for sample illumination, image formation and magnification: the condenser, objective and projector systems, respectively (Fig. 17.10). The electron source can be either a hot filament (e.g., W or LaB_6) for thermionic emission, or a hot or cold cathode (e.g., ZrO_2 coated W) for field emission. The gun and lenses are traditionally assembled in a column with a viewing screen and camera at the bottom to record images or diffraction patterns from an electron-transparent sample placed within the objective pole piece. A range of differently sized apertures are located in the condenser system to help collimate the electron probe, in the projector system to select a region of sample from which a diffraction pattern may be formed, and in the objective lens just below the sample to select the transmitted beam(s) used to form an image. The path of travel of the electron beam through the entire electron-optic column must be under conditions of high vacuum, considering the ease of absorption of electrons in air. The electrons are accelerated by high voltage, typically 100 or 200 kV, although there are TEMs that operate at MV conditions, depending on the intended application. An accelerated, high-energy electron acquires significant kinetic energy and momentum. It can also be represented by a wavelength (corrected to take into account relativity) that can be approximated in nanometers by $\lambda = [1.5/(V + 10^{-6}V^2)]^{0.5}$. By way of example, an electron beam within a conventional TEM operating at 200 kV has a relativistically corrected wavelength of 0.0025 nm, as compared with the range of $\approx 400-700$ nm for visible light. Since resolution is wavelength-limited, the technique of TEM is, in principle, able to provide a vast improvement in resolution over conventional optical systems. The equation for resolution becomes approximated by $r \approx 0.61\lambda/\alpha$ since $n = 1$ for electromagnetic lenses and $\sin\alpha \approx \alpha$ in view of the small angles associated with electron diffraction. A value of $\alpha \approx 150$ mrad would suggest a resolution approaching ≈ 0.01 nm. However, this level of resolution for TEM is generally not achievable in practice due to the effect of lens aberration, which degrades the resolution of an image to a *disc of least confusion* above the theoretical wavelength-limited resolution of the electron-optic system. Resolution values of $\approx 0.2-0.3$ nm are associated with the best conventional instruments. (When charting the historical development of TEM, improvements in resolution have depended on the construction of microscopes operating at higher voltages, in order to capitalize on the benefits from the reduced electron wavelength. However, it should be noted that recent improvements in techniques of compensating for spherical and chromatic aberration have now enabled a resolution of $\lesssim 0.1$ nm to be achieved for intermediate-voltage, field emission gun instruments.)

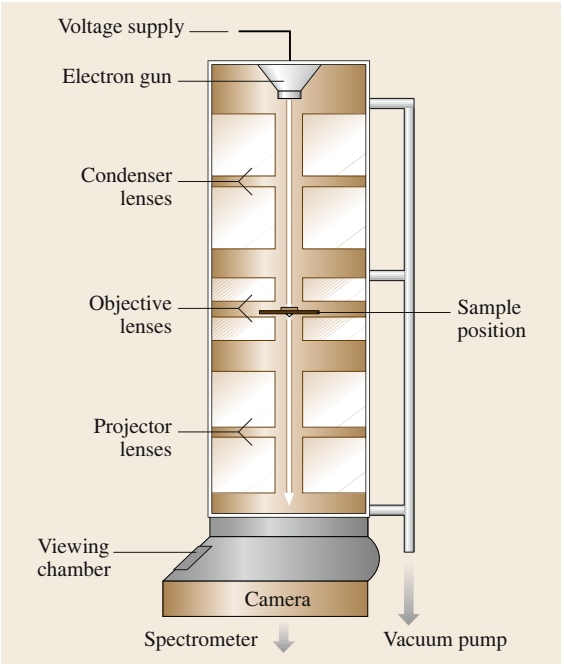


Fig. 17.10 Simplified lens configuration of a TEM

The variation in focal length of a lens as a function of the distance of a beam from the center of the lens results in rays traveling further from the optic axis being brought to focus closer to the lens (Fig. 17.11a). This spread in path lengths of rays traveling from an object to the image plane is termed spherical aberration. In this case, the radius of the disc of least confusion is given by $r_s = C_s\alpha^3$, when referred back to the object, where C_s is termed the spherical aberration coefficient and α is again the semiangle, in radians, subtended at the lens. Spherical aberration can be limited by reducing α; in other words by *stopping down* the lens by using smaller apertures (in this case, the term α corresponds to the *aperture collection angle*). However, this conflicts with the large value of α needed to optimize resolution, so a balance is required for optimum resolution, which is achieved when $\alpha_{opt} \sim (\lambda/C_s)^{1/4}$, corresponding to

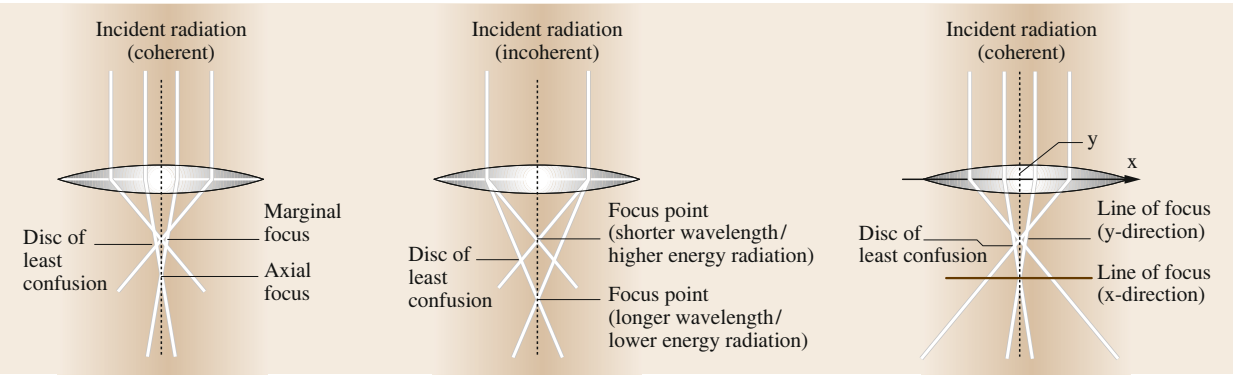


Fig. 17.11a–c Schematic illustration of **(a)** spherical aberration; **(b)** chromatic aberration and **(c)** astigmatism

an effective resolution of $r_{opt} \sim \lambda^{3/4} C_s^{1/4}$. Lenses also exhibit different focal lengths for electrons with different wavelengths; such a spread of wavelengths can arise from slight fluctuations in the accelerating voltage or from inelastic scattering (energy loss) processes within the specimen, and this effect is termed chromatic aberration (Fig. 17.11b). The radius of the disc of least confusion in this instance is given by $r_c = C_c \alpha (\Delta E/E_0)$, where C_c is the chromatic aberration coefficient, E_0 the accelerating voltage and ΔE the spread in electron energy.

An additional aberration, termed astigmatism, arises from the asymmetry of a lens about the optic axis. The different focal lengths of a lens for different orientations leads to a loss of sharpness of the image at focus. One measure of the astigmatism is the distance between the two foci formed at right angles along the optic axis (Fig. 17.11c), in contrast to a single point of focus. Astigmatism may be compensated for by using electromagnetic stigmators that generate a compensating field to bring the rays back into common focus. Distortion of an image can also occur due to slight variations in magnification with radial distance from the optic axis, leading to so-called *pin-cushion* or *barrel* distortion. This effect can become particularly noticeable at very low magnification.

The term *depth of field* is defined as the distance along the optic axis that an object can be moved without noticeably reducing the resolution. This effect is again dependent on the radius of the disc of least confusion that can be tolerated and α. The depth of field approximates to λ/α^2, and a typical value of a few tens of nm for electron microscopy means that every point within the thickness of a typical electron-transparent foil can be imaged at focus. Conversely, the *depth of focus* corresponds to the maximum permissible spacing between the imaging screen and the photographic plate or CCD used to record an image. The depth of focus approximates to $\lambda M^2/\alpha^2$, and since this works out at many meters, the viewing screen and recording system of an electron microscope need not coincide.

Following on from these general considerations affecting the process of imaging, we now move on to briefly describe the conventional modes of operation of a TEM. The accelerating stack and condenser system of the microscope defines the high-energy electron probe incident at a thin sample, within which many complex interactions occur and various signals are produced, as detailed earlier. A diffraction pattern is initially formed in the back focal plane of the objective lens and the recombination of diffracted beams allows the reconstruction of an inverted image in the first image plane. Changing the strengths of the intermediate lenses allows the back focal plane (corresponding to a projection of reciprocal space) or image plane (corresponding to a projection of real space) to be observed on the viewing screen (Fig. 17.12a,b). Placing an aperture with selected area in an intermediate image plane, around a feature of interest within a projected image, ensures that only beams from that particular area of the sample contribute to the diffraction pattern viewed on the screen.

It is sometimes convenient to think of the intersection of an Ewald sphere with a reciprocal lattice when describing the construction and projection of a diffraction pattern. A reciprocal lattice is constructed from a crystal lattice such that any vector from the origin to a diffracted spot is normal to a particular plane in the real lattice, with a reciprocal length given by the plane spacing. Thus, a three-dimensional crystal lattice can be transformed into a three-dimensional reciprocal lattice (mathematically, the process is a Fourier transform). The Ewald sphere can be thought of as a geometrical construction of radius $1/\lambda$ intersecting the reciprocal lattice. Figure 17.13a illustrates the construction of an Ewald sphere for the case of large-angle scattering (x-ray diffraction). The direction of the incident beam at the sample corresponds to the direction of the vertical radius of the Ewald sphere, and the point of emer-

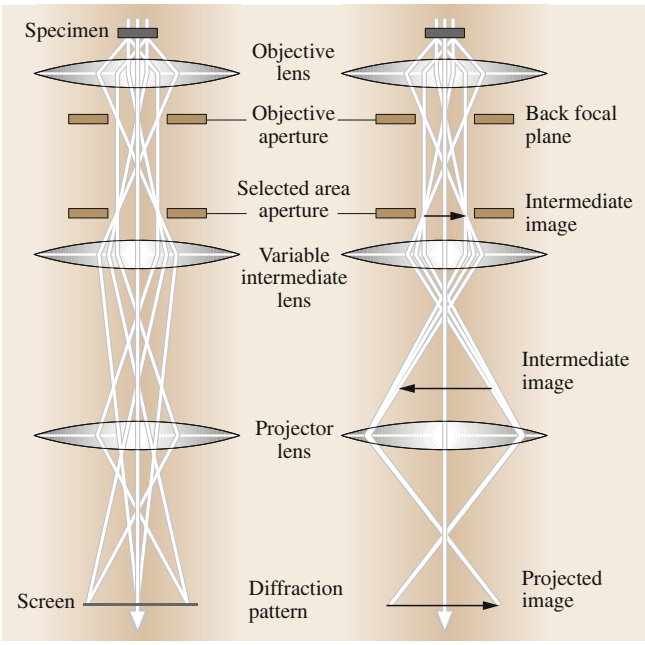


Fig. 17.12a,b Schematic ray diagrams showing the projection of **(a)** a diffraction pattern and **(b)** an image onto a TEM viewing screen

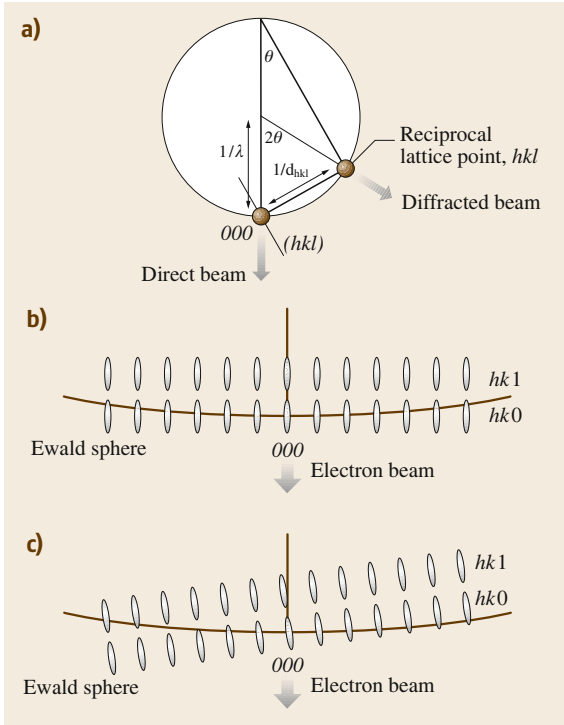


Fig. 17.13a–c Ewald sphere construction for **(a)** x-ray diffraction (*large angle scattering*); **(b)** electron diffraction (*zone axis projection*) and **(c)** electron diffraction (*tilted projection*)

gence of this vector at the sphere surface, coincident with a reciprocal lattice point, is taken as the origin 000 of the diffraction pattern. The same construction applies for a beam of electrons, but the sphere surface has very shallow curvature relative to the reciprocal lattice spacings, due to the very small value of λ relative to d_{hkl} (Fig. 17.13b). Further, diffraction from a thin crystal is associated with a lengthening of the reciprocal lattice spots into rods in a direction parallel to the electron beam. Consequently, if the electron beam is incident along a low index zone axis, the Ewald sphere approximates to a plane and intersects a layer of reciprocal lattice points, and a two-dimensional array of diffraction spots is projected. In very general terms, it is considered that a set of {*hkl*} planes is at the Bragg condition when the reciprocal lattice point corresponding to *hkl* falls on the surface of the Ewald sphere. For the case of larger angles of electron diffraction, an outer ring of diffraction spots may be observed, termed a high-order Laue zone (HOLZ), since the Ewald sphere has sufficient curvature to intersect with a neighboring layer of points within the reciprocal lattice (Fig. 17.13c).

We should also mention the Kikuchi lines that arise in diffraction patterns due to the elastic scattering of incoherently scattered electrons. The intensities of these Kikuchi lines increase with increasing thickness of sample foil, and the line spacings are the same as the spacings of the diffraction spots from the associated crystal planes. Kikuchi lines move as the sample is tilted and hence can be used to establish very precise crystal orientations for the purpose of image contrast or convergent beam electron diffraction (CBED) experiments.

Referring back to Fig. 17.12, a diffracted beam may be selected using an aperture inserted into the back focal plane of the objective lens and used to form an image. If the undeviated, transmitted beam is used, then a bright field image is formed where the areas that diffract strongly appear dark. Image contrast also arises from a mass-thickness effect, whereby thicker or high-density regions of material scatter more strongly and hence appear dark. Alternatively, a centered dark field image may be created by tilting the illumination to center the diffracted beam, termed a $\boldsymbol{g}_{hkl}$ reflection, down the optic axis of the microscope (in contrast to moving the aperture over the diffraction spot, which would increase aberrations and degrade resolution). Areas where the {*hkl*} planes diffract strongly appear bright in such cases. Weak beam images can also be produced under dark field imaging conditions, with the sample tilted slightly away from a strong Bragg condition. The trade-off, in this instance, is reduced image contrast for the

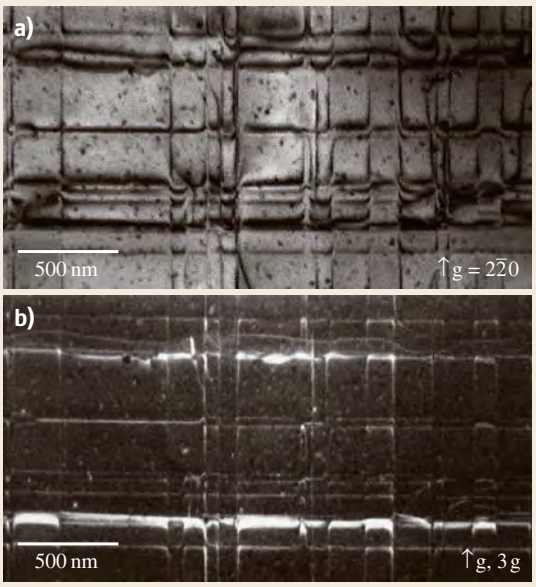


Fig. 17.14 (a) Bright-field and (b) weak-beam plan-view TEM images of Si/SiGe/Si(001)

improved resolution, which allows fine detailed features, such as partial dislocations bounding dissociated dislocations, to be delineated.

By way of example, Fig. 17.14 shows the different contrasts obtained from a plan-view Si/$Si_{0.96}Ge_{0.04}$/Si(001) sample foil imaged under bright-field and weak-beam diffraction conditions, respectively. The sample foil contains orthogonal arrays of interfacial misfit dislocations that are generated as a consequence of the relaxation of the strain arising from the different lattice parameters of Si and $Si_{0.96}Ge_{0.04}$ [17.23]. The strain fields around the dislocations lead to a local deviation away from the exact Bragg condition, which can be used to delineate the (approximate) positions of the dislocation cores. The weak beam technique allows for complex dislocation tangle to be resolved more clearly. Imaging such sample foils using different diffraction vectors produces different contrasts that allow the nature of the dislocations to be precisely determined. Hence, the dislocation reaction mechanisms responsible for microstructure development can, in principle, be identified.

Some examples are now presented that illustrate how electron diffraction can contribute to the structural characterization of semiconductors. The use of conventional TED combined with image contrast analysis to assess the defect microstructure of a semiconductor heterostructure is initially considered. The application of the CBED technique, sometimes termed microdiffraction, in order to determine the polarities of noncentrosymmetric crystals is then described. We also focus on the RHEED technique, which may be used for rapidly assessing the near-surface microstructures of semiconductor thin films.

17.4.1 Electron Diffraction and Image Contrast Analysis

Figure 17.15 shows a TED pattern corresponding to a highly symmetric, low-index, zone axis projection, acquired from an electron-transparent foil of epitaxial GaN grown on the {0001} basal plane of sapphire, viewed in cross-section under spread beam conditions, i. e., near-parallel illumination.

It is generally instructive to view such diffraction patterns alongside direct images of the sample. Figure 17.16 compares high-resolution electron microscopy (HREM) and conventional TEM (CTEM) images of heteroepitaxial GaN–sapphire sample acquired for this projection. The HREM mode of imaging, otherwise termed *phase contrast* or *lattice imaging*, makes use of several diffracted beams selected by a large objective aperture; the resulting interference patterns can be used to elucidate the locations of atomic columns (Fig. 17.16a). Careful simulation is required to precisely assign atomic positions within such images, since the contrast is strongly dependent on defocus, foil thickness and sample orientation. However, the structural integrity of the interface between GaN and sapphire can be appraised, along with the presence of nanometer-scale, three-dimensional growth islands formed during the initial stages of epitaxy, prior to layer coverage. Conversely, the conventional *many-beam, bright-field* image of heteroepitaxial GaN–sapphire (Fig. 17.16b) was created using a small objective aperture placed over the directly transmitted beam, and recorded at lower magnification. The image contrast is again complicated

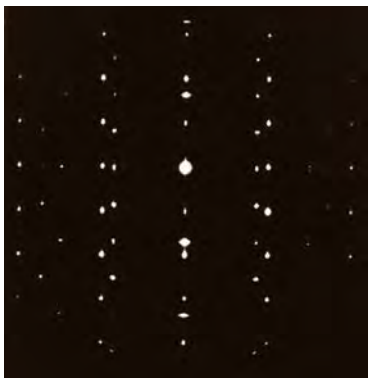


Fig. 17.15 TED pattern acquired from epitaxial GaN/sapphire{0001} viewed in cross-section, corresponding to the $\langle 11\bar{2}0\rangle_{\text{nitride}} \,||\, \langle 1\bar{1}00\rangle_{\text{sapphire}}$ axis projection

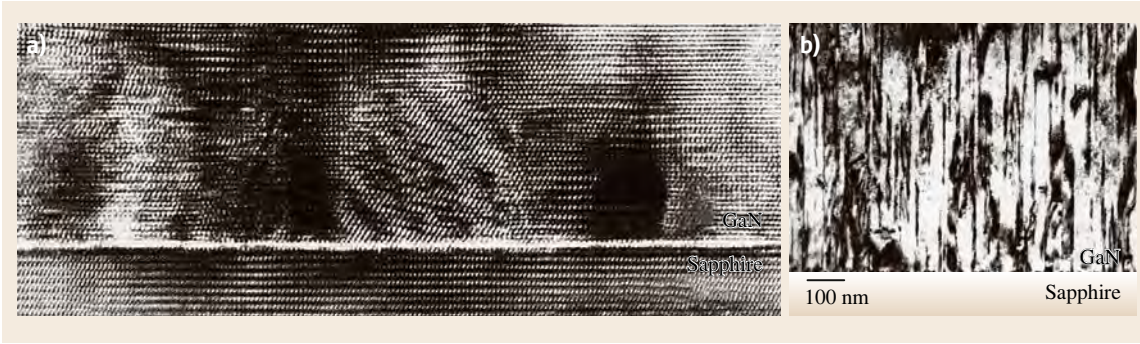


Fig. 17.16 **(a)** HREM and **(b)** many-beam CTEM images of GaN/sapphire{0001}

because there are many excited diffracted beams operating for this low-index crystal orientation combined with strain fields associated with the large number of threading defects ($\approx 10^{10}\,\mathrm{cm}^{-2}$) within this sample. Despite this level of defect content, the diffraction pattern of Fig. 17.15 still reflects a high level of crystallographic perfection of the matrix. Also, rotation of this sample foil by 30° about the growth axis enabled a $\langle 1\bar{1}00\rangle_{\mathrm{nitride}} \,||\, \langle 11\bar{2}0\rangle_{\mathrm{sapphire}}$ projection to be established, consistent with the hexagonal symmetry of this GaN–sapphire system. Indeed, establishing different diffraction patterns for different sample projections, for known angles of sample tilt, enables the phase and structural relationship of an epilayer and substrate to be clarified.

There is a need to characterize the precise nature of the fine-scale defect content within such samples. Dislocations, for example, generally act as nonradiative recombination centers, which can deleteriously affect the charge transport properties of a semiconductor [17.24]. Gaining an improved understanding of how dislocations are created and how they interact enables us to identify mechanisms that could be used to control their development, thereby improving the growth of a semiconductor and the resultant properties of devices made from it.

Dislocations are one-dimensional defects that can be pure edge, pure screw or *mixed edge-screw* in character. In certain circumstances, a dislocation can dissociate to form a pair of partial dislocations bounding a stacking fault ribbon, the separation of which depends on the material stacking fault energy. Dislocations can be described with reference to their line direction $\boldsymbol{u}$ and Burgers vector $\boldsymbol{b}$, representing the closure failure of a loop of equal atomic spacings around the defect core. Dislocations can move through a crystal lattice by glide or climb mechanisms; for example, when under deformation or as part of a strain relaxation mechanism during heteroepitaxial growth. Also, dislocations can interact to either self-annihilate or multiply to generate more dislocations depending on their respective type, habit plane and the slip system of the matrix.

The technique of diffraction contrast analysis allows the nature of dislocations to be ascertained. The approach used is to tilt the sample away from a highly symmetrical, low-index zone axis orientation in order to establish diffraction contrast images under selected *two-beam* conditions, with one strong diffraction spot excited ($\boldsymbol{g}$) in addition to the central transmitted beam, corresponding to one set of crystal planes at the Bragg condition. When $\boldsymbol{g}\cdot\boldsymbol{b} = 0$, the displacement associated with a dislocation does not affect the diffracting planes used to form the image and so the defect appears invisible. In practice, two examples of this invisibility condition are generally required to determine the precise displacement associated with a given defect. The dislocation will appear visible when $\boldsymbol{g}\cdot\boldsymbol{b} = 1$ and might show a more complex double image for the case of $\boldsymbol{g}\cdot\boldsymbol{b} \geq 2$. For example, since a screw dislocation is characterized by a Burgers vector $\boldsymbol{b}$ parallel to the line direction $\boldsymbol{u}$, the defect is invisible if a diffraction vector perpendicular to the line direction is chosen. In the case of an edge or mixed dislocation where $\boldsymbol{b}$ is not parallel to $\boldsymbol{u}$, there is a stricter requirement for both $\boldsymbol{g}\cdot\boldsymbol{b} = 0$ and $\boldsymbol{g}\cdot(\boldsymbol{b}\times\boldsymbol{u}) = 0$ for true invisibility, otherwise residual contrast might be present that acts to confuse image interpretation. We should also mention the deviation parameter $\boldsymbol{s}$ when establishing a diffracting condition, which represents a short distance in reciprocal space from the exact Bragg condition, since this is associated with imaging artefacts such as extinction contours and thickness fringes that may act to further complicate an image.

Bright-field, dark-field or weak-beam diffraction contrast imaging techniques can be used for defect analysis, depending on the resolution required. By way of example, epitaxial GaN–$\{\overline{111}\}$B GaAs grown by molecular-beam epitaxy (MBE) at 700 °C exhibits a mosaic cell structure with subgrain boundaries delineated by predominantly mixed-type threading dis-

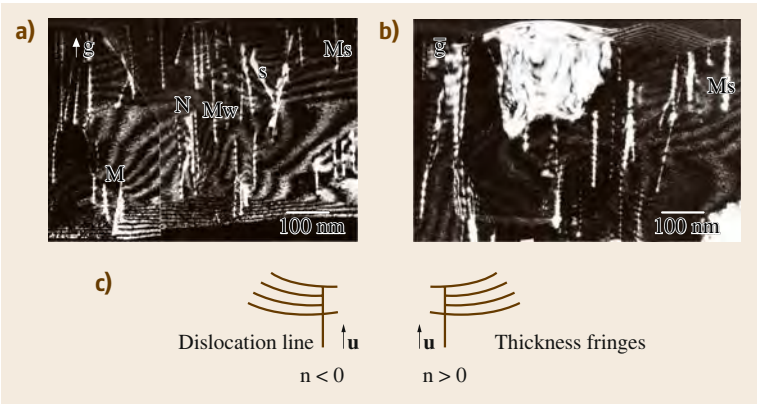


Fig. 17.17a–c Weak-beam images of the same regions of a GaN epilayer viewed in cross-section using **(a)** $\boldsymbol{g} = 000\overline{2}$ and **(b)** $\boldsymbol{g} = 1\overline{1}00$. **(c)** Schematic diagram summarizing Ishida's rule for determining $\boldsymbol{b}$ for $n = -2$ and $n = 2$

locations (typically $> 10^{11}\,\mathrm{cm}^{-2}$) [17.25]. The epitaxial relationship is given by $[000\overline{1}]_{\mathrm{GaN}}||[\overline{111}]_{\mathrm{GaAs}}$, $[1\overline{1}00]_{\mathrm{GaN}}||[2\overline{11}]_{\mathrm{GaAs}}$ and $[11\overline{2}0]_{\mathrm{GaN}}||[01\overline{1}]_{\mathrm{GaAs}}$, and the mismatch between GaN and GaAs is 38.2% between $\{1\overline{1}00\}_{\mathrm{GaN}}$ and $\{2\overline{2}0\}_{\mathrm{GaAs}}$, with the epilayer in tensile strain. Figure 17.17a,b are weak-beam images of the epilayer viewed in cross-section near the $[11\overline{2}0]$ zone axis, using $\boldsymbol{g} = 000\overline{2}$ and $\boldsymbol{g} = 1\overline{1}00$, respectively. Most of the dislocations have a line direction of $\langle 0001\rangle$. Examples of perfect edge, $\boldsymbol{b} = 1/3\langle 11\overline{2}0\rangle$, and screw-type dislocations, $\boldsymbol{b} = \langle 0001\rangle$, are apparent, but mostly dislocations (typically $\approx 70\%$) can be seen in both images, hence they are mixed-type, with Burger's vector components of $\boldsymbol{a}$ and $\boldsymbol{c}$, i. e., $1/3\langle 11\overline{2}3\rangle$.

It is possible to move this analysis forwards by applying Ishida's rule, which allows the magnitude and sense of a Burgers vector to be determined if there are terminating thickness fringes at the exit of a dislocation from a wedge-shaped sample foil [17.26]. If there are n fringes terminating at one end of a dislocation, then $\boldsymbol{g}\cdot\boldsymbol{b} = n$ and the sign of $n (n > 0$ or $n < 0)$ is defined according to Fig. 17.16c. For example, the screw dislocation S (Fig. 17.17a) imaged using $\boldsymbol{g} = 000\overline{2}$ has two thickness fringes terminating on the right-hand side of the dislocation, with the dislocation line pointing towards the $[000\overline{1}]$ growth direction. Hence, $\boldsymbol{g}\cdot\boldsymbol{b} = -2L = 2$, so $\boldsymbol{b}$ is $[000\overline{1}]$ for this screw-type dislocation. For dislocations of mixed type, Fig. 17.17a,b show that some have strong contrast in both images, while others have strong contrast when imaged using $\boldsymbol{g} = 000\overline{2}$ and weak contrast when $\boldsymbol{g} = 1\overline{1}00$. For the former, for example the dislocation denoted as $\boldsymbol{M}_{\mathrm{s}}$ in Fig. 17.17a, thickness fringes terminate at the left-hand side of the dislocation with reference to the dislocation line pointing downwards along [0001]. Hence, $\boldsymbol{g}\cdot\boldsymbol{b} = -2L = -2$ for $\boldsymbol{g} = 000\overline{2}$ and $\boldsymbol{g}\cdot\boldsymbol{b} = -1$ for $\boldsymbol{g} = 1\overline{1}00$, so its Burgers vector is either $1/3[\overline{2}113]$ or $1/3[\overline{1}2\overline{1}3]$. For the cases of strong contrast by $\boldsymbol{g} = 000\overline{2}$ and weak contrast by $\boldsymbol{g} = 1\overline{1}00$, such as the dislocation labeled $\boldsymbol{M}_{\mathrm{w}}$ in Fig. 17.17a, such dislocations have $\boldsymbol{b}$ of $\pm 1/3[11\overline{2}3]$ or $\pm 1/3[11\overline{23}]$, and the weak contrast is due to $\boldsymbol{g}\cdot(\boldsymbol{b}\times\boldsymbol{u}) \neq 0$. Because the thickness fringes connected with this defect type are not clear in the images, a more precise value for $\boldsymbol{b}$ cannot be determined. It is also noted that some dislocations have opposite values of $\boldsymbol{b}$ to others, such as those denoted as $\boldsymbol{M}$ and $\boldsymbol{N}$ in Fig. 17.17a. These two opposite types of dislocations delineate a subgrain, and the region showing bright contrast when imaged with $\boldsymbol{g} = 1\overline{1}00$ in Fig. 17.17b is a misoriented subgrain, tilted close to the Bragg condition.

17.4.2 Microdiffraction and Polarity

If an electron beam is converged to form a focused probe at the sample, then there will be a range of beam directions within the incident probe and within the transmitted and diffracted beams. This leads to the formation of diffraction patterns comprising discs rather than spots. The fine detail within such microdiffraction or CBED patterns contains space and point group information. Such patterns also provide a sensitive measure of lattice parameters and strain within a sample, and can be used also to assess defect type.

Two examples showing how focused probe diffraction patterns can be used to determine the absolute polarities of sphalerite and wurtzite noncentrosymmetric crystals, in situ in the TEM, are now illustrated. It is often important to know the polar orientation of a heterostructure since this can strongly influence the mode and rate of growth, the incorporation of dopants or impurities, and the development of extended defects, and hence the functional performance of the resulting device structure. The opposite polar faces of noncentrosymmetric crystals [17.27] may be distinguished in practice using appropriate chemical etchants. However, such reagents are discriminatory in their action and need to be correlated with some other experimental technique for the purpose of absolute polarity deter-

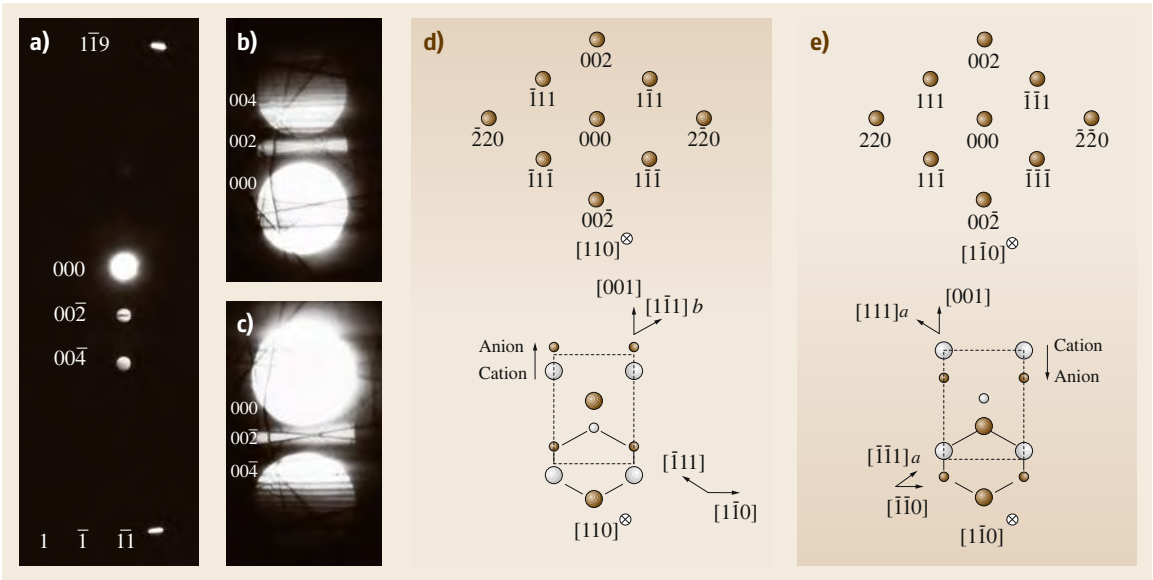


Fig. 17.18 **(a–c)** Microdiffraction patterns from [110] GaAs. **(a)** Destructive interference through an $00\bar{2}$ diffraction disc due to the interaction of doubly diffracted $1\bar{1}9$ and $1\bar{1}\,\bar{1}1$ beams. **(b,c)** Constructive and destructive interference through 002 and $00\bar{2}$ diffraction discs. **(d,e)** Sphalerite space lattice and reciprocal lattice projections for [110] and $[1\bar{1}0]$, indexed for an electron beam traveling down into the page

mination. For most cases of diffraction, there is no difference between the intensities of beams scattered by *hkl* and $\overline{hkl}$ diffracting planes. This indeterminacy is known as Friedel's law. However, the technique of microdiffraction coupled with a breakdown in Friedel's law allows the absolute polarities of sphalerite crystals to be determined when the anion and cation sizes are very similar (as with GaAs, ZnSe and CdTe) [17.28]. The diffraction condition shown in Fig. 17.18a corresponds to a projection tilted $\approx 10°$ off a GaAs [110] zone axis, arrived at following an $\approx 00\bar{4}$ Kikuchi band. The interaction of the doubly diffracted high-order, odd-index $1\bar{1}9$ and $1\,\bar{1}\,\bar{1}1$ beams with directly scattered $00\bar{2}$ reflection gives rise to destructive interference through the $00\bar{2}$ disc when all of the reflections are close to the Bragg position. Figure 17.18c, corresponding to a more highly converged incident electron beam, emphasizes the formation of the dark cross through the $00\bar{2}$ diffraction disc. Similarly, Fig. 17.18b shows the bright cross formed through the opposite 002 diffraction disc, by constructive interference following interaction with the corresponding $1\,\bar{1}\,11$ and $1\bar{1}9$ high-order, odd-index reflections. Figure 17.18d,e, showing the [110] and $[1\bar{1}0]$ projections of the sphalerite space lattice and reciprocal lattice, respectively, are given, to clarify this particular sample geometry. Establishing such microdiffraction patterns requires carefully balancing crystal orientation, probe convergence and layer thickness, and the technique is found to work best with freshly plasma-cleaned sample foils to minimize the effect of extraneous hydrocarbon contamination (which diffuses out fine-scale contrast under the imaging electron beam).

The dynamical equations of electron diffraction demonstrate that the 002 reflection, which exhibits constructive interference effects (bright cross), always occurs in the sense of the cation to anion bond (Fig. 17.18d), and hence is always associated with the sense of advancing $\{\overline{1}\overline{1}1\}_b$ planes. To emphasise this, defining [001] as the growth direction of an epilayer on an (001) oriented GaAs substrate, and establishing a bright cross through an 002 diffraction disc, for example through the interaction of $1\bar{1}9$ and $1\,\bar{1}\,\bar{1}1$ reflections, corresponds to the sense of advancing $(\bar{1}11)_b$ and $(1\bar{1}1)_b$ planes in the growth direction, which corresponds to the absolute [110] projection of the sample foil. Conversely, a dark cross through an 002 diffraction disc, such as that obtained through the interaction of $11\bar{9}$ and $1\;1\;11$ reflections, corresponds to the sense of advancing $(111)_a$ and $(\bar{1}\bar{1}1)_a$ planes in the growth direction, which corresponds to the $[1\bar{1}0]$ projection of the sample foil. Since microdiffraction patterns are directly sensitive to crystallographic polarity, a qualitative interpretation of such results allows the crystal orientation to be determined unambiguously, in situ within the TEM, without reference to any other technique.

Contrast reversals within diffraction discs of systematic row CBED patterns can similarly be used to

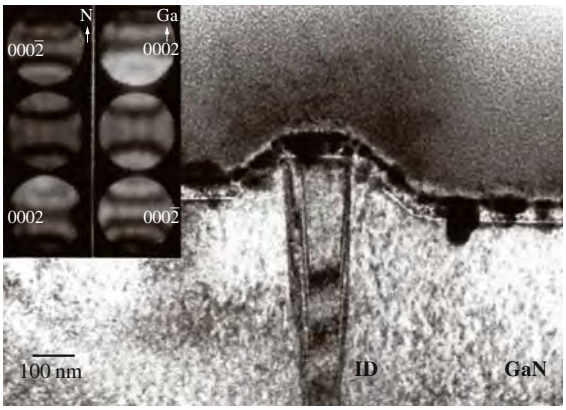


Fig. 17.19 Inversion domain within GaN with CBED patterns recorded across the boundary plane (*inset*)

determine absolute crystal polarity, since Friedel's law again breaks down due to dynamical scattering. In practice, this approach is found to be most effective when used with noncentrosymmetric crystals that have much larger differences in anion and cation sizes, such as wurtzite GaN [17.29]. Figure 17.19 presents a thin section through a columnar defect, imaged within homoepitaxial GaN grown on an N-polar $(000\bar{1})$ GaN substrate, with two associated CBED patterns inset, recorded either side of the boundary plane corresponding to matrix and core material, respectively. Simulation of the contrast within the 0002 diffraction discs, for a known sample foil thickness, demonstrates that the central bright and dark bands correspond to N and Ga-polar growth directions, respectively. The reversal of contrast within the 0002 diffraction discs across the boundary plane therefore indicates an inversion in crystal polarity, confirming the defect to be an inversion domain in this instance.

17.4.3 Reflection High-Energy Electron Diffraction

Electron-transparent samples are required for TED investigations. For the case of semiconductor heterostructures, sample preparation requires sequential mechanical polishing and ion beam thinning, which can be very time-consuming, but the advantage of this approach is that diffraction data can be directly correlated with projected images of the internal structure of the sample. A complementary approach is to use glancing angle electron diffraction techniques to characterize the near-surface microstructure of a bulk sample or an as-grown heterostructure. Coupling of the electron beam directly with the wafer surface at low angle allows scattering of the electrons to occur, producing a diffraction pattern that may be viewed or recorded directly. This provides valuable information on the near-surface crystallography of the sample, which can be correlated with the growth conditions used or surface modifications applied, without the need for time-consuming sample preparation.

There are three variants of reflection electron diffraction (RED) depending on the accelerating voltage available:

- Low (< 1 kV)
- Medium (1–20 kV)
- High (20–200 kV) energy.

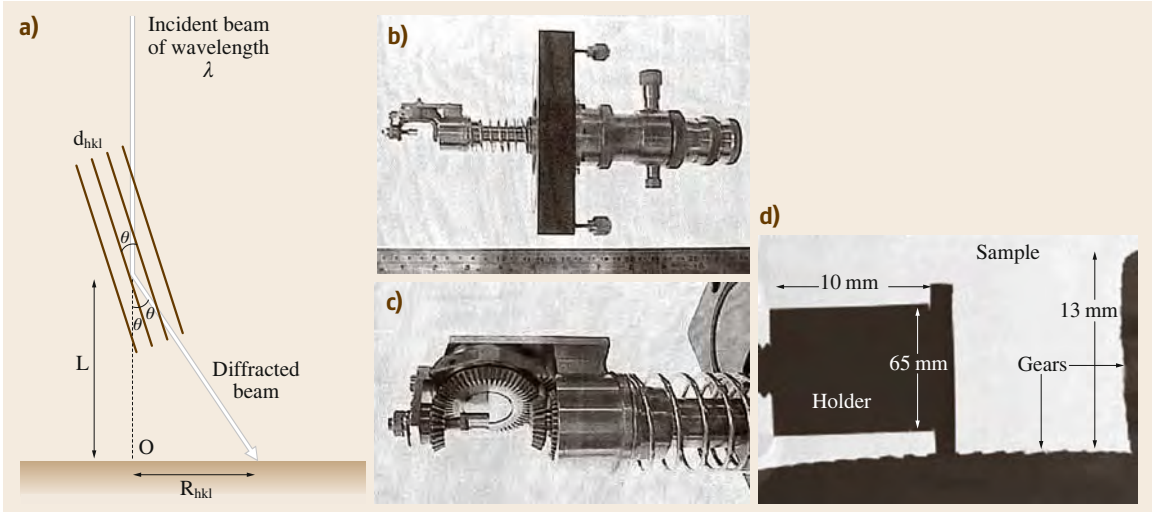


Fig. 17.20 **(a)** Schematic illustration of the RHEED diffraction geometry; **(b)** RHEED stage; **(c)** magnified view of the sample holder; and **(d)** projected shadow image of a specimen on TEM phosphor screen

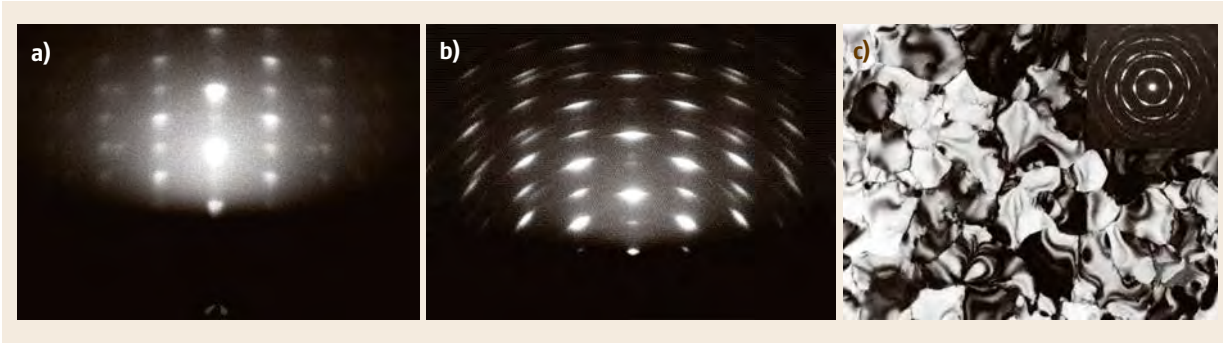


Fig. 17.21a–c RHEED patterns from Si-doped GaN/GaAs(001) **(a)** before and **(b)** after plasma cleaning. **(c)** Plan-view TEM image confirming the presence of rotated columnar grains (*TED pattern inset*)

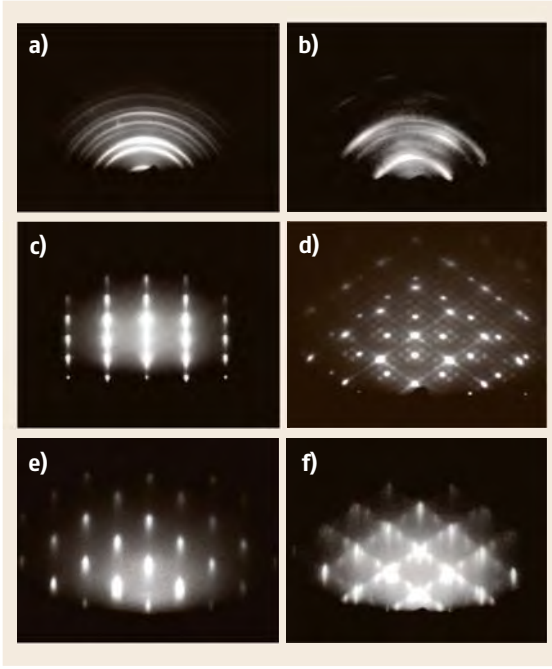


Fig. 17.22a–f RHEED patterns from **(a)** GaAs/AlN(As)/GaP(001) indicating polycrystalline growth; **(b)** GaN(As)/GaP(001) grown under low As flux at 620 °C showing disordered growth with some degree of preferred orientation; **(c)** Mg-doped GaN/GaAs $\{\overline{1}11\}$B showing α-GaN single-crystal growth ($\langle 11\overline{2}0\rangle$ projection); **(d)** GaN(As)/GaP (001) grown under high As flux at 650 °C showing spots due to single-crystal GaAs and β-GaN; **(e,f)** Be-doped GaN/GaAs (001) grown at 700 °C showing orthogonal $\langle 110\rangle$ and $\langle 1\overline{1}0\rangle$ projections. *Extra spots* indicate a high degree of anisotropy for the distribution of microtwin defects within this sample

Medium–energy electron diffraction (MEED) systems are commonly associated with UHV growth chambers, whilst a variant of RHEED can be performed using a conventional TEM. In general terms, the coupling of an electron beam with a flat single crystal tends to be associated with the formation of diffraction streaks normal to the surface, providing information on the reconstruction of atomic layers at the surface. More precisely, streaky RED patterns are indicative of a surface that is not quite perfectly flat, but has slight local misorientations combined with some degree of surface disorder [17.30]. When electrons are coupled with a slightly rougher surface, there is a tendency for more three-dimensional information to be obtained from interactions with the near-surface microstructure. This leads to the production of spotty diffraction patterns from crystalline materials (half-obscured by the sample shadow edge), in an analogous fashion to TED. It is noted that UHV-MEED systems allow for intensity fluctuations within the central beam to be monitored in real time, providing a way to monitor and control layer-by-layer growth.

In practice, RHEED is used most effectively for rapid comparative studies of samples sets; for appraising the effect of changing the growth parameters on the structural integrity of deposited thin films [17.31]. This approach to process mapping provides a convenient way to identify appropriate samples for more detailed TEM investigation prior to sample foil preparation. There are two variants of the RHEED technique that may be used in a TEM: one where small samples are mounted vertically in the objective lens pole piece, and the other where larger samples are mounted vertically below the projector lens; we focus on the latter variant here.

A schematic illustration of the diffraction geometry for a RHEED experiment is shown in Fig. 17.20a. The diffraction spacing R_{hkl} can be measured and the associated crystal plane spacing d_{hkl} determined using the equation, $\lambda L = R_{hkl}d_{hkl}$, where λL is known as the camera constant. Figure 17.20b,c shows a RHEED stage made to interface with a TEM through an existing cam-

era port at the base of the projector lens. The vacuum system of the TEM vents the space below the projector lens upon opening the camera chamber, and so the only waiting time is for the chamber to pump down upon changing a specimen. The RHEED stage is able to support centimeter-square sections of a crystalline specimen, whilst full tilt, twist and lateral movements enable any zone axis within the growth plane of the sample to be accessed. A shadow image of a sample, as projected onto the microscope phosphor screen, is shown in Fig. 17.20d. The area sampled by the glancing electron beam is typically $\approx 1\,\mathrm{mm}^2$, and so this experimental arrangement is closely associated with glancing angle XRD, although the acquisition time for RHEED data is extremely short, with photographic plate exposure times of ≈ 1 s. Three or four samples can typically be examined within 1 h.

The 100 kV RHEED patterns shown in Fig. 17.21a,b were acquired from a highly Si-doped GaN/GaAs(001) heterostructure, grown by MBE at 700 °C, before and after plasma cleaning. Surface hydrocarbon deposits arising from specimen handling can generate an amorphous background glow that can hinder RHEED pattern acquisition (Fig. 17.21a). The oxygen-argon plasma acts to remove such contaminants, allowing high-contrast RHEED patterns to be obtained (Fig. 17.21b). In this instance, the generated RHEED pattern remained effectively constant as the sample was rotated, indicating a random distribution of columnar grains. This predicted microstructure was subsequently confirmed by means of conventional plan-view TEM imaging (Fig. 17.21c, with TED pattern inset), which revealed a fine-scale distribution of rotated columnar grains ($\approx 20°$ of arc).

The RHEED patterns shown in Fig. 17.22a–f, acquired from a variety of III–V heterostructures grown by MBE, are presented to illustrate the variety of microstructures that can be readily distinguished. In the first instance, variations in spot, arc or ring spacings from the central beam provide evidence of the different phases present within a sample and show the general nature of the structural integrity of the near-surface layer (presumed to be representative of the bulk or thin film). Thus, polycrystalline, preferred orientation or single-crystal growth (Fig. 17.22a–c) can be rapidly distinguished, whilst embedded phases (Fig. 17.22d) and anisotropic defect distributions within zinc blende thin films (Fig. 17.22e,f) are also revealed.

17.5 Characterizing Functional Activity

There are many solid state analytical techniques that employ x-ray or electron probes, generating a variety of signals for chemical microanalysis. However, techniques for performing correlated assessment of the structural and functional performance of a material are perhaps less well-covered in mainstream texts. Accordingly, we introduce briefly the techniques of scanning transmission electron beam induced conductivity (STEBIC) and TEM-cathodoluminescence (TEM-CL), since these allow us to make correlated structure-property investigations of electrical and optical activity within a semiconductor, respectively.

When an electron beam is incident on a semiconductor specimen, electron-hole pairs are created by excitation of crystal electrons across the band-gap. These electron-hole pairs can, for example, recombine to emit light that may be detected by a photomultiplier providing, e.g., information on the semiconductor band-gap. A CL image can then be obtained by displaying the detected photomultiplier signal as a function of the position of the incident electron beam as it is scanned across the specimen. CL spectra can also be acquired in spot mode, which show features attributable to excitons, donor–acceptor pairs or impurities. The *information content* of CL images and spectra includes the location of recombination sites such as dislocations and precipitates, and the presence of doping-level inhomogeneities. Similarly, if the sample is configured to incorporate a collection junction, such as a Schottky-contacted semiconductor or an ohmic-contacted p–n junction, electron-hole pairs that sweep across the built-in electric field constitute a current flow. This can be amplified and an image of the recombination activity displayed as the electron beam is rastered across the sample. If the dislocations within a semiconductor act as nonradiative recombination centers, then they appear as dark lines in both CL and EBIC images because of the reduced specimen luminescence or reduced current that is able to flow through the collection junction when the beam is incident at a defect.

The techniques of CL and EBIC are most commonly performed in an SEM, but this precludes the direct identification of features responsible for a given optical or electronic signature. The resolution of extended defects achieved using EBIC and CL techniques is limited by the penetration depth of the electron beam, the effect of beam spreading and the diffusion length of minority carriers. Conversely, the resolutions of the STEBIC and TEM-CL techniques, as applied to

electron-transparent sample foils, are essentially limited by specimen geometry. The constraint of minority carrier diffusion length is removed due to the close proximity of the sample foil surfaces, and resolution depends on the incident probe size, the width of the electron hole pair generation zone and recombination velocity at the free surface. For the case of STEBIC, resolution also depends on defect position relative to the collecting junction. The trade-off is low electrical signal and a degraded signal-to-noise ratio due to the small generation volume and surface recombination effects, in addition to the practicality of contacting and handling thin foils.

Before presenting some material characterization case studies based on electron beam techniques, we now discuss the preparation of electron transparent foils that are free from artefacts and suitable for TEM investigation.

17.6 Sample Preparation

We should initially consider whether destructive or nondestructive preparative techniques need to be applied. Some characterization techniques allow samples to be examined with a minimal amount of preparation, provided they are of a form and size that will fit within the apparatus. For example, the crystallography of bulk or powder samples may be investigated directly by XRD, since the penetration depth of energetic x-rays within a sample is on the scale of $\approx 100\,\mu m$. The surface morphology and near-surface bulk chemistry of a sample may be investigated directly within the SEM, noting the interaction volume of electrons [on the scale of $\approx 1\,(\mu m)^3$] associated with the EDX and WDX techniques. It might, however, be necessary to coat insulating samples with a thin layer of carbon or gold prior to SEM investigation to avoid charging effects. Similarly, minimal preparation might be required before surface assessment using XPS or RHEED, such as cleaning using a degreasing protocol or plasma cleaning. Accordingly, the focus of this section is to introduce techniques used to prepare samples for TEM investigation, since the requirement is for specimens that are typically submicrometer in thickness and free of preparation artefacts. For example, a complex sequence of sequential mechanical polishing, dimpling, ion beam thinning and plasma cleaning may be required to produce a pristine semiconductor heterostructure sample, with each stage of the process

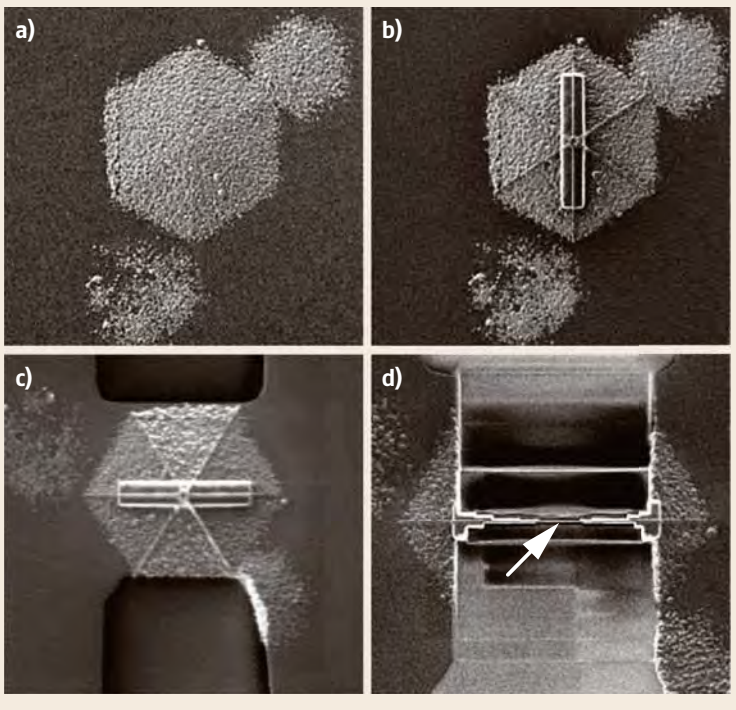


Fig. 17.23a–d SE images showing the application of FIB milling to target a source of defects buried beneath the emergent core of a growth hillock. **(a)** An etched CVD-grown GaN/sapphire hillock ($\approx 5\,\mu m$ in size; CVD – chemical vapor deposition). **(b)** Platimum stripes deposited along the hillock facet edges in order to retain sight of the defect core ($\approx 100\,nm$ in size). **(c)** A high Ga flux is used to create an access trench to the defect. **(d)** Decreasing the Ga flux enables an electron-transparent membrane (*arrowed*) to be defined at the position of the defect core

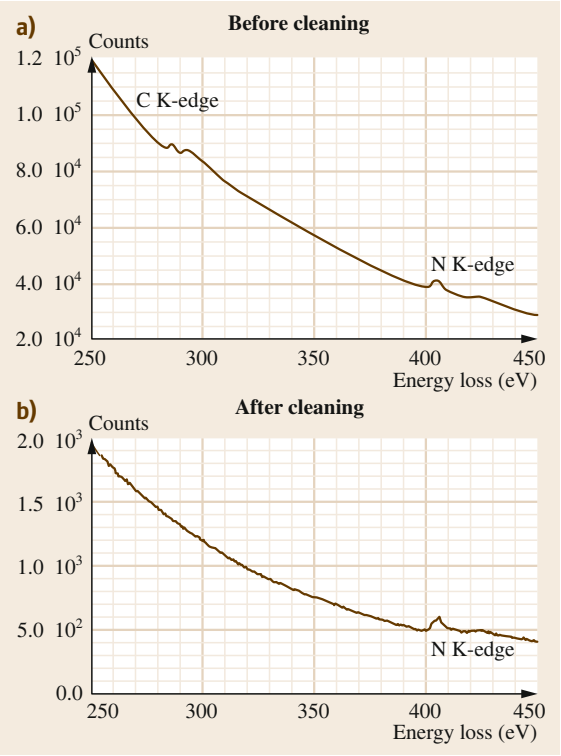


Fig. 17.24a,b EELS spectra from GaN (a) before and (b) after plasma cleaning

being designed to minimize artefacts from the previous stage of the preparation process. The idea being to minimize or eliminate artefacts from the preparation process, to ensure that the sample being investigated is representative of the starting bulk material. Care is also needed to avoid artefacts that might be introduced through interaction of the high-energy electron beam with the sample.

In this context, it is interesting to note how TEM sample preparation techniques have developed over the years. Small particles of MgO, produced by igniting the metal and drifting a specimen grid through the smoke, were typical of samples investigated in the 1940s, along with sample replicas made by a dry stripping technique using formvar films. Biological samples fashioned by enzymatic digestion, staining and microincineration were also possible by 1945. Glass and diamond knife microtomes were introduced in the 1950s and used to section soft biological materials. Advances in the controlled preparation of inorganic materials were made upon the introduction of argon ion beam thinning in the late 1960s, which allowed for cross-sectional observation of semiconductor heterostructures when combined with sequential mechanical polishing and dimpling. Significant development work in this area appears throughout the literature from the 1970s. The problem of surface amorphization, introduced by the argon sputtering process, was minimized by adopting low-voltage milling techniques to define the final electron-transparent sample foil. Low stacking fault energy semiconductors, such as II–VI compounds which are easily damaged or InP-based compounds that suffer from In droplet formation with conventional milling techniques, were also successfully prepared for TEM observation using the technique of iodine reactive ion beam etching (RIBE), otherwise known as chemically assisted ion beam etching (CAIBE), developed in the 1980s. The 1990s, however, saw the development of the most effective raft of sample preparation techniques for functional materials and complex semiconductor device structures in particular: i. e., tripod polishing, focused ion beam (FIB) milling and plasma cleaning.

The fine adjustment of micrometer supports is the key to tripod polishing that allows for direct mechanical polishing of specimens down to $< 10\,\mu m$ thickness, using diamond-impregnated polishing cloths on a stable, high-torque, low-speed polishing wheel. A brief final stage of low-voltage argon ion beam thinning (with liquid nitrogen cooling) then enables electron-transparent sample foils to be defined, free of differential mechanical polishing and shadowing artefacts, whilst minimizing remnant surface amorphization.

FIB instrumentation (now commonly integrated with SEM) was originally developed for the semiconductor industry as a diagnostic tool for silicon chip fabrication. The application of a focused beam of gallium ions enables sample material to be sputtered away in a controlled fashion to produce an electron-transparent membrane, for example through a specific device within a complex microprocessor. Samples may be periodically observed using SEs generated by gallium ion/material interactions in order to maintain control over and the precision of the sputtering process. In particular, the ability of FIB instruments to prepare site-specific TEM membranes provides a unique opportunity to access the subsurface microstructure of complex device structures or specific regions within a heterostructure identified as being particularly interesting from observations of the surface. By way of example, Fig. 17.23 shows a TEM membrane sectioned through the apex of a hillock identified within a sample of heteroepitaxial GaN–sapphire. Prior to sectioning, the sample was coated with a thin layer of gold and a platinum alkyl decomposed under the rastered Ga beam to define protective metallic stripes. Sequential sputtering and reduction of the incident Ga beam current density allowed a supported thin membrane to be defined (Fig. 17.23d). Care is required to minimize amorphization artefacts at the surfaces of

such membranes, arising from the sputtering action of the glancing, high-energy Ga ion beam. Further, charging effects associated with insulating samples can act to compromise the fine-scale control of the ion milling procedure. However, this particular problem can be addressed by using low ion dose and shadowing techniques, combined with charge neutralization procedures to inhibit the deflection of the incident ion beam.

Plasma cleaning using oxygen-argon gas, commonly used for the final stage of TEM sample preparation, enables pristine electron-transparent membranes to be obtained, suitable for detailed chemical microanalysis. The disassociated oxygen component of the plasma reacts with organic surface contaminants to produce CO, CO_2 and H_2O reaction products that can be pumped away conveniently. To illustrate the effectiveness of this procedure, it was found that the quality of EELS data from GaN electron transparent foils was significantly improved following plasma cleaning. Figure 17.24 shows the removal of the artefact carbon *K*-edge and enhancement of the sample nitrogen *K*-edge after a few minutes of exposure to the plasma. It is worth noting that in principle the argon component of the plasma permits gentle sputtering of the sample to occur if low-pressure conditions are used due to the larger mean free path and hence increased energy of the ions. However, this can be problematic due to the possibility of sample cross-contamination with material sputtered from the supporting sample stage. Another cautionary note on plasma cleaning relates to semi-insulating samples that can become too clean and consequently more susceptible to charging effects under the imaging electron beam.

17.7 Case Studies – Complementary Characterization of Electronic and Optoelectronic Materials

There are four general levels of interest when characterizing a given sample:

- What is it made of?
- What additional imperfections does it contain?
- How did it get to be that way?
- How does the microstructure influence the functional properties of the material/device?

For example, one might wish to identify the chemical constituents and crystal structure of a sample to start with, such as whether it is a compound or an alloy, and whether it is single-crystal, polycrystalline, exhibits preferred orientation, or is amorphous. One might then wish to characterize the additional fine-scale defect microstructure within the sample, including precipitates or extended structural defects such as dislocations, stacking faults or domain boundaries, since these could deleteriously affect the functional properties of the material. The next level of understanding focuses on how the material has formed via growth, processing or device usage, and seeks to make sense of the process of dynamic evolution, since this potentially enables us to find ways to improve the material in a controlled manner. A clear record of sample history is generally helpful in this context, particularly when trying to identify defect sources. The final level of understanding is possibly the most challenging, since it seeks to associate microstructure with the functional properties exhibited by the material, and thereby to make sense of its process–structure–property interrelationship at the fundamental level.

In this context, the structural characterization of semiconductor heterostructures tends to focus on the following issues:

- The integrity of the layer growth and the orientation relationship with the substrate.
- The nature of the structural defects within the epilayer, arising, e.g., from lattice mismatch or from differential thermal contraction following cool-down from the growth temperature.
- The structural integrities of critical interfaces within the device's active region, such as multiple quantum wells, and the chemical homogeneity of the associated alloy layers.
- Modification of the microstructure due to subsequent processing, such as contact formation and device usage.

XRD techniques are well-suited to assessments of the structural integrities of bulk and epitaxial thin film semiconductors. However, we should note the inherent averaging over a large number of microscopic features, such as dislocation distributions responsible for the twist and tilt of mosaic grains, associated with such techniques. Conversely, TEM and related techniques are more suited to assessing the fine-scale defect microstructure of heterostructures. For example, conventional weak-beam, HREM or CBED analysis can be used for fine-scale defect structural analysis, whilst EELS and EDX analysis can be used to profile composition. However, the ability to perform atomic-level structural characterization and chemical analysis on

the nanometer scale is offset by concerns about statistical significance and whether the small volume of material analyzed is truly representative of the larger object. Hence, electron microscopy-based techniques combined with FIB procedures for site-specific sample preparation tend to be favored when investigating integrated device structures.

The examples provided so far illustrate how various diffraction and imaging techniques can provide information on the structural integrity of a given sample. The following examples emphasize the need to apply complementary material characterization techniques in support of the development of semiconductor science and technology.

17.7.1 Identifying Defect Sources Within Homoepitaxial GaN

The emergence of the (In,Ga,Al)N system for short-wavelength light-emitting diodes, laser diodes and high-power field effect transistors has been *the* semiconductor success story of recent years. In parallel with the rapid commercialization of this technology, nitride-based semiconductors continue to provide fascinating problems to be solved for future technological development. In this context, a study of homoepitaxial GaN, at one time of potential interest for high-power blue–UV lasers, is presented.

The reduction in extended microstructural defects permitted by homoepitaxial growth was considered to be beneficial in the development of nitride-based technology, particularly in view of evidence confirming that dislocations do indeed exhibit nonradiative recombinative properties. However, in the case of metal organic chemical vapor deposition (MOCVD)-grown homoepitaxial GaN on chemomechanically polished $(000\bar{1})$, N-polar substrates, gross hexagonally shaped surface hillocks were found to develop, considered problematic for subsequent device processing [17.32]. The homoepitaxial GaN samples examined in this case study were grown at 1050 °C. The bulk GaN substrate material was grown under a high hydrostatic pressure of nitrogen (15–20 kbar) from liquid Ga at 1600 °C. Prior to growth, the $(000\bar{1})$ surfaces were mechanically polished using 0.1 μm diamond paste and then chemomechanically polished in aqueous KOH solution. Epitaxial growth was performed using trimethylgallium and NH_3 precursors with H_2 as the carrier gas, under a total pressure of 50 mbar. Figure 17.25 shows an optical micrograph of the resultant homoepitaxial GaN/GaN$(000\bar{1})$ growth hillocks, typically 5–50 μm in size depending on the layer thickness (and therefore the time of growth).

Electron-transparent samples were prepared in plan view using conventional sequential mechanical polishing and argon ion beam thinning procedures applied from the substrate side, whilst cross-sectional samples were prepared using a Ga-source FIB workstation. As shown earlier, the selectivity of the FIB technique enables cross-sections through emergent cores of the hillocks to be obtained, thereby allowing nucleation events associated with these features to be isolated and characterized. When prepared in plan-view geometry for TEM observation, each hillock exhibited a small faceted core structure at the center (Fig. 17.25b), but otherwise the layers were generally found to be defect-free. Low-magnification cross-sectional TEM imaging also revealed the presence of faceted column-shaped defects beneath the apices of these growth hillocks (Fig. 17.25c). It was presumed that these features originated at the original epilayer-substrate interface since no other contrast delineating the region of this homoepitaxial interface could be discerned. A reversal of contrast within 0002 diffraction discs from CBED patterns acquired across the boundary walls of such features (Fig. 17.19) confirmed that they were inversion domains. Thus, the defect cores were identified as having Ga-polar growth surfaces embedded within an N-polar GaN matrix. Once nucleated, the inversion domains exhibited a much higher growth rate than the surrounding matrix, being directly responsible for the development of the *circus tent* hillock structures around them. Competition between growth and desorption rates of Ga and N-polar surfaces allowed the gross hexagonal pyramids to evolve.

This initial approach of applying electron diffraction and imaging techniques thus enabled the nature of the inversion domains to be identified and their propagation mechanism established in order to explain the development of the hillocks. However, more detailed chemical analysis was required to ascertain the nature of the source of the inversion domains and how this related to substrate preparation and the growth process. A high-angle annular dark field (HAADF) image of the inversion domain nucleation event is shown in Fig. 17.25d. HAADF is a scanned electron probe imaging technique with a resolution defined by the size of the incident probe, with the scattering (and hence contrast) governed by local average atomic number. In this instance, the sample, tilted slightly to minimize the effects of diffraction contrast, showed dark contrast at the position of the inversion domain source, confirming the presence of a low atomic number material associated with the nucleation event. HREM subsequently confirmed that such nucleation events were

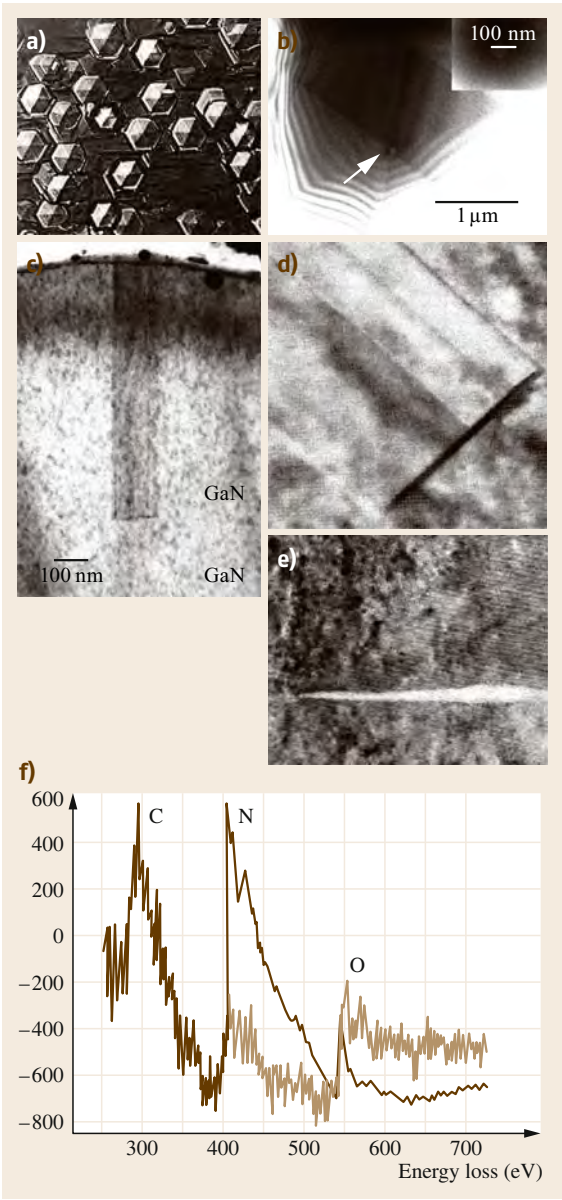


Fig. 17.25 **(a)** Optical micrograph of MOCVD-grown GaN on a KOH chemomechanically polished, N-polar, bulk GaN substrate showing a high density of hillocks. **(b)** Slightly tilted plan-view TEM image through a growth hillock revealing a central defect core (*arrowed*). An enlarged ⟨0001⟩ projected image of the hillock core is shown in the *inset*. **(c)** Cross-sectional, weak-beam TEM image through a hillock core. CBED analysis confirmed the feature to be an inversion domain. **(d)** HAADF image indicating the presence of a low atomic number material at the inversion domain source. **(e)** HREM image indicating the presence of a narrow band of amorphous material at the inversion domain source. **(f)** EEL second difference spectrum indicating the presence of oxygen at the nucleating event ◄

due to narrow bands of amorphous material, 2–5 nm in thickness (Fig. 17.25e), whilst EELS confirmed the presence of oxygen (Fig. 17.25f) within these narrow amorphous bands. Accordingly, these defect sources were attributed to remnant contamination from the chemomechanical polishing technique used to prepare the substrates prior to growth. The oxygen-containing residue was presumed to be gallium oxide or hydroxide – being products of the reaction of the KOH etchant with GaN. An improved surface preparation method incorporating a short, final deoxidizing polishing procedure in an aqueous solution of NaCl led to a dramatic reduction in these nucleation sources and thus allowed N-polar homoepitaxial GaN films to be grown virtually free of these hillock structures.

17.7.2 Cathodoluminescence/Correlated TEM Investigation of Epitaxial GaN

The CL technique is ideally suited to studies of luminescence uniformity and spectral purity. The following case study illustrates how the defect microstructure of mixed-phase epitaxial GaN/GaAs($\overline{111}$)B can be correlated with the layer of luminescent properties [17.33].

Even though the majority of developments in GaN technology to date have come from material grown by MOCVD on sapphire and SiC substrates, there is much interest in exploring alternative growth techniques. The MBE technique offers a lower growth temperature than MOCVD and hence enables a greater range of candidate substrate materials to be investigated. One general issue for the MBE growth of heteroepitaxial GaN is the need for direct control of the process of nucleation, since this impacts on the phase and polarity of the deposit and the resultant structural integrity of the film. Epitaxial GaN preferentially adopts the wurtzite phase, with a band-gap of 3.4 eV_{hex}, although zincblende inclusions, with a band-gap of 3.2 eV_{cubic}, are sometimes associated with MBE-grown material (for conditions of high Ga flux); Fig. 17.26. The CL spectrum shown in Fig. 17.26a was recorded from a plan-view, electron-transparent foil of nominally single-crystal wurtzite GaN, cooled to liquid nitrogen temperature. The peak in the CL spectrum at 386 nm (3.21 eV) was used to create the image in Fig. 17.26b indicating the distribution of cubic phase inclusions throughout the hexagonal GaN matrix. The complementary conventional TEM image of this plan-view sample foil (Fig. 17.26c) allowed the nature of the fine-scale microstructure to be

characterized, with diffraction patterns confirming the presence of embedded sphalerite GaN growth. The size and distribution of these inclusions correlated nicely with the distribution of bright spots in the CL image of Fig. 17.26b. A subsequent cross-sectional investigation of this GaN/GaAs($\overline{1}\overline{1}\overline{1}$)B specimen confirmed that cubic inclusions nucleated at the epilayer/substrate interface (Fig. 17.26d).

17.7.3 Scanning Transmission Electron Beam Induced Conductivity of $Si/Si_{1-x}Ge_x/Si(001)$

The STEBIC technique was originally demonstrated in the late 1970s, using dedicated STEM instrumentation to obtain information on the electrical properties of dislocation core structures within (Ga,Al)(As,P), thereby providing the first evidence that nonradiative recombination processes at dislocations are related to jogs and kink sites. Dissociated 60° dislocations showed the highest electrical activity, while sessile Lomer–Cottrell edge dislocations were found to be electrically neutral, indicative of reconstructed core structures. STEBIC imaging of an electron-transparent foil allows the electrical and structural properties of defects to be observed simultaneously. The availability of electron sources with high brightness in modern scanning TEM instruments compensates for the main problem of small generation volume and provides an accessible way to perform STEBIC experiments.

The next case study illustrates how the electrical nonradiative recombination properties of MBE-grown $Si/Si_{1-x}Ge_x/Si(001)$ heterostructures correlate with the distribution of interfacial misfit dislocations [17.34]. The $Si_{1-x}Ge_x$ system has potential application in devices with high electron and hole mobilities. However, the introduction of dislocation networks, or the multiplication of existing dislocations, driven by the 4% misfit strain between Ge and Si, is generally regarded as being detrimental to device operation. With a view to gaining an improved understanding of the relationship between fine-scale structural defects and electron transport properties, the structures of capped MBE-grown $Si/Si_{1-x}Ge_x/Si(001)$ were investigated using STEBIC and a range of complementary microscopies. The samples incorporated buried p–n junctions to assist with charge collection and surmount the problem of surface recombination effects.

To prepare for the STEBIC investigation of the $Si/Si_{1-x}Ge_x/Si(001)$, evaporated Al contacts were attached to the top surface prior to preparing electron-transparent foils in plan-view by sequential mechanical polishing and argon ion milling of the substrate (Fig. 17.27a). Ohmic contact to the lower surface was made using an InGa eutectic. Electrical activity images were acquired using an electrical contact stage and a scanning TEM. Signal amplification was performed using an amplifier with low noise current. By controlling the STEM scan rate, the beam could be rastered at a rate compatible with the low bandwidth constraint of the amplifier. STEBIC signals were typically $\approx 100\,\mathrm{pA}$ for an electron-transparent $Si/Si_{1-x}Ge_x/Si$ foil imaged at a magnification of $\times 1000$.

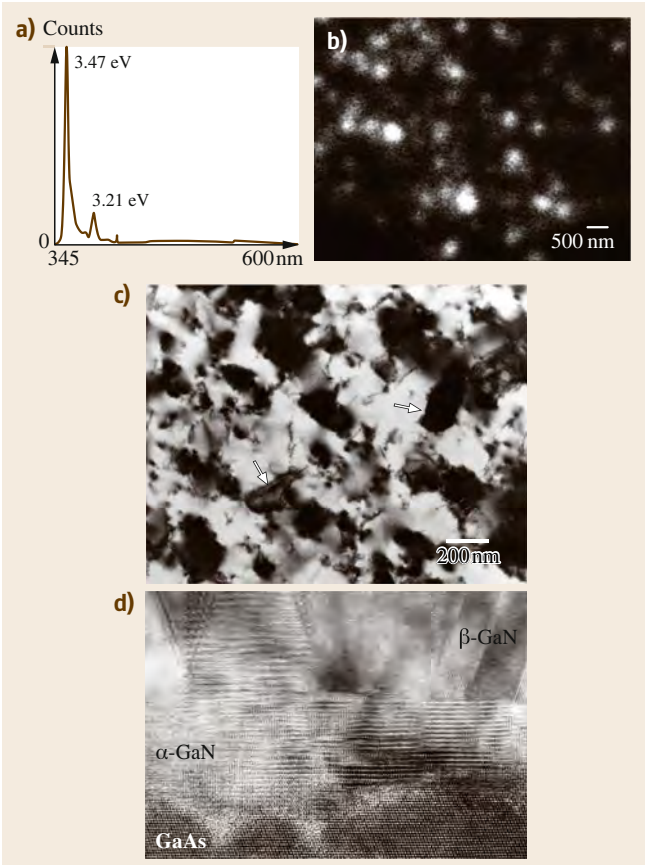


Fig. 17.26 **(a)** CL spectrum acquired from an electron-transparent foil showing peaks at 357.8 nm and 386 nm, corresponding to wurtzite and zinc blende GaN, respectively; **(b)** CL image formed at 386 nm and **(c)** complementary TEM image confirming a distribution of cubic GaN inclusions (*arrowed*) embedded within the hexagonal GaN matrix; **(d)** HREM image of epitaxial GaN/GaAs ($\overline{1}\overline{1}\overline{1}$)B viewed in cross-section, indicating the nucleation of cubic phase inclusions at the epilayer-substrate interface

Figure 17.27a illustrates the physical parameters relevant to this STEBIC experiment. The silicon substrate was n-type $10^{18}\,\mathrm{cm}^{-3}$, whilst the epilayer was p-type boron-doped to a level of $10^{16}\,\mathrm{cm}^{-3}$. These values of sample doping were chosen to create a depletion region of $\approx 30\,\mathrm{nm}$ width to assist with charge

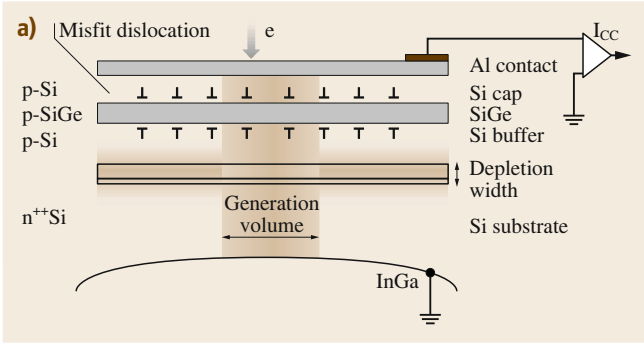


Fig. 17.27 **(a)** Schematic of the Al/Si/$Si_{1-x}Ge_x$/Si(001)/InGa sample configuration used for STEBIC investigation

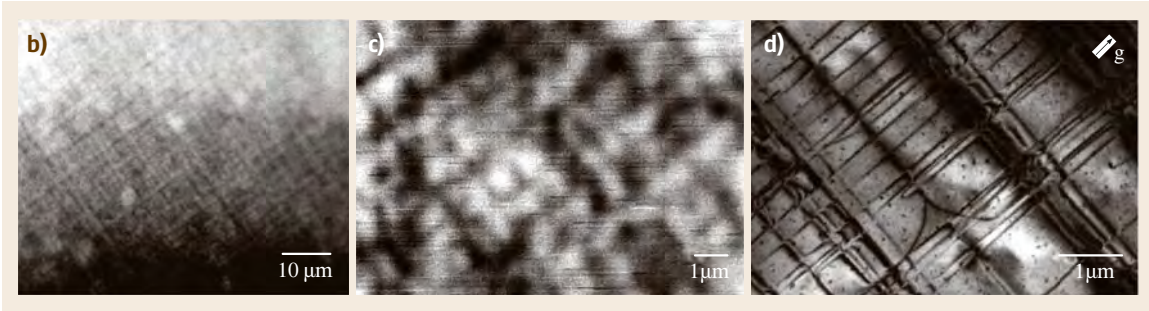


Fig. 17.27 **(b)** STEBIC image showing recombination activity due to ⟨110⟩ orthogonal arrays of misfit dislocations within relaxed Si/$Si_{0.96}Ge_{0.04}$/Si (001). **(c)** Higher magnification STEBIC image illustrating sub-µm resolution of the electrical activity. **(d)** Bright-field TEM image demonstrating the presence of bundles of misfit dislocations ($\boldsymbol{g} = 220$)

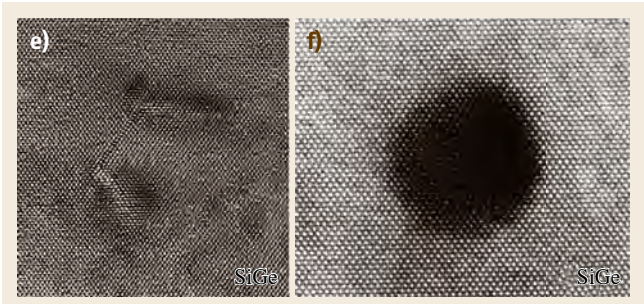


Fig. 17.27 **(e)** HREM image of a decorated dissociated dislocation viewed in cross-section at the Si/$Si_{1-x}Ge_x$ interface following sample annealing and relaxation. **(f)** HREM image of a precipitate within as-grown, metastable Si/$Si_{1-x}Ge_x$/Si

collection, away from the misfit dislocations delineating the Si/$Si_{1-x}Ge_x$ and $Si_{1-x}Ge_x$/Si interfaces. The low-magnification STEBIC image shown in Fig. 17.27b illustrates the recombination activity within a relaxed Si/$Si_{0.96}Ge_{0.04}$/Si(001) sample imaged in plan view. Submicron resolution of the recombination activity is readily achievable using this technique (Fig. 17.27c), with line scans from digitized images indicating a resolution of $\approx 0.3\,\mu m$ in this case. The spacing of $\approx 1\,\mu m$ striations in the STEBIC image is much greater than the spacing of individual misfit dislocations shown in the associated TEM image (Fig. 17.27d), and is more closely associated with the spacing of dislocation bundles. Thus, correlation with structural images shows that bunched arrays of orthogonal ⟨110⟩ misfit dislocations are primarily responsible for the enhanced recombination. For this particular sample, detailed $\boldsymbol{g} \cdot \boldsymbol{b}$ analysis confirmed the presence of bands of predominantly 60° misfit dislocations with a few 90° segments arising from dislocation interactions.

Supporting evidence for these dislocations being dissociated and probably decorated by transition metal impurities was obtained from complementary HREM and EDX investigations of metastable and relaxed Si/$Si_{1-x}Ge_x$/Si(001) samples from the same growth trial. As the relaxation of a low Ge content, metastable Si/$Si_{1-x}Ge_x$/Si structure proceeds, extensive arrays of orthogonal ⟨110⟩ dislocations form and interact, with

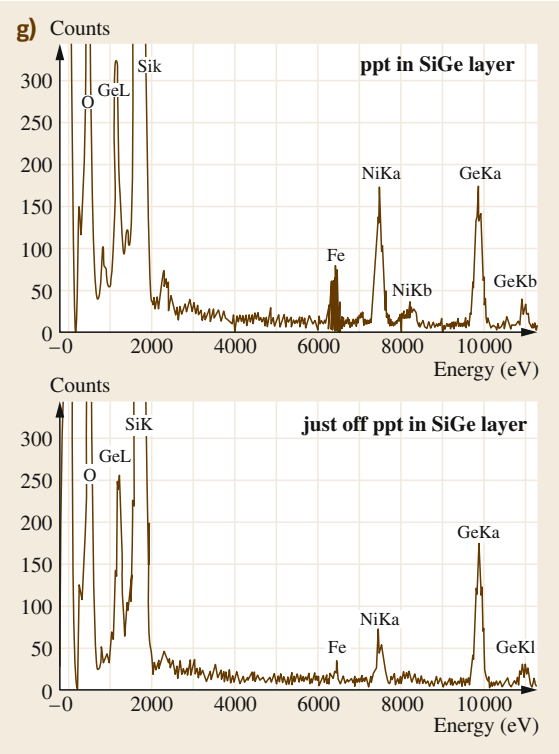


Fig. 17.27 (g) EDX spectra confirming the presence of Fe transition metal impurities

dislocations generated in the strained $Si_{1-x}Ge_x$ layer being pushed by repulsive dislocation forces into the Si substrate and cap on {111} glide planes. HREM observations of relaxed $Si/Si_{1-x}Ge_x/Si$ samples in cross-section confirmed that the misfit dislocations were dissociated, with tails associated with each of the partials, indicative of impurity decoration (Fig. 17.27e). This particular image was acquired before the development of electron beam-induced damage artefact structures within this sample foil and so is considered representative of the as-grown material. A distribution of small precipitates, of typical size 5 nm, was also identified within metastable samples prior to strain relaxation. These precipitates showed strong scattering in HREM (Fig. 17.27f) and revealed the presence of Fe when analyzed using EDX within a dedicated STEM (1 nm probe size), as shown in Fig. 17.27g. No Fe was present in spectra acquired immediately to the side of the precipitates. (The Ni signal was considered to be an artefact of EDX acquisition and attributed to fluorescence from x-rays and electrons interacting with the specimen's Ni support ring.) Hence, the suggestion is that enhanced transition metal impurity segregation at dissociated dislocations is responsible for the STEBIC contrast observed. This emphasizes the need to control dopants or impurity sources in the vicinity of heterostructure interfaces during the development of functional device structures.

17.8 Concluding Remarks

The above commentary has attempted to convey the framework underpinning a variety of analytical techniques used to investigate the structures of semiconductors. It is emphasized that an appropriate combination of assessment techniques should be applied generally, since no single technique of assessment will provide information on the morphology, composition, microstructure and (opto)electronic properties of a given functional material or processed device structure.

This type of considered approach to materials characterization is required in order to break free of the *black-box* mentality that can develop if one is too trusting of the output generated by automated or computerized instrumentation systems. One should always bear in mind the process of signal generation that provides the information content. This in turn helps us to develop an appreciation of performance parameters such as spatial or spectral resolution, in addition to sensitivity, precision and detection limits. One should consider technique calibration and the appropriate use of standards in order to ensure that the data acquired is appropriate (and reproducible) to the problem being addressed. Consideration should also be given to the form and structure of the data being acquired and how the data sets are analyzed. In this context, distinction should be made between the processing of analog and digital information and the consequences of data conversion. Issues regarding the interpretation (or misinterpretation) of results often stem from the handling of experimental errors. On a practical level, rigorous experimental techniques should be applied to ensure that the data generated is both meaningful and representative of the sample being investigated, free from artefacts from the preparation and investigation processes. There are clearly differences between qualitative assessment and the more rigorous demands of quantitative analysis. The level of effort invested often reflects the nature of the problem that is being addressed. A comparative assessment of a number of samples may simply require a qualitative investigation (for example, in order to solve a specific materials science problem within a growth or device fabrication

process). Alternatively, quantitative analysis may be required to gain a more complete understanding of the nature of a given sample, such as the precise composition. To summarize, an awareness of the methodology used in any investigation is required to establish confidence in the relevance of the results obtained. A range of complementary analysis techniques should ideally be applied to gain a more considered view of a given sample structure.

Acknowledgments. As ever, there are many people one wishes to acknowledge for their involvement in the growth, processing and underpinning characterisation research programmes drawn from to illustrate this chapter. University of Nottingham: with thanks to Tom Foxon, T.S. Cheng, Sergei Novikov and Chris Statton for the provision of MBE GaN samples and supporting XRD analysis; and to Mike Fay for GaAs CBED patterns. University of Cambridge: with thanks to Colin Humphreys for provision of instrumentation; Chris Boothroyd for STEM data on the SiGe and GaN samples; Michael Natusch for GaN EELS data; Robin Taylor for RHEED stage development; David Tricker for the Si-doped GaN micrograph; and Yan Xin for the GaN images used for dislocation analysis. University of Warwick: with thanks to Richard Kubiak and E.H.C. Parker for supplying SiGe/Si samples. Polish academy of Sciences, Warsaw: with thanks to Jan Weyher for homoepitaxial GaN samples. With thanks also to the EPSRC for funding support.

References

17.1 R.W. Cahn, E. Lifshin: *Concise Encyclopedia of Materials Characterization* (Pergamon, New York 1992)
17.2 J.M. Cowley: *Electron Diffraction Techniques*, Vol. 1 (Oxford Univ. Press, Oxford 1992)
17.3 J.M. Cowley: *Electron Diffraction Techniques*, Vol. 2 (Oxford Univ. Press, Oxford 1993)
17.4 B.D. Cullity, S.R. Stock: *Elements of X-Ray Diffraction*, 3rd edn. (Addison Wesley, New York 1978)
17.5 J.W. Edington: *Practical Electron Microscopy in Materials Science* (Philips Electron Optics, Eindhoven 1976)
17.6 R.F. Egerton: *Electron Energy-Loss Spectroscopy in the Electron Microscope* (Plenum, New York 1996)
17.7 P.J. Goodhew, F.J. Humphreys, R. Beanland: *Electron Microscopy and Analysis* (Taylor Francis, New York 2001)
17.8 P.J. Grundy, G.A. Jones: *Electron Microscopy in the Study of Materials* (Edward Arnold, London 1976)
17.9 P.B. Hirsch, A. Howie, R.B. Nicholson, D.W. Pashley, M.J. Whelan: *Electron Microscopy of Thin Crystals* (Butterworths, London 1965)
17.10 I.P. Jones: *Chemical Microanalysis Using Electron Beams* (Institute of Materials, London 1992)
17.11 D.C. Joy, A.D. Romig, J.I. Goldstein: *Principles of Analytical Electron Microscopy* (Plenum, New York 1986)
17.12 M.H. Loretto, R.E. Smallman: *Defect Analysis in Electron Microscopy* (Chapman Hall, London 1975)
17.13 D. Shindo, K. Hiraga: *High-Resolution Electron Microscopy for Materials Science* (Springer, Berlin, Heidelberg 1998)
17.14 J.C.H. Spence: *High-Resolution Electron Microscopy* (Oxford Univ. Press, New York 2013)
17.15 G. Thomas, M.J. Goringe: *Transmission Electron Microscopy of Metals* (Wiley, New York 1979)
17.16 D.B. Williams, C.B. Carter: *Transmission Electron Microscopy: A Textbook for Materials Science* (Springer, New York 2009)
17.17 R.F. Egerton: *Physical Principles of Electron Microscopy: An Introduction to TEM, SEM and AEM* (Springer, New York 2005)
17.18 R. Brydson: *Aberration-Corrected Analytical Transmission Electron Microscopy* (Wiley, Chichester 2011)
17.19 P. Echlin: *Handbook of Sample Preparation for Scanning Electron Microscopy and X-Ray Microanalysis* (Springer, New York 2009)
17.20 L.A. Giannuzzi: *Introduction to Focused Ion Beams: Instrumentation, Theory, Techniques and Practice* (Springer, New York 2005)
17.21 R. Brydson: *Electron Energy Loss Spectroscopy* (Garland Science, New York 2001)
17.22 J. Ayeche, L. Beaunier, J. Boumendil, G. Ehret, D. Laub: *Sample Preparation Handbook for Transmission Electron Microscopy* (Springer, New York 2010)
17.23 R. Hull, J.C. Bean: Crit. Rev. Solid State **17**, 507 (1992)
17.24 T. Sugahara, H. Sato, M. Hao, Y. Naoi, S. Kurai, S. Tattori, K. Yamashita, K. Nishino, L.T. Romano, S. Sakai: Jpn. J. Appl. Phys. **37**, 398 (1997)
17.25 Y. Xin, P.D. Brown, T.S. Cheng, C.T. Foxon, C.J. Humphreys: Inst. Phys. Conf. Ser. **157**, 95 (1997)
17.26 Y. Ishida, H. Ishida, K. Kohra, H. Ichinose: Philos. Mag. A **42**, 453 (1980)
17.27 D.B. Holt: J. Mater. Sci. **23**, 1131 (1988)
17.28 K. Ishizuka, J. Taftø: Acta. Cryst. B **40**, 332 (1984)
17.29 D. Cherns, W.T. Young, M. Saunders, J.W. Steeds, F.A. Ponce, S. Nakamura: Philos. Mag. A **77**, 273 (1998)
17.30 J.M. Cowley: *Electron Diffraction: An Introduction*, Vol. 1 (Oxford Univ. Press, Oxford 1992)
17.31 G.J. Russell: Prog. Cryst. Growth Ch. **5**, 291 (1982)
17.32 J.L. Weyher, P.D. Brown, A.R.A. Zauner, S. Muller, C.B. Boothroyd, D.T. Foord, P.R. Hageman, C.J. Humphreys, P.K. Larsen, I. Grzegory, S. Porowski: J. Cryst. Growth **204**, 419 (1999)
17.33 P.D. Brown, D.M. Tricker, C.J. Humphreys, T.S. Cheng, C.T. Foxon, D. Evans, S. Galloway, J. Brock: Mater. Res. Soc. Symp. Proc. **482**, 399 (1998)
17.34 P.D. Brown, C.J. Humphreys: J. Appl. Phys. **80**, 2527 (1996)

18. Surface Chemical Analysis

David Sykes

The physical bases of surface chemical analysis techniques are described in the context of semiconductor analysis. Particular emphasis is placed on the SIMS (secondary ion mass spectrometry) technique, as this is one of the more useful tools for routine semiconductor characterization. The practical application of these methods is addressed in preference to describing the frontiers of current research.

18.1 **Electron Spectroscopy** 414
18.1.1 Auger Electron Spectroscopy 414
18.1.2 X-Ray Photoelectron Spectroscopy (XPS) 415
18.2 **Glow-Discharge Spectroscopies (GDOES and GDMS)** 416
18.3 **Secondary Ion Mass Spectrometry (SIMS)** 417
18.4 **Conclusion** 423

Surface chemical analysis is a term that is applied to a range of analytical techniques that are used to determine the elements and molecules present in the outer layers of solid samples. In most cases, these techniques can also be used to probe the depth distributions of species below the outermost surface. In 1992 the International Standards Organisation (ISO) established a technical committee (TC) on surface chemical analysis (ISO TC 201) to harmonize methods and procedures in surface chemical analysis. ISO TC 201 has a number of subcommittees that deal with different surface chemical analytical techniques and this chapter will discuss the applications of these different methods, defined by ISO TC 201, in the context of semiconductor analyses. In particular, this discussion is intended to deal with practical issues concerning the application of surface chemical analysis to routine measurement rather than to the frontiers of current research. Standards relating to surface chemical analysis developed by the ISO TC201 committee can be found on the ISO TC201 web site www.iso.org (under *standards development*).

Traditional surface chemical analysis techniques include the electron spectroscopy-based methods Auger electron spectroscopy (AES or simply Auger) and x-ray photoelectron spectroscopy (XPS, once also known as ESCA – electron spectroscopy for chemical analysis), and the mass spectrometry method SIMS (secondary ion mass spectrometry). The ISO TC 201 committee also has a subcommittee that deals with glow discharge spectroscopies. Whilst these latter methods have been used more for bulk analysis than surface analysis, the information they produce comes from the surface of the sample as that surface moves into the sample, and so they have been finding applications in depth profiling studies.

One thing that is common to all of these surface chemical analysis techniques is that they are vacuum-based methods. In other words, the sample has to be loaded into a high or ultrahigh vacuum system for the analysis to be carried out. With the one exception of glow discharge optical emission spectroscopy (GDOES), where the analysis relies upon the detection of photons, all of the techniques also depend upon the detection of charged particles. This requirement for vacuum operation necessarily imposes limits on the types and sizes of samples that can be analyzed, although of course instruments capable of handling semiconductor wafers do exist. The quality of the vacuum environment around the sample can also affect the quality of the analysis, especially with regard to the detection of elements that exist in the atmosphere around us. The size and complexity of surface chemical analysis equipment has arguably tended to limit the wider use of these powerful methods.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_18

18.1 Electron Spectroscopy

In the electron spectroscopies, Auger and XPS, the surface of the sample is probed by an exciting beam which causes electrons to be ejected from the atoms in the sample. These electrons are collected and their energies analyzed. The two techniques are similar but subtly different.

18.1.1 Auger Electron Spectroscopy

In Auger, a beam of electrons is used to excite the sample. During the interaction of the primary electron beam with the sample atoms, core electrons are knocked out, creating vacancies in the inner electron shells. Electrons from outer shells can fall into the vacancy, thus leaving the atom in an unstable state and, in order to return to equilibrium, the excess energy the atom possesses can be dissipated in one of two ways: either by the emission of an x-ray photon with a characteristic energy (the basis of energy- or wavelength-dispersive x-ray analysis), or by the emission of a third electron (the Auger electron), with an energy determined by the difference in energy between the original core state and those of the two other levels involved. Clearly these energies are uniquely determined by the energy levels in the atom and thus provide a route for analysis. As three electrons are necessary for the Auger process to occur, AES is incapable of detecting hydrogen or helium, but all other elements produce characteristic Auger electrons. The energies of the Auger electrons can range from a few tens of electron volts to a few thousand electron volts. In this energy regime, electrons can only travel of the order of monolayers through a solid before an interaction occurs which causes a loss of energy, thus destroying the analytical information the electron possessed. It is this property which gives the Auger technique its surface sensitivity – although Auger electrons will be produced as far into the material as the primary beam can penetrate, only those produced in the top few monolayers can escape with their characteristic energy intact. The Auger electrons excited deeper into the sample lose energy before escaping and contribute to a background signal, as do the scattered primary electrons and the initial core electrons ejected at the start of the process. These scattered electrons produce a background signal against which the Auger electrons must be detected; this limits the analytical sensitivity that can be achieved by Auger. This is also why Auger spectra are sometimes displayed as the differential of the number of electrons against energy, because the small Auger peaks are more rapidly varying functions than the larger, slowly changing background signal and hence are enhanced by the differentiation process. Figure 18.1 shows an Auger spectrum from GaInAsP, where the spectrum plotted as the number of electrons as a function of energy is compared to the differential of the number of electrons as a function of energy.

As the number of Auger electrons produced is proportional to the number of atoms excited, Auger offers the ability to perform quantitative as well as qualitative analyses, although some form of calibration is required, either through the use of local reference materials or instrument calibration and standard databases. So Auger offers quantitative analysis for all elements from lithium to uranium from layers only a few atoms in thickness. What gives Auger an extra dimension is the ability to profile into the sample by removing the outermost surface layers with an inert gas ion beam (usually argon) in order to expose the layers below. This sputter

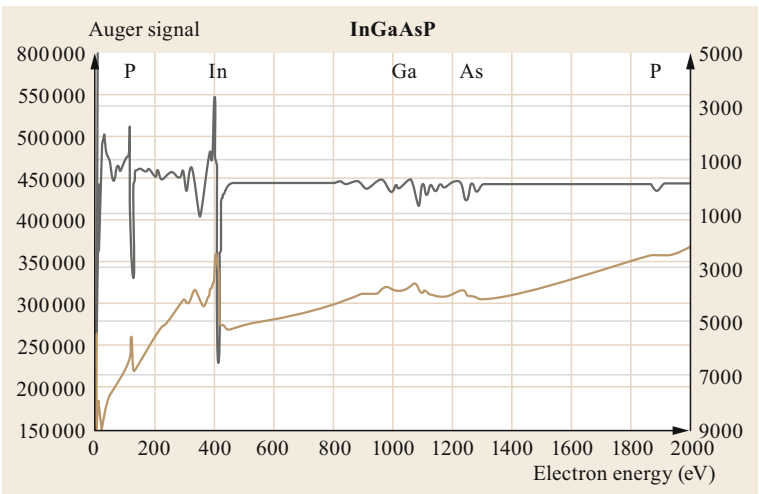


Fig. 18.1 Auger spectrum from GaInAsP, shown as the number of electrons and the differential of the number of electrons as a function of energy

depth profiling is the inverse of the sputter deposition widely used to deposit thin layers of material. Argon is the most widely used ion beam in Auger depth profiling, as its effects are physical rather than chemical, although it should be noted that some chemicals can be modified by the sputtering process. For example, some metal oxides can be reduced by sputtering while others are not, and so any chemical state data inferred from atomic compositions in sputter depth profiles should be treated with caution. Physical effects can also occur, for example atomic mixing and the development of surface topography, which can distort the shape of buried features.

One of the great advantages of Auger is that, as the excitation is provided by electrons, the primary beam can be easily focused and scanned over the surface of the sample. By detecting the scattered electrons, a physical image of the sample is produced as in the scanning electron microscope, and maps of elemental distributions can be obtained by detecting the Auger signal as a function of beam position. These elemental maps can be time-consuming to acquire but can be useful when making a point or illustrating a book chapter; however, for practical analyses it is often adequate to identify the feature of interest from the physical image. Figure 18.2 shows aluminium and oxygen Auger maps from a contaminated bond pad. In the region of the contamination, the aluminium signal is reduced and the oxygen signal is higher compared with the uncontaminated regions.

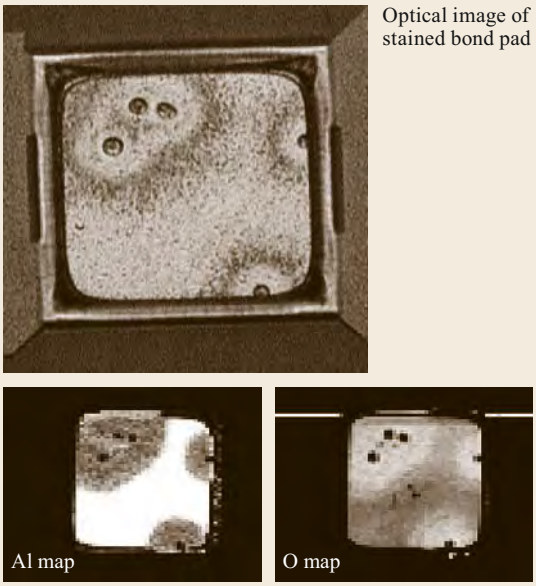


Fig. 18.2 Optical image and Auger maps of a stained aluminium bond pad

In terms of hardware, Auger systems can be standalone systems comprising an electron beam column, an electron energy analyzer, an inert gas ion gun and sample handling stage and an associated vacuum chamber (or chambers), or form a part of a multitechnique system with x-ray sources for XPS analysis or a mass spectrometer for basic SIMS studies.

18.1.2 X-Ray Photoelectron Spectroscopy (XPS)

XPS is very similar to Auger in terms of the instrumentation and physics involved. The primary excitation, as the name implies, is, in this case, a beam of x-rays. The x-rays, often magnesium or aluminium Kα, eject core electrons from the surface atoms by the photoelectron effect. The kinetic energy of the emitted photoelectrons will be equal to the difference between the x-ray photon energy and the core level binding energy, and thus will be less than $\approx 1100\,\mathrm{eV}$ or $\approx 1400\,\mathrm{eV}$ for magnesium and aluminium Kα respectively. So, as with Auger, the mean free paths of the photoelectrons are of the order of monolayers in solid materials. As a core level vacancy is produced by the excitation, Auger electrons will also be present in the XPS spectra but, by convention, Auger spectroscopy refers to the electron-excited situation. Whereas in Auger electron spectroscopy the electron energy is usually referred to in terms of the electron's kinetic energy, in XPS the electron binding energy is usually plotted as the ordinate in the spectra. This means that, whatever x-ray excitation is used, be it magnesium, aluminium or a more exotic material, the photoelectrons will appear at the same binding energy in the spectra but the apparent positions of the Auger peaks will change (on the binding energy scale) as their kinetic energy is independent of the excitation source. There are no scattered primary electrons (which are always present in Auger spectra) in the XPS spectra, so these spectra have better signal-to-background, and the photoelectron peaks are easily distinguished against the background arising from scattered photoelectrons produced deeper into the sample and other secondary processes.

As the problem of focusing electron beams is much simpler than focusing x-ray beams, XPS is perceived as a large-area technique, whereas Auger is the technique of choice for small area analysis. However, the relentless advances made in the performance and design of instruments means that XPS instruments can achieve spatial resolutions of the order of 1 to 10 μm. In XPS no charge is brought to the sample by the primary excitation, and so insulating samples are easier to analyze with XPS than with Auger; also, the photoelectron peaks show small energy differences in the peaks po-

sitions depending upon the local chemical environment of the atom from which they originated, the so-called chemical shift. While chemical shifts are present in some Auger peaks, this is the exception rather than the rule, and XPS is the technique of choice where information on the local chemical state of the surface is required. The physics of the XPS process is probably even better understood than the Auger process, and quantification of the spectra is relatively routine.

As with Auger, composition depth profiles can be produced by sputtering the surface of the sample with an inert gas ion beam. Again caution is advised when interpreting chemical state information from a surface that has been subject to ion bombardment. With both of the electron spectroscopies it is the surface of the sample that remains after sputtering that is analyzed in a depth profile, and there are two factors to be aware of (if not more). Once the passivating surface layer has been sputtered away, the surface of the sample may become chemically active and getter residual gas from the vacuum system. If the sample is a multicomponent material, one component may have a higher sputtering rate than the other, so as the sputtering process proceeds the surface will become depleted in the higher sputtering rate material. This process will continue until an equilibrium state is reached where the material leaving the surface is in the same ratio as the bulk composition; the corollary of this is that the surface will be enriched in the lower sputter rate material and so the composition of the material as measured by either XPS or Auger will be in error unless this effect is understood and accounted for.

Both Auger and XPS are capable of detecting all elements from lithium to uranium (and beyond), and have sensitivities in the parts per hundred to parts per thousand regime. The responses vary from element to element but typically sensitivities remain within an order of magnitude or so between elements.

18.2 Glow-Discharge Spectroscopies (GDOES and GDMS)

These are two apparently similar but quite unrelated techniques that rely upon glow discharges as the excitation source. In glow discharge optical emission spectroscopy (GDOES) a high-pressure glow discharge is used to sputter material from the surface of a sample, and this sputtered material is detected by the optical emissions it produces in the glow discharge. In glow discharge mass spectrometry (GDMS), a low pressure dc glow discharge is used to sputter material from the sample surface and ionized material from the discharge is extracted into a mass spectrometer for analysis.

GDOES is probably the simplest of all the surface chemical analysis techniques, at least as far as the vacuum requirements are concerned. There is no complex vacuum system, as needed for all of the other methods, and the sample itself sits with atmospheric pressure on one side of it while the opposite face acts as one electrode of a glow discharge cell that is pumped by a simple vacuum pump. A flow of high-purity argon gas flows through the cell, providing the sputtering and discharge gas and purging the cell of impurities and material removed from the sample. A window at the other end of the discharge cell transmits light from the discharge into an optical spectrometer. By using a spectrometer with a number of photomultiplier detectors, prepositioned at the known wavelengths of the expected element emission lines, data from a large number of elemental channels can be collected in parallel, making GDOES an extremely efficient analytical system. A scanning spectrometer can also be included in the instrument to provide a continuous spectral scan to detect emission lines from elements other than those built into the instrument, but this is, of course, a serial detection device and the advantages of parallel acquisition are lost.

Traditionally GDOES has been widely used for the analysis of metals and coatings on metals, but it is currently also finding application in the area of semiconductor materials. With the development of radio frequency glow discharge sources, the technique is capable of analyzing insulators, and with the high sputtering rates available, it can depth profile tens of μm into dielectric layers. With fast electronics and parallel detection, thin oxide layers can also be profiled. The sensitivity of the GDOES technique lies between that of Auger or XPS and that of SIMS. One potential weakness of the method is that the technique has no spatial resolution and the analysis area is millimeters in diameter. It is thus useful for large-area plain samples, but cannot be used with patterned material or to probe small features. The technique is useful for bulk analysis, but it is the ability to depth-profile into material, providing an insight into layer structures, that is its greatest appeal. However, because the technique has no ability to discriminate where the analytical signal is coming from, the quality of the depth profiles produced will be compromised by crater edge effects. In other words, while most of the analytical signal will originate from the bottom of the sputtered crater, there will always be some information

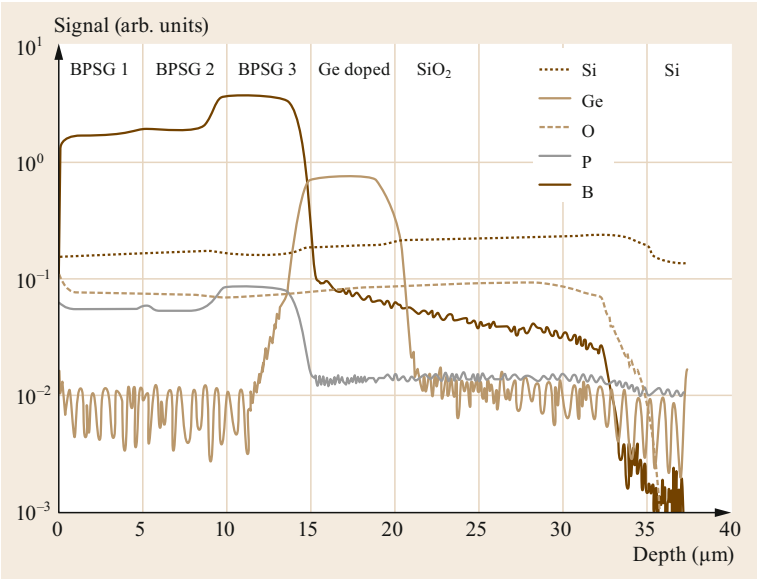


Fig. 18.3 GDOES depth profile through a DWDM multilayer glass structure

that comes from the crater side wall. The consequence of this is that, with layered structures, layers closer to the surface will appear to tail into layers beneath them, even though the interface between the layers is abrupt. This effect can be seen in the depth profile shown in Fig. 18.3, which shows a GDOES profile into a dense wavelength division multiplexing (DWDM) structure.

Glow discharge mass spectrometry (GDMS) is a considerably more complex technique, at least from an instrumental point of view. Originally developed as a method of bulk analysis, GDMS is probably the most sensitive, in terms of the detection limit achievable, of all of the techniques being considered here. As with GDOES, in GDMS the sample forms one electrode in a simple glow discharge cell. However, in the case of GDMS, the discharge cell is mounted within a high-vacuum system. In its original form, the sample (typically be 1 mm^2 by about 15 mm long) is placed in the center of a cylindrical cell into which argon is leaked at low pressure. By applying a dc voltage between the sample and the cell, an argon plasma is created which sputters the outside of the sample, removing material. This material, some of which is ionized but the majority of which is neutral as it leaves the surface, is ionized by a variety of processes as it passes through the glow discharge plasma. These ions are then accelerated into a high-resolution magnetic sector mass spectrometer where they are mass-analyzed and counted. Instruments can also be based on quadrupole mass spectrometers, but it is the magnetic sector instruments which offer the greater sensitivity. By sweeping the mass spectrometer through a range of masses, which can cover the entire periodic table, the major, minor and trace elements present in a sample can be determined. GDMS is a particularly powerful method of detecting the trace elements present in bulk semiconductor materials at levels down to parts per billion.

It is also possible to analyze flat, rather than matchstick-shaped, samples in GDMS. Just as in GDOES, the flat sample is positioned at the end of the discharge cell, and a cylindrical crater is etched into the sample surface. As with GDOES, with GDMS there is no spatial resolution, and the depth information from layered structures will be distorted by crater edge effects and loss of crater base flatness as it is not possible to discriminate between ions produced from the base of the crater and those produced from the sidewalls.

18.3 Secondary Ion Mass Spectrometry (SIMS)

SIMS is probably the most powerful and versatile of all of the surface analysis techniques and comes in the widest variety of instrumentations, from big, stand-alone instruments to bench-top instruments and add-ons to electron spectrometers. SIMS can offer chemical identification of submonolayer organic contamination, measurement of dopant concentrations, and can produce maps and depth profile distributions from nanome-

ters to tens of μm in depth. However, no one instrument is going to be capable of all of these tasks, and even if it could it would not be able to achieve all of them at the same time.

SIMS, in its simplest form, requires an ion gun and a mass spectrometer. The sample is placed in a vacuum chamber and ions from the ion gun sputter the sample surface. Material is sputtered from the sample surface and some of this will be ionized, although in most cases the major part of the sputtered material will be in the form of a neutral species. The ionized component of the sputtered material is mass-analyzed with the mass spectrometer.

The technique has evolved in various directions from this common origin to produce a variety of subtly different variants of the SIMS technique, including dynamic SIMS (DSIMS), static SIMS (SSIMS) and time of flight SIMS (ToFSIMS), each of which has its own distinct attributes. There are three main types of mass spectrometer used for SIMS analysis: the magnetic sector, the quadrupole and the time of flight, ToF. Dedicated depth-profiling SIMS machines, dynamic SIMS instruments, tend to employ either magnetic sector or quadrupole mass spectrometers. Magnetic sector instruments offer high transmission and high mass resolution capabilities, useful for separating adjacent mass peaks with a very small mass difference, for example ^{31}P from ^{30}SiH. Quadrupole mass spectrometers offer ultrahigh vacuum compatibility and, as well as being used in DSIMS instruments, smaller versions are also found as add-ons to Auger/XPS instruments and bench-top instruments. Time of flight instruments are remarkably efficient in their use of material in that the entire mass spectrum is sampled in parallel, whereas in the magnetic sector and quadrupole instruments the spectrum is produced by sequentially scanning through the mass range of interest. ToFs can also offer high mass resolution but profile relatively slowly because it is necessary to use a pulsed ion beam with a low duty cycle. Whilst static SIMS can be carried out on either a quadrupole or magnetic sector instrument, the ToF-based instruments are more suited to the task as less material is used in the course of the analysis. However, when dynamic SIMS, where only a few elements need to be monitored, is the goal, the magnetic sector or quadrupole instrument is the appropriate choice, although ToF machines can be used for shallow profiling.

Figure 18.4 shows a schematic diagram of the basis of these SIMS techniques. Ions from the ion gun bombard the sample surface and transfer momentum to the atoms in the sample creating a collision cascade. This cascade distributes the energy of the incoming ion amongst the atoms in the sample, causing them to be displaced from their original sites, possibly breaking

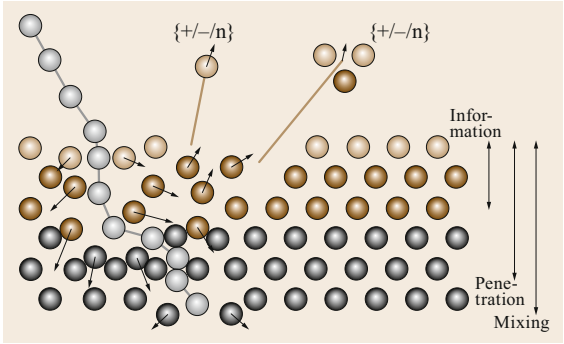


Fig. 18.4 Schematic diagram of the SIMS process

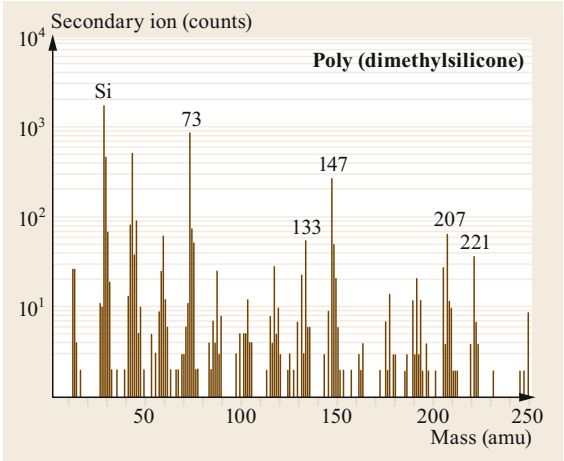


Fig. 18.5 Static SIMS spectrum from a thin film of poly(dimethylsilicone)

some bonds between atoms whilst creating others. The bombarding ion also can become implanted into the target material, modifying the chemistry of the material in the process. None of this is peculiar to SIMS – it happens in any sputtering process, be it in Auger depth-profiling, GDMS analysis or sputter deposition systems. It is probable that some of the energy deposited in the sample will cause atoms or molecules to be ejected from the surface of the sample, and those atoms or molecules that leave the surface as ions can be collected and detected by the mass spectrometer.

In the early stages of sputtering, the atoms and molecules sputtered from the surface will originate from areas of virgin surface – in other words from sites that have not yet been damaged by the primary ion – and thus carry with them information about the chemistry of the outer molecular layers of the sample. As the sputtering process proceeds, the probability of the sputtered particles being emitted from an area that has been modified by earlier ion impacts increases. Thus, at the start of the process the sputtered particles are characteristic

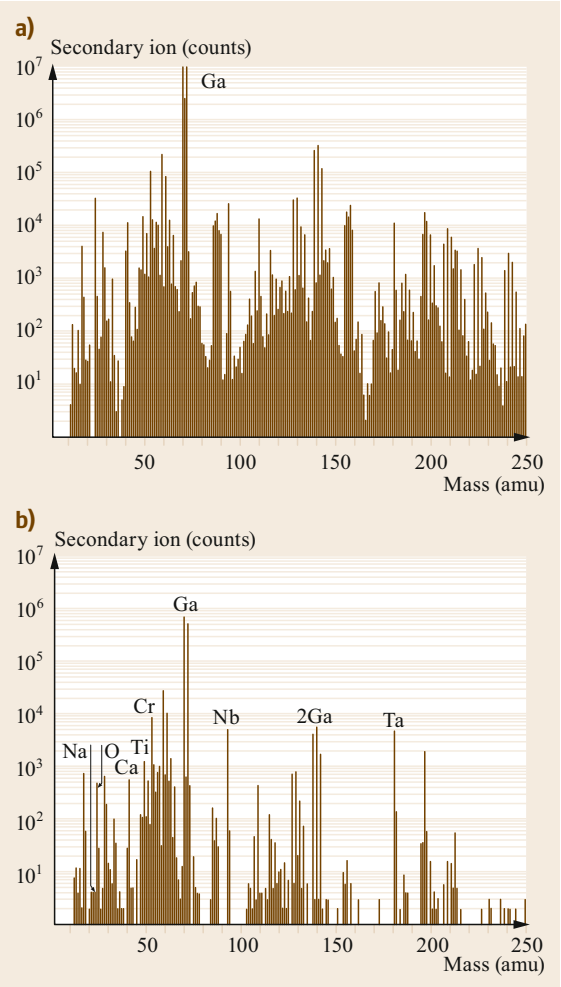


Fig. 18.6a,b SIMS mass spectrum from an unknown sample **(a)** without energy filtering, **(b)** with energy filtering

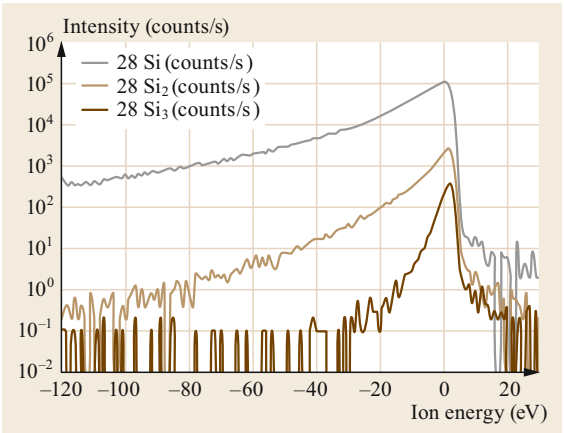


Fig. 18.7 The ion energy distributions of Si^+, Si_2^+ and Si_3^+ ions

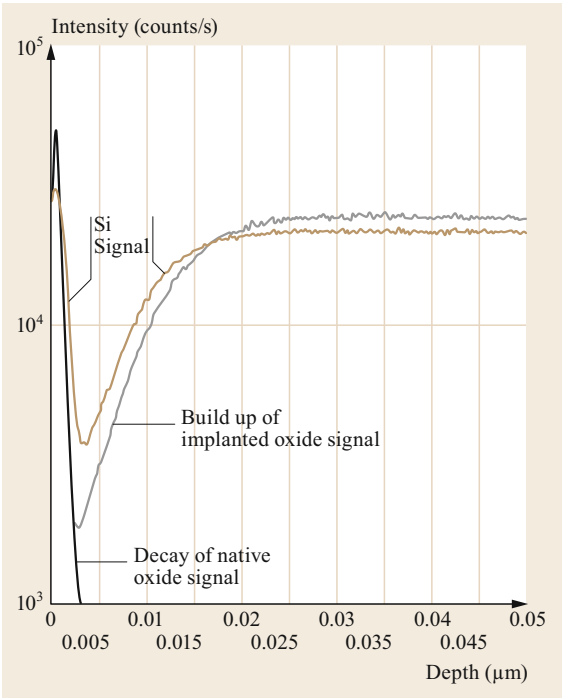


Fig. 18.8 The evolution of the Si^+ signal from a clean silicon wafer as a function of sputtering time with an oxygen primary ion beam, showing the decay of the surface oxide signal and the build-up of the implanted oxide layer

of the virgin surface, but they will eventually become characteristic of the ion beam-modified surface.

Static SIMS is concerned with the measurement of the sputtered molecules produced at the start of the process, where information about the surface chemistry of the sample can be obtained. In SSIMS a very low dose of primary ions is used, typically less than 10^{13} primary ions per square centimeter, so that there is a very low probability that the ions that are detected come from damaged material. Used in this way, SIMS can be used to identify organic contamination on surfaces and obtain information about the molecular structure of the sample. Figure 18.5 shows a typical SSIMS mass spectrum from poly(dimethylsilicone), a common surface contaminant, which can be recognized by prominent peaks in the mass spectrum at masses 73, 133, 147, 207 and 221, which originate from fragmentation of the parent molecule $(CH_3{-}Si(CH_3)_2{-}O{-}Si(CH_3)_2{-}O{-}\ldots{-}Si(CH_3)_2{-}CH_3)$.

Whilst the abundance of molecular fragments that can be produced by the sputtering process is of great value for revealing chemical information about the sample surface in the SSIMS context, it can also be something of a problem when carrying out elemental analysis

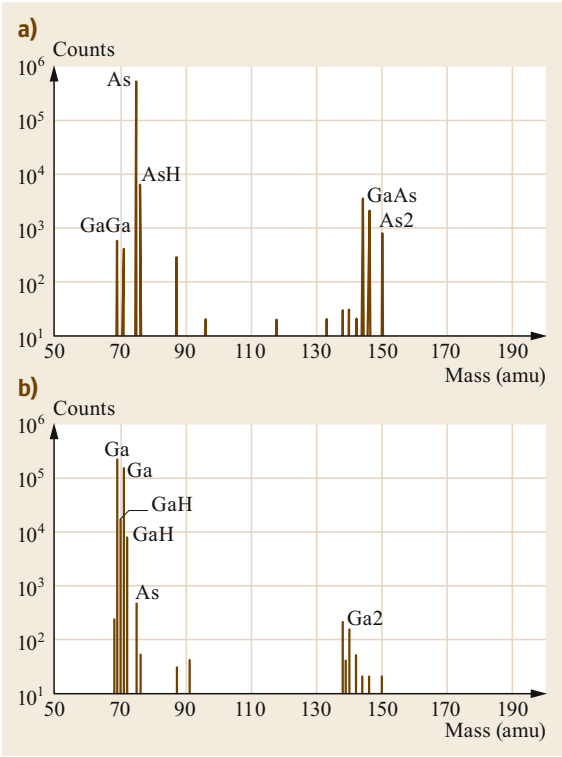


Fig. 18.9a,b Mass spectra from a GaAs wafer recorded **(a)** using Cs^+ primary ion bombardment and negative secondary ion detection, and **(b)** using O_2^+ primary ion bombardment and positive secondary ion detection

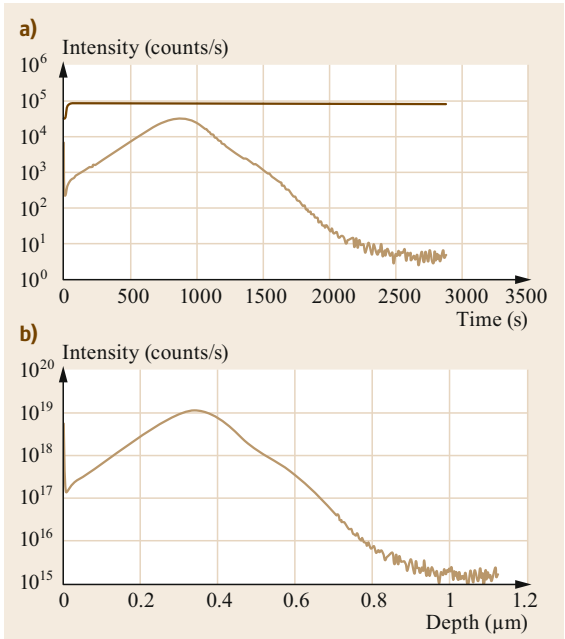


Fig. 18.10a,b Reference sample of boron ion-implanted silicon, **(a)** raw data profile and **(b)** quantified depth profile

of trace impurities in simple matrices. For example, the spectrum shown in Fig. 18.6a, from an unknown Ga-based material, illustrates the number and complexity of molecular species produced by the SIMS process, which can make interpretation difficult. Fortunately, the distribution of the number of secondary ions with energy is different between atomic ions and molecular ions. The atomic ions, in general, have a broader energy distribution than the molecular ions, and its energy distribution tends to become sharper as the complexity of the molecular ion increases. For example, Fig. 18.7 shows the ion energy distributions of the Si^+, Si_2^+ and Si_3^+ ions, illustrating the narrower energy distributions of the molecular ions. Thus, by selecting the ion energy range from which the spectrum is recorded, the molecular information can be suppressed, allowing the elements present in the material to be identified, see Fig. 18.6b.

The intensities of the peaks in the SIMS mass spectra reveal more about the relative ionization probability and instrument transmission function than the number of species of that mass on the sample surface. The signals that are measured in SIMS are proportional to the numbers of ions produced, and this is not simply related to the amount of material present. The degree of ionization can vary enormously from element to element and matrix to matrix. For example, the numbers of positive ions formed by the inert gas bombardment of clean metal samples can be several orders of magnitude lower than the numbers produced from oxidized surfaces of the same metals under identical bombardment conditions. Clearly it is not the concentration of the metal atoms on the surface that is important, as there will be fewer metal atoms in the oxide layer than in the pure metal. It is the presence of oxygen that increases the ionization probability for material leaving the surface as positive ions.

In SSIMS, the chemical nature of the primary ion species is of relatively little consequence in terms of its effect on the number of ions produced by the surface, but in DSIMS, where it is the deliberately modified surface that is of interest, oxygen ion beams are widely used for analyses of electropositive species. Figure 18.8 shows how the silicon matrix signal from a silicon wafer with a native oxide layer behaves as the sputtering process proceeds. Initially there is a strong signal from silicon as the surface oxide layer is still present on the sample surface. The signal then falls as the oxide layer is sputtered away, and then slowly recovers in intensity as oxygen from the primary ion beam is implanted into the silicon surface. Eventually a steady state is reached

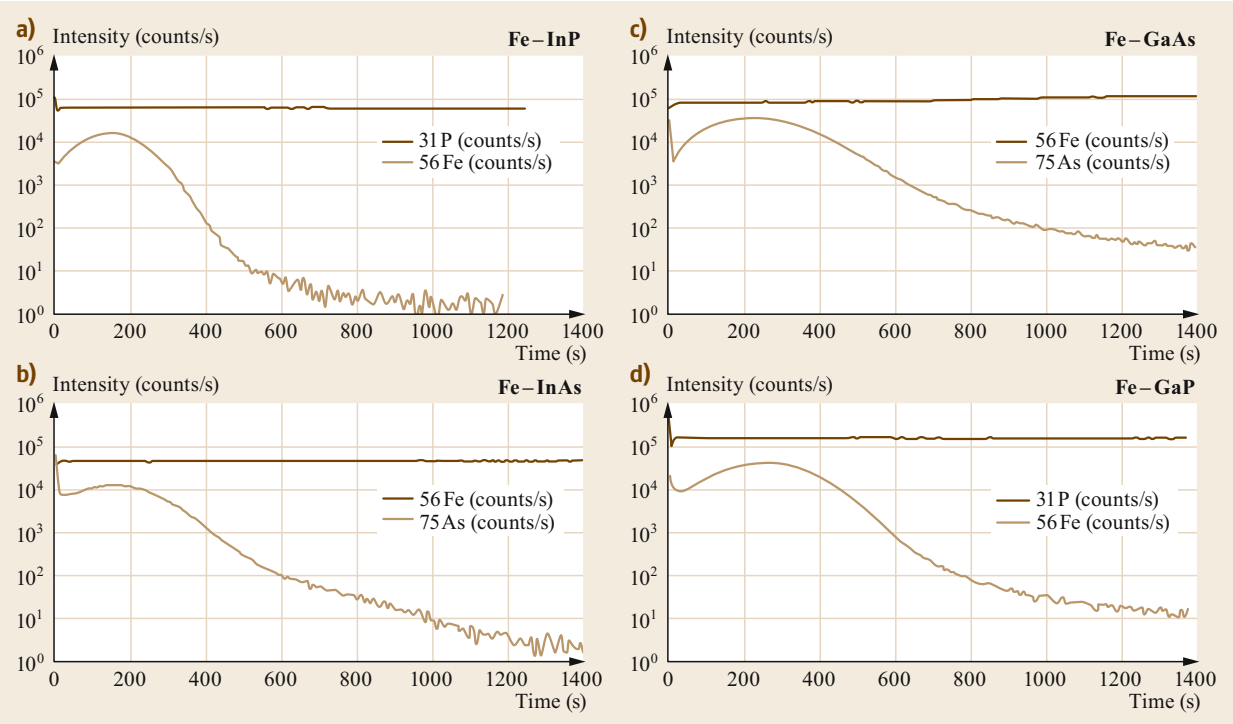


Fig. 18.11a–d Iron profiles, recorded under identical conditions, from implants into **(a)** InP, **(b)** InAs, **(c)** GaAs and **(d)** GaP, respectively

where the material that is being sputtered is constant-composition oxygen-implanted silicon. The thickness of this transient region will depend upon the energy and angle of incidence of the primary oxygen ion beam. The higher the ion beam energy, the thicker the transient region. In the transient region the sputtering rate of the material and the ionization probability may well be changing, and precise quantification in this part of the profile may be difficult. This is of particular importance when the features of interest are very close to the surface and within the transient region. The technological importance of shallow structures has driven the development of very low energy primary ion beam columns to enable the characterization of shallow implants.

Clearly, as is illustrated in Fig. 18.8, the number of positive ions produced in the sputtering process depends upon the concentration of oxygen in the sample surface. In order to maximize the positive secondary ion yield, the sample surface needs to be saturated with oxygen, and this can be achieved with normal-incidence primary ion beams or by deliberately allowing oxygen to flood the sample surface as well as using an oxygen primary ion beam. In the example shown in Fig. 18.8, the primary ion beam was incident at approximately 45° and so complete oxidation of the silicon surface was not achieved, meaning that the silicon and oxygen signals in the equilibrium region are not as high as from the fully oxidized native oxide layer.

Not all species prefer to form positive ions in the sputtering process; some prefer to produce negative ions. Oxygen is a case in point, and indeed it would be difficult to conduct an analysis for oxygen using an oxygen beam. For electronegative species, it turns out that high negative secondary ion yields can be achieved if the surface is sputtered using cesium ions. An illustration of the relative ion yields of gallium and arsenic as positive and negative ions from a gallium arsenide surface with oxygen and cesium ion bombardment are shown in Fig. 18.9. Some species can be reluctant to produce either positive or negative secondary ions, but in some cases (such as zinc), good sensitivity can be achieved by using cesium ion bombardment and monitoring the cesium–element molecular ion, $CsZn^+$ in the case of zinc.

Despite the wide variations in secondary ion yields from element to element and matrix to matrix, the quantification of impurities in semiconductors can be relatively straightforward when reference materials are used. For example, Fig. 18.10a shows a raw data depth profile of ^{11}B ions implanted in silicon; knowing the areal dose of ions implanted into the sample and the depth of the crater sputtered into the sample, it is a rel-

atively simple task (usually buried in the instrument software) to convert boron counts into concentrations (Fig. 18.10b). By recording a signal from the matrix element, either during the profile as in Fig. 18.10a, or in the crater after the profile, a relative sensitivity factor for the analyte (boron) in the matrix of interest (silicon) can be obtained which can then be used to determine the concentration of that analyte in another sample of the same matrix. Errors or uncertainties can arise from both the crater depth measurement and the implanted dose measurement. At the time of writing, there are only three metrologically traceable reference samples for SIMS (boron, arsenic and phosphorus in silicon) produced by NIST in the USA. For other species and other matrices, the analyst must rely upon locally produced materials with no traceability.

Care must be exercised when the analyte of interest is present in different matrices. For example, Fig. 18.11a–d show iron profiles, recorded under identical conditions (positive ion detection with O_2^+ primary ion bombardment), from implants into InP, InAs, GaAs and GaP produced in the same implant run. While the relative sensitivity of iron to the phosphorus matrix signal is the same for InP and GaP matrices, there is a variation of the order of 20% in the arsenide matrices. The useful ion yield of iron (the total number of ions produced as a fraction of the number of atoms present) increases approximately four-fold in the gallium-based matrices compared to the indium-based matrices. The profile in the GaAs sample, Fig. 18.11c, also shows a further complication: the arsenic matrix signal increases at about 1000 s into the profile. This could be interpreted as an upwards drift in the primary ion beam current, but it is actually caused by roughening of the GaAs surface as a result of the sputtering process. This effect is more common with oxygen primary ion bombardment than with cesium bombardment, and it has been shown that it is possible to reduce such effects by rotating the sample during analysis. Not only does the roughening modify the analytical signal, it also causes an apparent broadening of sharp features present in the sample as the surface roughness is convoluted with the feature width. The problems associated with sputter-induced roughness can become more severe when dealing with metal films but, once again, can be overcome by sample rotation.

Another example of how relative sensitivity factors vary with the matrix is shown in Fig. 18.12. Here a comparison is shown between the behavior of the Si^- and $AsSi^-$ signals as the matrix is changed from GaAs to $Ga_{0.16}Al_{0.84}As$. Molecular signals such as $AsSi^-$ are sometimes used to give increased signal compared to the atomic signal but, as in the example shown, the backgrounds are often similarly increased. What is interesting here is that whilst the ion yields from the Si^- and As^- signals increase, but at different rates with increasing Al content, the $AsSi^-$ signals decrease. Ideally the analyst needs to be aware of, and able to take account of, such variations in response. Simply using the

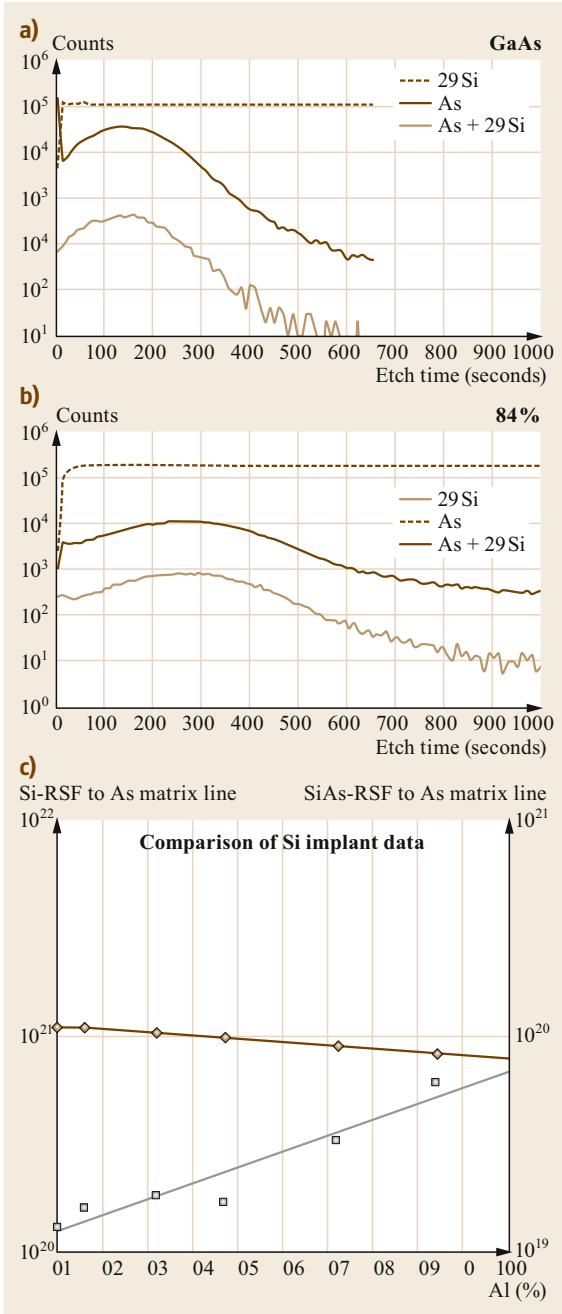


Fig. 18.12a–c Depth profiles of silicon as Si^- and $AsSi^-$ in **(a)** GaAs and **(b)** $Ga_{0.16}Al_{0.84}As$, and **(c)** variation in relative sensitivity factor as a function of aluminium content

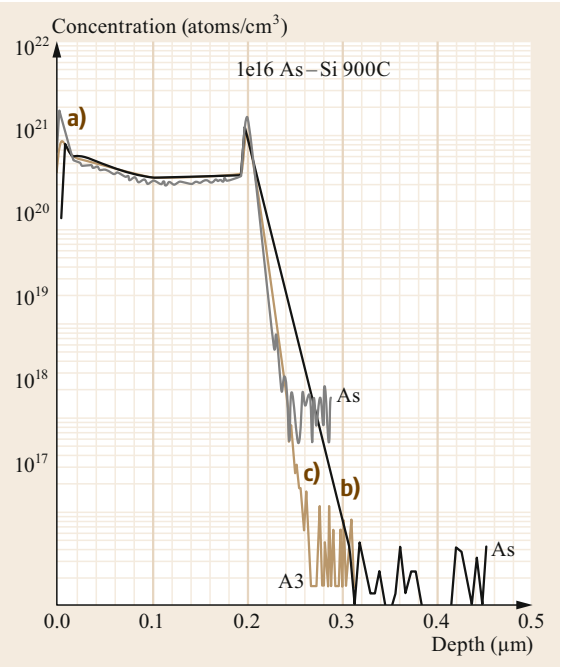


Fig. 18.13a–c Depth profiles of As in Si for different analytical conditions. **(a)** Oxygen ion bombardment, **(b)** cesium ion bombardment, **(c)** oxygen ion bombardment with oxygen flooding of the sample surface

relative sensitivity factor derived for a GaAs matrix and applying this to high Al content material using a point-by-point normalization to the As^- signal can lead to a 30% overestimate of the silicon content. However, monitoring the $AsSi^-$ signal instead and using the same normalization procedure can underestimate the silicon content by over 350%.

With an ideal analytical technique the analytical signal will vary linearly with concentration over all concentrations. With SIMS, for uniform matrices, the measured signal is indeed linear with concentration for dilute systems. However, once the analyte concentration begins to exceed a few percent of the material, the signal may begin to vary in a nonlinear way with concentration. Thus, whilst SIMS is the ideal technique for quantifying minor and trace contaminants, it is less than ideal for determining matrix element concentrations. The most successful method of measuring elemental concentrations at high levels was achieved by monitoring the cesium molecular species ($CsAl^+$ and $CsGa^+$ and $CsAs^+$) in GaAlAs. In GaAlAs the ($CsAl^+/CsAs^+$) and ($CsGa^+/CsAs^+$) ratios do vary linearly with the Al content.

Returning to practical analysis, another factor that needs to be considered in relation to DSIMS, and one that is related to the primary bombarding particles, is atomic mixing. The primary ions used to bombard the surface transfer momentum to the atoms in the sample surface. While some of the atoms will be sputtered from the surface, others will be driven deeper into the sample, thus broadening any concentration distributions that were present within the original sample. The higher the primary ion beam energy, the more severe the atomic mixing. Consequently, as the need to analyze shallower and shallower structures has increased, there has been a move to develop instruments that use lower bombardment energies and sources producing polyatomic ions. A simple example of how the analytical conditions can cause profile distortion is shown in Fig. 18.13. These profiles were recorded from an arsenic-implanted and annealed polysilicon layer on a thin oxide on silicon. Using an oxygen primary ion beam striking the surface at oblique incidence with an impact energy of 5.5 keV (a) produces a profile that shows a spike followed by a rapid fall at the polysilicon/oxide interface. Using a cesium primary ion beam striking the surface at close to normal incidence with an impact energy of 14.5 keV (b) gives a much better detection limit with smaller interface spikes but with a much less abrupt fall in the arsenic signal. Profile (c), which uses the same bombardment conditions as profile (a) but combined with oxygen flooding of the sample surface (and consequently took longer to acquire), shows good agreement with profile (b) in terms of the size of the interface spike – which is exaggerated in profile (a) because of the increased ion yield from the oxide layer compared to the polysilicon – and good agreement with profile (a) in terms of the decay of the arsenic signal in the interface region, and has better sensitivity than profile (a) but less than profile (b). These profiles demonstrate the trade-off that the analyst must make when choosing analysis conditions.

18.4 Conclusion

The various surface chemical analysis techniques have their own strengths and weaknesses. No one method is suitable for all of the tasks the analyst faces; sometime one technique is sufficient to address the problem at hand, sometimes a combination of them is required. However, the approach should be successful if the technique(s) is (are) fit for the purpose of the task.

19. Thermal Properties and Thermal Analysis: Fundamentals, Experimental Techniques and Applications

Safa Kasap, Jiří Málek, Roman Svoboda

The chapter provides a summary of the fundamental concepts that are needed to understand the heat capacity C_P, thermal conductivity κ, and thermal expansion coefficient α_L of materials. The C_P, κ, and α_L of various classes of materials, namely, semiconductors, polymers, and glasses, are reviewed, and various typical characteristics are summarized. A key concept in crystalline solids is the Debye theory of the heat capacity, which has been widely used for many decades for calculating the C_P of crystals. The thermal properties are interrelated through Grüneisen's theorem. Various useful empirical rules for calculating C_P and κ have been used, some of which are summarized. Conventional differential scanning calorimetry (DSC) is a powerful and convenient thermal analysis technique that allows various important physical and chemical transformations, such as the glass transition, crystallization, oxidation, melting etc. to be studied. DSC can also be used to obtain information on the kinetics of the transformations, and some of these thermal analysis techniques are summarized. Temperature-modulated DSC, TMDSC, is a relatively recent innovation in which the sample temperature is ramped slowly and, at the same time, sinusoidally modulated. TMDSC has a number of distinct advantages compared with the conventional DSC since it measures the complex heat capacity. For example, the glass-transition temperature T_g measured by TMDSC has almost no dependence on the thermal history, and corresponds to an almost step life change in C_P.

19.1 **Heat Capacity** 426
19.1.1 Fundamental Debye Heat Capacity of Crystals 426
19.1.2 Specific Heat Capacity of Selected Groups of Materials 428
19.2 **Thermal Conductivity** 431
19.2.1 Definition and Typical Values 431
19.2.2 Thermal Conductivity of Crystalline Insulators 432
19.2.3 Thermal Conductivity of Noncrystalline Insulators 433
19.2.4 Thermal Conductivity of Metals 435
19.3 **Thermal Expansion** 436
19.3.1 Grüneisen's Law and Anharmonicity 436
19.3.2 Thermal Expansion Coefficient α_L 437
19.4 **Enthalpic Thermal Properties** 438
19.4.1 Enthalpy, Heat Capacity and Phase Transformations 438
19.4.2 Conventional Differential Scanning Calorimetry (DSC) 440
19.5 **Temperature-Modulated DSC (TMDSC)** 446
19.5.1 TMDSC Principles 446
19.5.2 TMDSC Applications 447
19.5.3 Tzero Technology 448
References 449

The new Tzero DSC has an additional thermocouple to calibrate better for thermal lags inherent in the DSC measurement, and allows more accurate thermal analysis.

The selection and use of electronic materials, one way or another, invariably involves considering such thermal properties as the specific heat capacity (c_s), thermal conductivity (κ), and various thermodynamic and structural transition temperatures, for example, the melting or fusion temperature (T_m) of a crystal, glass transformation (T_g) and crystallization temperature (T_c) for glasses and amorphous polymers. The thermal expansion coefficient (α_L) is yet another important material property that comes into full play in applications of electronic materials inasmuch as the thermal expansion mismatch is one of the main causes of electronic device failure. One of the most important thermal characterization tools is the differential scanning calorimeter (DSC), which enables the heat capacity, and various structural transition temperatures to be determined.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_19

Modulated-temperature DSC in which the sample temperature is modulated sinusoidally while being slowly ramped is a recent powerful thermal analysis technique that allows better thermal characterization and heat-capacity measurement. In addition, it can be used to measure the thermal conductivity. The present review is a selected overview of thermal properties and the DSC technique, in particular MTDSC. The overview is written from a materials science perspective with emphasis on phenomenology rather than fundamental physics.

The thermal properties of a large selection of materials can be found in various handbooks [19.1, 2]. In the case of semiconductors, *Adachi*'s book is highly recommended [19.3] since it provides useful relationships between the thermal properties for various group IV, II–V and II–VI semiconductors.

19.1 Heat Capacity

19.1.1 Fundamental Debye Heat Capacity of Crystals

The heat capacity of a solid represents the increase in the enthalpy of the crystal per unit increase in the temperature, usually, per unit mass of the crystal; in which case, it is called the *specific heat capacity*. The heat capacity is usually defined either at constant volume or at constant pressure, C_V and C_P, respectively. C_V represents the increase in the internal energy of the crystal when the temperature is raised because the heat added to the system increases the internal energy U without doing mechanical work by changing the volume. On the other hand C_P represents the increase in the enthalpy H of the system per unit increase in the temperature. Thus,

$$C_V = \left(\frac{\partial H}{\partial T}\right)_V = \left(\frac{\partial U}{\partial T}\right)_V \quad \text{and} \quad C_P = \left(\frac{\partial H}{\partial T}\right)_P . \tag{19.1}$$

The exact relationship between C_V and C_P is

$$C_V = C_P - \frac{9T\alpha_L^2}{\rho K} , \tag{19.2}$$

where T is the temperature, ρ is the density, α_L is the linear expansion coefficient, K is the compressibility and the heat capacities in (19.2) are for unit mass of the crystal (specific heat capacities). For solids, C_V and C_P are approximately the same. The increase in the internal energy U is due to an increase in the energy of lattice vibrations. This is generally true for all solids except metals at very low temperatures where the heat capacity is due to the conduction electrons near the Fermi level becoming excited to higher energies. For most practical temperature ranges of interest, the heat capacity of most solids is determined by the excitation of lattice vibrations. The molar heat capacity C_m is the increase in the internal energy U_m of a crystal of Avogadro's number N_A atoms per unit increase in the temperature at constant volume, that is, $C_m = (\mathrm{d}U_m/\mathrm{d}T)_V$. The Debye heat capacity is still the most successful model for understanding the heat capacity of crystals, and is based on the thermal excitation of lattice vibrations, that is phonons, in the crystal [19.4]; it is widely known as a conventional heat capacity model in many textbooks [19.5, 6]. The vibrational mean energy at a frequency ω is given by

$$\bar{E}(\omega) = \frac{\hbar\omega}{\exp\left(\frac{\hbar\omega}{k_B T}\right) - 1} , \tag{19.3}$$

where k_B is the Boltzmann constant. The energy $\bar{E}(\omega)$ increases with temperature. Each phonon has an energy of $\hbar\omega$ so that the phonon concentration in the crystal increases with temperature. To find the internal energy due to all the lattice vibrations we must also consider how many vibrational modes there are at various frequencies. That is, the distribution of the modes over the possible frequencies: the spectrum of the vibrations. Suppose that $\mathrm{g}(\omega)$ is the number of modes per unit frequency, that is, $\mathrm{g}(\omega)$ is the vibrational density of states or modes. Then $\mathrm{g}(\omega)\mathrm{d}\omega$ is the number of vibrational states in the range $\mathrm{d}\omega$. The internal energy U_m of all lattice vibrations for 1 mole of solid is,

$$U_m = \int_0^{\omega_{max}} \bar{E}(\omega)\mathrm{g}(\omega)\mathrm{d}\omega . \tag{19.4}$$

The integration is up to certain allowed maximum frequency ω_{max}. The density of states $\mathrm{g}(\omega)$ for the lattice vibrations in a periodic three-dimensional lattice, in its most highly simplified form, is given by

$$\mathrm{g}(\omega) \approx \frac{3V}{2\pi^2}\frac{\omega^2}{v^3} , \tag{19.5}$$

where v is the mean velocity of longitudinal and transverse waves in the solid and V is the crystal volume. The maximum frequency is ω_{max} and is determined by the fact that the total number of modes up to ω_{max} must be $3N_A$. It is called the *Debye frequency*. Thus, integrating $\mathrm{g}(\omega)$ up to ω_{max} we find,

$$\omega_{max} \approx v\left(\frac{6\pi^2 N_A}{V_m}\right)^{1/3} , \tag{19.6}$$

where V_m is the molar volume. This maximum frequency ω_{max} corresponds to an energy $\hbar\omega_{max}$ and to a temperature T_D defined by,

$$T_D = \frac{\hbar\omega_{max}}{k_B} , \tag{19.7}$$

and is called the *Debye temperature*. Qualitatively, it represents the temperature above which all vibrational frequencies are executed by the lattice waves.

Thus, by using (19.3) and (19.5) in (19.4) we can evaluate U_m and hence differentiate U_m with respect to temperature to obtain the molar heat capacity at constant volume,

$$C_m = 9R\left(\frac{T}{T_D}\right)^3 \int_0^{T_D/T} \frac{x^4 e^x dx}{(e^x - 1)} , \tag{19.8}$$

which is the well-known Debye heat capacity expression.

Figure 19.1 represents the constant-volume Debye molar heat capacity C_m for a perfect crystal (19.8) as a function of temperature, normalized with respect to the Debye temperature. The well-known classical *Dulong–Petit rule* ($C_m = 3R$) is only obeyed when $T > T_D$. The Dulong–Petit (DP) rule is a direct consequence of the applications of Maxwell's theorem of equipartition of energy and the classical kinetic molecular theory to vibrations of atoms in a crystal. Notice that C_m at $T = 0.5T_D$ is $0.825(3R)$, whereas at $T = T_D$ it is $0.952(3R)$. For most practical purposes, C_m is within 6% of $3R$ when the temperature is $0.9T_D$. For example, for copper $T_D = 315$ K and above about $0.9T_D$, that is above 283 K (or 10 °C), $C_m \approx 3R$, as borne out by experiments. Table 19.1 provides typical values for T_D, and heat capacities for a few selected elements. At the lowest temperatures, when $T \ll T_D$, (19.8) predicts that $C_m \propto T^3$ and this is indeed observed in low-temperature heat capacity experiments on a variety of crystals. On the other hand, well-known exceptions are glasses, noncrystalline solids, whose heat capacity is proportional to $a_1T + a_2T^3$, where a_1 and a_2 are constants (e.g., [19.6, 7]).

It is useful to provide a physical picture of the Debye model inherent in (19.8). As the temperature increases from near zero, the increase in the crystal's vibrational energy is due to more phonons being created and higher frequencies being excited. The phonon concentration increases as T^3 and the mean phonon energy increases as T. Thus, the internal energy increases as T^4. At temperatures above T_D, increasing the temperature creates more phonons but does not increase the mean phonon energy and does not excite higher frequencies. All frequencies up to ω_{max} have been now excited. The internal energy increases only due to more phonons being created. The phonon concentration and hence the internal energy increases as T; the heat capacity is constant, as expected from (19.8).

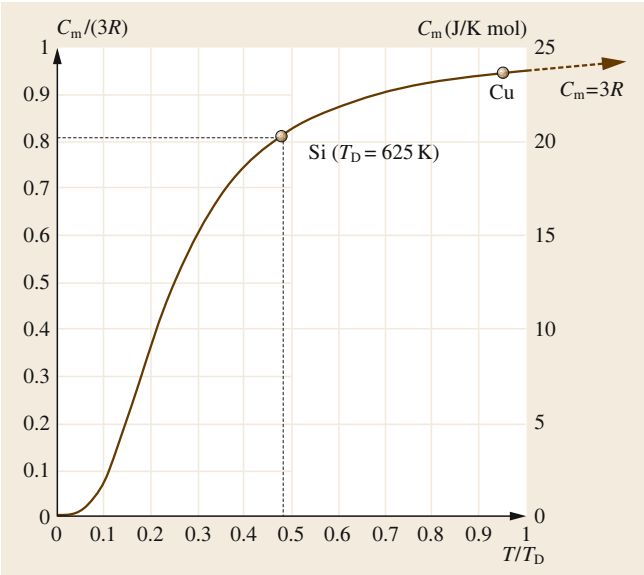


Fig. 19.1 Debye constant-volume molar heat capacity curve. The dependence of the molar heat capacity C_m on temperature with respect to the Debye temperature: C_m versus T/T_D. For Si, $T_D = 625$ K, so at room temperature (300 K), $T/T_D = 0.48$ and C_m is only $0.81(3R)$

In general terms, elements with higher atomic mass tend to have lower T_D, and their molar heat capacities are closer to the DP limit. Usually T_D is taken to depend on the temperature to accurately model the experimental C_P–T behavior; the dependence of T_D on T is only important at low temperatures, and can be neglected in most practical applications in engineering.

There are two important assumptions in the Debye theory. First is that all thermal excitations are lattice vibrations (i. e., involving phonons), which is usually the case in solids. Secondly, the interatomic potential-energy (PE) curve as a function of interatomic displacement x from equilibrium, is assumed to be parabolic, that is *harmonic* or symmetrical: $PE = PE_{min} + (1/2)\beta x^2$, where β is a constant (the spring constant in the equation of net force versus displacement, $F = -\beta x$). Within the harmonic approximation, the lattice vibrations are *plane waves* that are independent of each other; they are vibrational normal modes of the lattice. Further, a symmetrical PE curve does not allow thermal expansion, which means that C_V and C_P are identical. However, the actual interatomic PE is anharmonic, that is, it has an additional x^3 term. It is not difficult to show that in this case the vibrations or phonons interact. For example, two phonons can *mix* to

Table 19.1 Debye temperatures (T_D), heat capacities, thermal conductivities and linear expansion coefficients of various selected metals and semiconductors. C_m, c_s, κ, and α_L are at 25 °C. For metals, T_D is obtained by fitting the Debye curve to the experimental molar heat capacity data at the point $C_m = \frac{1}{2}(3R)$. T_D data for metals from [19.8]. Other data from various references, including [19.2] and the Goodfellow metals website

Metals	**Ag**	**Al**	**Au**	**Bi**	**Cu**	**Ga**	**Hg**	**In**	**Pd**	**W**	**Zn**
T_D (K)	215	394	170	120	315	240	100	129	275	310	234
C_m (J/K mol)	25.6	24.36	25.41	25.5	24.5	25.8	27.68	26.8	25.97	24.45	25.44
c_s (J/K g)	0.237	0.903	0.129	0.122	0.385	0.370	0.138	0.233	0.244	0.133	0.389
κ (W/m K)	420	237	317	7.9	400	40.6	8.65	81.6	71.8	173	116
α_L (K^{-1}) $\times 10^{-6}$	19.1	23.5	14.1	13.4	17	18.3	61	24.8	11	4.5	31
Semiconductors	**Diamond**	**Si**	**Ge**	**AlAs**	**CdSe**	**GaAs**	**GaP**	**InAs**	**InP**	**ZnSe**	**ZnTe**
T_D (K)	1860	643	360	450	135	370	560	280	425	340	260
C_m (J/K mol)	6.20	20.03	23.38	43.21	53.77	47.3	31.52	66.79	46.95	51.97	49.79
c_s (J/K g)	0.540	0.713	0.322	0.424	0.281	0.327	0.313	0.352	0.322	0.360	0.258
κ (W/m K)	1000	156	60	91	4	45	77	30	68	19	18
α_L (K^{-1}) $\times 10^{-6}$	1.05	2.62	5.75	4.28	7.43	6.03	4.89	5	4.56	7.8	8.33

generate a third phonon of higher frequency or a phonon can decay into two phonons of lower frequency etc. Further, the anharmonicity also leads to thermal expansion, so that C_V and C_P are not identical as is the case in the Debye model. As a result of the anharmonic effects, C_P continues to increase with temperature beyond the $3R$ Dulong–Petit rule, though the increase with temperature is usually small.

19.1.2 Specific Heat Capacity of Selected Groups of Materials

Many researchers prefer to quote the heat capacity for one mole of the substance, that is quote C_m, and sometime express C_m in terms of R. The limit $C_m = 3R$ is the DP rule. It is not unusual to find materials for which C_m can exceed the $3R$ limit at sufficiently high temperatures for a number of reasons, as discussed, for example, by *Elliott* [19.6]. Most applications of electronic materials require a knowledge of the specific heat capacity c_s, the heat capacity per unit mass. The heat capacity per unit volume is simply c_s/ρ, where ρ is the density. For a crystal that has only one type of atom with an atomic mass M_{at} (g/mol) in its unit cell (e.g., Si), c_s is C_m/M_{at} expressed in J/K g.

While the Debye heat capacity is useful in predicting the molar heat capacity of a crystal at any temperature, there are many substances, such as metals, both pure metals and alloys, and various semiconductors (e.g., Ge, CdSe, ZnSe etc.) and ionic crystals (e.g., CsI), whose room-temperature heat capacities approximately follow the simple DP rule of $C_m = 3R$, the limiting value in Fig. 19.1. For a metal alloy, or a compound such as $A_xB_yC_z$, that is made up of three components A, B and C with molar fractions x, y and z, where $x + y + z = 1$, the overall molar heat capacity can be found by adding individual molar heat capacities weighted by the molar fraction of the component,

$$C_m = xC_{mA} + yC_{mB} + zC_{mC}\,, \tag{19.9}$$

where C_{mA}, C_{mB} and C_{mA} are the individual molar heat capacities. Equation (19.9) is *the additive rule of molar heat capacities*. The corresponding specific heat capacity is

$$c_s = \frac{3R}{\bar{M}_{at}}\,, \tag{19.10}$$

where $\bar{M}_{at} = xM_A + yM_B + zM_C$ is the mean atomic mass of the compound, and M_A, M_B and M_C are the atomic masses of A, B and C. For example, for ZnSe, the average mass $\bar{M}_{at} = (1/2)(78.96 + 65.41) = 72.19$ g/mol, and the DP rule predicts $c_s = 3R/\bar{M}_{at} = 0.346$ J/K g, which is almost identical to the experimental value at 300 K.

The modern Debye theory and the classical DP rule that $C_m = 3R$ are both based on the addition of heat increasing the vibrational energy of the atoms or molecules in the solid. If the molecules are able to rotate, as in certain polymers and liquids, then the molar heat capacity will be more than $3R$. For example, the heat capacity of molten Bi−BiI_2 mixture is more than the DP rule, and the difference is due to the rotational energies associated with various molecular units in the molten state [19.9].

The heat capacity of polymers are usually quite different than the expected simple Debye heat capacity behavior. Most polymers are not fully crystalline but either amorphous or semicrystalline, a mixture of amorphous and crystalline regions. Secondly, and most significantly, the polymer structures usually have a main chain (a backbone) and various side groups attached to this backbone. Stretching and possible bending vibrations and possible partial rotations of these side groups

provide additional contributions to the overall heat capacity. The extent of freedom the main chain and side groups have in executing stretching and bending vibrations and rotations depends on whether the structure is above or below the glass-transition temperature T_g. (The glass-transition temperature is defined and discussed later.) Above T_g, the polymer structure is floppy and has sufficient free volume for various molecular motions to be able to absorb the added heat. The heat capacity is therefore larger above T_g than that below T_g, where the structure is frozen.

The heat capacity of most polymers increases with temperature. For most solid polymers (below the glass-transition temperature), the fractional rate of increase in the heat capacity with temperature follows

$$(1/C_P)(dC_P/dT) \approx 3 \times 10^{-3}\,K^{-1} \quad (T < T_g) \tag{19.11}$$

to within about 5%. This empirical rule is particularly useful for estimating C_P at a desired temperature from a given value at another temperature e.g., at room temperature. For liquid polymers (above T_g) on the other hand, $(1/C_P)(dC_P/dT)$ is less ($\approx 1 \times 10^{-3}\,K^{-1}$) and has a large variability (more than 30%) from polymer to polymer. There are various empirical rules for predicting the heat capacity of a polymer based on the backbone chain structure and types of side groups, as discussed by *van Krevelen* and *Hoftyzer* [19.10]. The most popular is to simply sum heat capacity contributions per mole from each distinguishable molecular unit in the polymer structure's repeat unit. The heat capacity contribution from each type of identifiable molecular unit (e.g., $-CH_2-$ in the backbone or $-CH_3$ side group) in polymer structures are listed in various tables. For example, the molar heat capacity of polypropylene can be calculated approximately by adding contributions from three distinguishable units, CH_2, CH and CH_3 in the repeating unit that makes up the polypropylene structure, as illustrated schematically in Fig. 19.2. Once the contributions are added, the sum is 71.9 J/K mol. The molecular mass of the repeat unit corresponds to three C atoms and six H atoms, or to 42 g/mol so that the specific heat is 71.9 J/K mol/42 g/mol or 1.71 J/K g, which is close to the experimental value. (Polypropylene is normally semicrystalline and its specific heat also depends on the crystalline-to-amorphous phase ratio.) The term *heat capacity per mole* in polymers normally refers to a mole of repeat units or segments in the polymer.

More rigorous characterization of the heat capacity of polymers tends to be complicated by the fact that most polymers are a mixture of the crystalline and amorphous phases, which have different C_P versus T characteristics, and the exact volume fractions of the phases may not be readily available. There have been a number of examples where the heat capacity characteristics of semicrystalline polymers have been critically examined [19.11]. A comprehensive overview of the heat capacity of polymers has been given by *Wunderlich* [19.12], which is highly recommended.

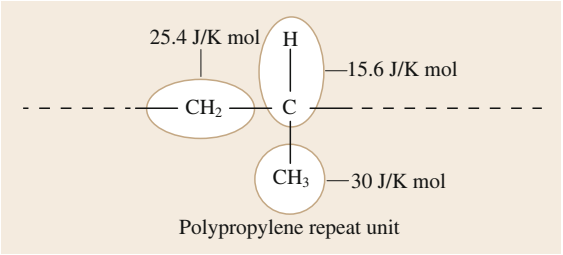


Fig. 19.2 Molar heat capacity of solid polypropylene calculated from various distinguishable units in the repeat unit

The overall or effective heat capacity of polymer blends can be usually calculated from the individual capacities and the volume fractions of the constituents. The simplest is a linear mixture rule in which individual capacities are weighted by the volume fraction and then summed.

The heat capacity of glasses depends on the composition and increases with temperature. There is a sharp change in the dependence of C_P on T through the glass-transition temperature T_g. For example, for zirconium, barium, lanthanum, aluminum, sodium (ZBLAN) glasses, C_m is 24 J/K mol below T_g and slightly less than $3R$. On the other hand, C_m is 38.8 J/K mol above T_g, and greater than $3R$. (The C_P change at T_g is discussed below.) The dependence of C_P on T below T_g has been modeled by a number of expressions dating back to the 1950s. For example, *Sharp* and *Ginther* [19.13] in 1951 proposed the following empirical expressions for the mean and true specific heat capacity c_m and c respectively, of oxide glasses in the range 0–1300 °C,

$$c_m = \frac{H_2 - H_1}{T_2 - T_1} = \frac{aT + c_o}{0.00146T + 1} \quad \text{and} \tag{19.12a}$$

$$c = \frac{dH}{dT} = \frac{aT + c_m}{0.00146T + 1} \tag{19.12b}$$

in which T is in °C, and a and c_o are calculated from the glass composition as described in the original paper [19.13]. The mean heat capacity in (19.12a) then predicts the heat capacity to within about 1%. More recently, for various oxide glasses from 300 K to T_g, *Inaba* and *Oda* [19.14] have proposed an exponential

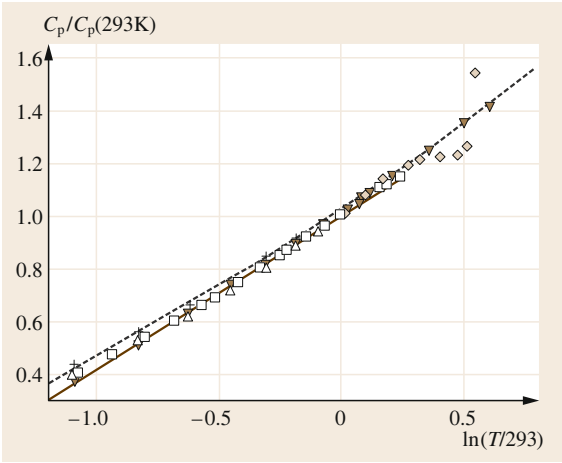


Fig. 19.3 Normalized heat capacity $C_P(T)/C_P(298)$ versus $\ln(T/298)$ for various glasses. *Open square:* $12.5Na_2O$, $12.5Al_2O_3$, $75SiO_2$; *solid circle:* 25MgO, 25CaO, $50SiO_2$; *open triangle:* $33Na_2O$, $67B_2O_3$; *solid triangle:* 24BaO, $1Fe_2O_3$, $75B_2O_3$; *open diamond:* B_2O_3 (includes T_g); $+30K_2O$, $70GeO_2$. Adapted from *Khalimovskaya-Churkina* and *Priven* [19.15]. The *solid straight line* is (19.14), whereas the *dashed line* is the experimental trend line

empirical expression of the form,

$$C_P = 3R\left[1-\exp\left(-1.5\frac{T}{T_D}\right)\right], \qquad (19.13)$$

where T_D is the Debye temperature. Equation (19.13) predicts a lower C_P than the Debye equation itself at the same temperature for the same T_D. There are also other empirical expressions for C_P (for example, a three-term polynomial in T) that typically involve at least three fitting parameters. *Khalimovskaya-Churkina* and *Priven* [19.15] have shown that a simple two-parameter empirical expression of the form

$$C_P = a + b\ln T\,, \qquad (19.14)$$

where a and b are composition-dependent parameters, can be fruitfully used to represent the dependence of C_P on T from 100 K to the lower boundary of T_g for various oxide glasses to within about 2.5% error as shown in Fig. 19.3. The authors further show that, to within about 4%,

$$C_P = C_P(293\,K)\,[1+0.52\ln(T/293)]\,. \qquad (19.15)$$

What is significant in this equation is that the composition dependence of C_P has been incorporated into $C_P(293\,K)$, and once the latter is calculated, the temperature dependence of C_P is then given by (19.15).

Various empirical rules have been proposed for approximately calculating the heat capacity of oxide glasses from its components. The simplest is the additive rule stated in (19.9), in which the overall heat capacity per mole is found from adding individual molar heat capacities of the constituents, so-called partial molar heat capacities, weighted by their molar fractions. However, for a heterogeneous mixture, that is for a solid in which there are two or more distinct phases present, the specific heat capacity may involve more complicated calculations.

The most common technique for heat capacity measurements is the heat-flow differential scanning calorimetry (DSC) – the instruments are relatively cheap and easy to operate, there is no specialized sample preparation, and the measurements are fast. Nonetheless, the lowest achieved uncertainties associated with the C_P values are approximately 1.5% even for standardized materials data obtained under ideal conditions using specialized DSC instruments [19.16] (typical uncertainties vary between 3 and 5%). Since the heat capacity belongs to the most important set of thermodynamic quantities (entropy, enthalpy, and free Gibbs energy can be calculated from the C_P–T dependences), often an increased need for maximum precision and accuracy arises. This is especially true for the low-temperature C_P data (0−100 K), where specialized experimental techniques like adiabatic calorimetry and relaxation calorimetry have to be used.

The most common and easy-to-apply methodology utilized for obtaining the C_P–T dependences by means of DSC measurements is the so-called *continuous method* [19.17]. This method is based on the comparison of two identical heating scans performed for the studied sample and referent material with known C_P–T dependence (sapphire). The heat-flow signal obtained for the studied sample can then be transformed to heat capacity according to the following equation (both measurements need to be corrected for a zero-line prior to the calculation)

$$C_P^{sample} = C_P^{ref}\frac{\Phi^{sample}m^{ref}}{\Phi^{ref}m^{sample}}\,, \qquad (19.16)$$

where C_P, Φ, and m correspond to the heat capacity, raw heat-flow gained from the DSC measurement, and mass of the measured material, respectively. The superscripts denote the sample and reference. Note that, while in this way the whole DSC curve can be easily transformed, true heat capacity is obtained only in the *thermodynamic regions*, where no overlapping kinetic effect (phase transformations or chemical reactions) occurs on the measured DSC curve.

In the case of a narrow temperature range, the obtained C_P–T dependence can be usually described

by means of a second-degree polynomial function, whereas for a wider temperature ranges more complicated functions need to be used. Below, an ultimate equation for description of the C_P–T dependences is shown

$$C_P = a + bT + cT^2 + dT^{-1} + eT^{-2} + f \ln T + gT \ln T \,. \quad (19.17)$$

The importance of the particular terms is usually determined by means of numerical optimization and consequent statistical analysis. The most commonly applied equations include the following combinations of the particular terms (denoted according to the corresponding composition-dependent parameters): $a+b+c$, $a+b+e$, $a+b+c+e$, $a+b+c+d+e$, and $a+b+c+f$.

19.2 Thermal Conductivity

19.2.1 Definition and Typical Values

Heat conduction in materials is generally described by *Fourier's heat conduction law*. Suppose that J_x is the heat flux in the x-direction, defined as the quantity of heat flowing in the x-direction per unit area per unit second: the thermal energy flux. Fourier's law states that the heat flux at a point in a solid is proportional to the temperature gradient at that point and the proportionality constant depends on the material,

$$J_x = -\kappa \frac{\mathrm{d}T}{\mathrm{d}x} \,, \quad (19.18)$$

where κ is a constant that depends on the material, called the thermal conductivity (W/m K or W/m °C), and $\mathrm{d}T/\mathrm{d}x$ is the temperature gradient. Equation (19.18) is called Fourier's law, and effectively defines the thermal conductivity of a medium. Table 19.2 provides an overview of typical values for the thermal conductivity of various classes of materials.

The thermal conductivity depends on how the atoms in the solid transfer the energy from the hot region to the cold region. In metals, the energy transfer involves the conduction electrons. In nonmetals, the energy transfer involves lattice vibrations, that is atomic vibrations of the crystal, which are described in terms of phonons.

The thermal conductivity, in general, depends on the temperature. Different classes of materials exhibit different κ values and dependence of κ on T, but we can generalize very roughly as follows:

- *Most pure metals*: $\kappa \approx 50$–400 W/m K. At sufficiently high T, e.g., above ≈ 100 K for copper, $\kappa \approx$ constant. In magnetic materials such as iron and nickel, κ decreases with T.
- *Most metal alloys*: κ lower than for pure metals; $\kappa \approx 10$–100 W/m K. κ increases with increasing T.
- *Most ceramics*: Large range of κ, typically 10–200 W/m K with diamond and beryllia being exceptions with high κ. At high T, typically above ≈ 100 K, κ decreases with increasing T.
- *Most glasses*: Small κ, typically less than $\approx$ 5 W/m K and increases with increasing T. Typical examples are borosilicate glasses, window glass, soda-lime glasses, fused silica etc. Fused silica is noncrystalline SiO_2 with $\kappa \approx 2$ W/m K.
- *Most polymers*: κ is very small and typically less than 2 W/m K and increases with increasing T. Good thermal insulators.

Table 19.2 Typical thermal conductivities of various classes of materials at 25 °C

Pure metals	**Nb**	**Sn**	**Fe**	**Zn**	**W**	**Al**	**Cu**	**Ag**
κ (W/mK)	53	64	80	116	173	237	400	420
Metal alloys	**Stainless steel**	**55Cu–45Ni**	**Manganin (86Cu–12Mn–2Ni)**	**70Ni–30Cu**	**1080 Steel**	**Bronze (95Cu–5Sn)**	**Brass (63Cu–37Zn)**	**Dural (95Al–4Cu–1Mg)**
κ (W/mK)	12–16	19.5	22	25	50	80	125	147
Ceramics and glasses	**Glass-borosilicate**	**Silica-fused (SiO_2)**	**S_3N_4**	**Alumina (Al_2O_3)**	**Magnesia (MgO)**	**Saphire (Al_2O_3)**	**Beryllia (BeO)**	**Diamond**
κ (W/mK)	0.75	1.5	20	30	37	37	260	1000
Polymers	**Poly-propylene**	**Polystyrene**	**PVC**	**Poly-carbonate**	**Nylon 6,6**	**Teflon**	**Polyethylene low density**	**Polyethylene high density**
κ (W/mK)	0.12	0.13	0.17	0.22	0.24	0.25	0.3	0.5

19.2.2 Thermal Conductivity of Crystalline Insulators

In nonmetals heat transfer involves lattice vibrations, that is phonons. The heat absorbed in the hot region increases the amplitudes of the lattice vibrations which is the same as generating more phonons. These new phonons travel towards the cold regions and thereby transport the lattice energy from the the hot to cold region. The thermal conductivity κ measures the rate at which heat can be transported through a medium per unit area per unit temperature gradient. It is proportional to the rate at which a medium can absorb energy, that is, κ is proportional to the heat capacity. κ is also proportional to the rate at which phonons are transported which is determined by their mean velocity v_{ph}. In addition, of course, κ is proportional to the mean free path Λ_{ph} that a phonon has to travel before losing its momentum, just as the electrical conductivity is proportional to the electron's mean free path. A rigorous classical treatment gives κ as,

$$\kappa = \frac{1}{3} C_v v_{ph} \Lambda_{ph} \,, \tag{19.19}$$

where C_v is the heat capacity per unit volume. The mean phonon velocity v_{ph} is constant and approximately independent of temperature so that κ is governed primarily by C_v and Λ_{ph}. The mean free path Λ_{ph} depends on various processes that can scatter the phonons and hinder their propagation along the direction of heat flow. Phonons can collide with other phonons, crystal defects, impurities and crystal surfaces.

At high temperatures, κ is controlled by phonon–phonon scattering. The phonon–phonon collisions that are normally responsible for limiting the thermal conductivity, that is scattering the phonon momentum in the opposite direction to the heat flow, are due to the anharmonicity (asymmetry) of the interatomic potential energy curve. Stated differently, the net force F acting on an atom is not simply βx (where β is a spring constant) but also has an x^2 term, i. e., it is nonlinear. The greater the asymmetry or nonlinearity, the larger is the effect of such momentum-flipping collisions. The same asymmetry that is responsible for thermal expansion of solids is also responsible for determining the thermal conductivity. When two phonons 1 and 2 interact in a crystal region as in Fig. 19.4, the nonlinear behavior and the periodicity of the lattice, cause a new phonon 3 to be generated. This new phonon 3 has the same energy as the sum of 1 and 2 but it is traveling in the wrong direction. (The frequency of 3 is the sum of frequencies of 1 and 2.) The flipping of the phonon momentum is called an *Umklapp process* and is a result of the anharmonicity in the interatomic bond. Figure 19.5 shows the dependence of κ on T for three semiconductors. We can identify three types of behavior.

At low temperatures there are two factors that control κ. The phonon concentration is too low for phonon–phonon collisions to be significant. Instead, the mean free path Λ_{ph} is determined by phonon collisions with crystal imperfections, most significantly, crystal surfaces and grain boundaries. Thus, Λ_{ph} depends on the sample geometry and crystallinity. Further, as we expect from the Debye model, C_v depends on T^3 so that κ has the same temperature dependence as C_v, that is $\kappa \propto T^3$. Figure 19.5 shows the dependence of κ on T for Ge, Si and GaAs. Notice that κ increases with increasing T at low temperatures, and that κ also depends on the concentration of dopants (GaAs) or impurities in the crystal, which can scatter phonons. An extensive

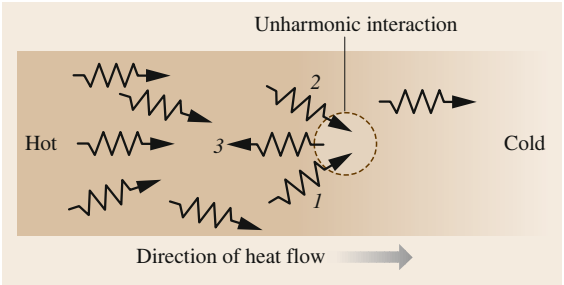


Fig. 19.4 Phonons generated in the hot region travel towards the cold region and thereby transport heat energy. Phonon–phonon anharmonic interaction generates a new phonon whose momentum is towards the hot region. (After [19.5])

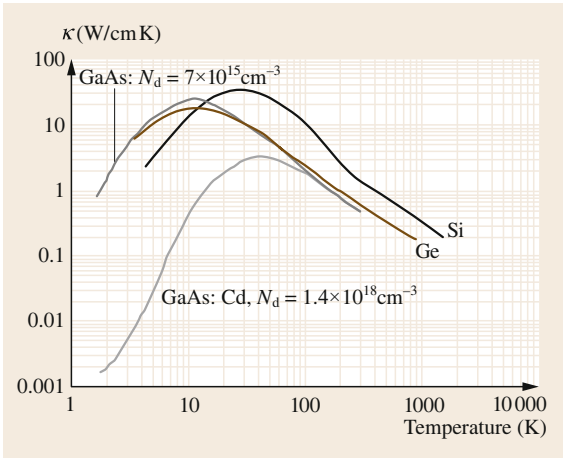


Fig. 19.5 Thermal conductivity of Si, Ge and GaAs. GaAs for two different levels of doping. Data extracted and replotted: Si [19.18]; Ge [19.19]; GaAs [19.20]

treatise on the thermal conductivity of semiconductors is given by *Bhandari* and *Rowe* [19.21].

As the temperature increases, the temperature dependence of κ becomes controlled by Λ_{ph} rather than C_v, which changes only slowly with T. The mean free path becomes limited by phonon–phonon collisions that obey the Umklapp process. These phonons have energies $\approx \frac{1}{2}k_B T_D$. The concentration of such phonons is therefore proportional to $\exp(T_D/2T)$. Since Λ_{ph} is inversely proportional to the Umklapp-obeying phonon concentration, κ decreases with increasing temperature, following a $\kappa \propto \exp(T_D/2T)$ type of behavior. At temperatures above the Debye temperature, C_v is, of course, constant, and the phonon concentration n_{ph} increases with temperature, $n_{ph} \propto T$. Thus, the mean free path decreases as $\Lambda_{ph} \propto 1/T$, which means that $\kappa \propto T^{-1}$, as observed for most crystals at sufficiently high temperatures. Figure 19.6 illustrates the expected dependence of κ on T for nonmetals, with three temperature regimes

$$\kappa \propto \begin{cases} T^3\,; & \text{Low } T \\ \exp\left(\dfrac{-T_D}{2T}\right)\,; & \text{Intermediate } T \\ T^{-1}\,; & \text{High } T\,. \end{cases} \qquad (19.20)$$

Most crystalline semiconductors tend to follow this type of behavior. The thermal conductivity of important semiconductors has been discussed by *Adachi* [19.3]. In the temperature range above the peak of κ with T, Adachi used a $\kappa = AT^n$ type of power law to model the κ–T data, and has provided an extensive table for A and n for various group IV, III–V and II–V semiconductors [19.3]. For example, for GaAs, over $150 \le T \le 1500$, $A = 750$ and $n = -1.28$ so that $\kappa\,(\mathrm{W/cm\,K}) = (750)T^{-1.28}$, where T is in Kelvin. Further, Adachi has been able to correlate κ with the lattice parameter a, mean atomic mass $\bar{M}$ in the unit cell and the Debye temperature T_D.

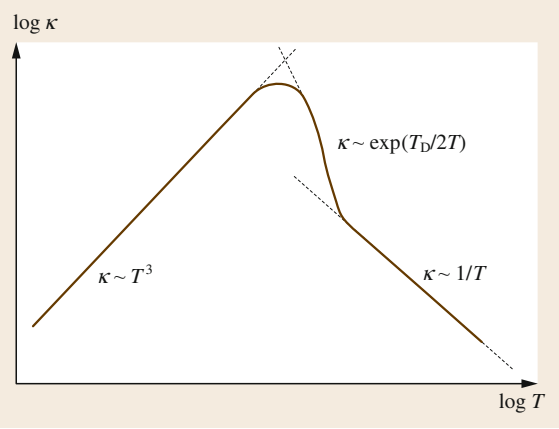


Fig. 19.6 Lattice thermal conductivity versus temperature on a log–log plot. κ is due to phonons

19.2.3 Thermal Conductivity of Noncrystalline Insulators

Glasses

The thermal conductivity of glasses tends to be substantially smaller than their crystalline counterparts, and tends to increase with the temperature almost monotonically. The most striking difference at room temperature is that, as the temperature is decreased, κ for a glass decreases whereas for a crystal it increases. This difference has been qualitatively attributed to the phonon mean free path Λ_{ph} in a glass remaining roughly constant as the temperature is lowered and being limited by the range of disorder in the structure, which is a few atomic spacings [19.23, 24]. While intuitively this explanation seems reasonable, a more rigorous derivation excludes (19.19) since concepts like phonon velocity and wavelength tend to become meaningless over very short distances [19.25]. Most of the theories that have been proposed for describing κ more rigorously in the noncrystalline state are quite mathematical, and beyond the scope of this review based on materials science. (These models include the Kubo transport equation [19.26], hopping of localized vibrations [19.27], simulations based on molecular dynamics [19.28] etc.) Figure 19.7 shows κ versus T for a few selected glasses [19.22, 29]. Other glasses

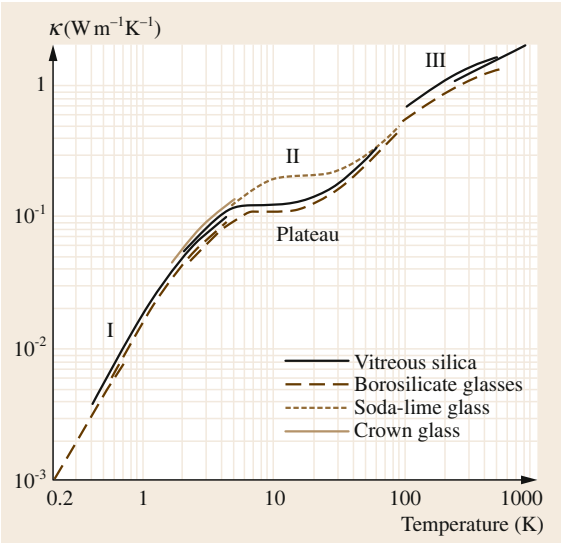


Fig. 19.7 The temperature dependence of κ for four types of glasses. Data in Figure 2 of *Zeller* and *Pohl* [19.22] replotted

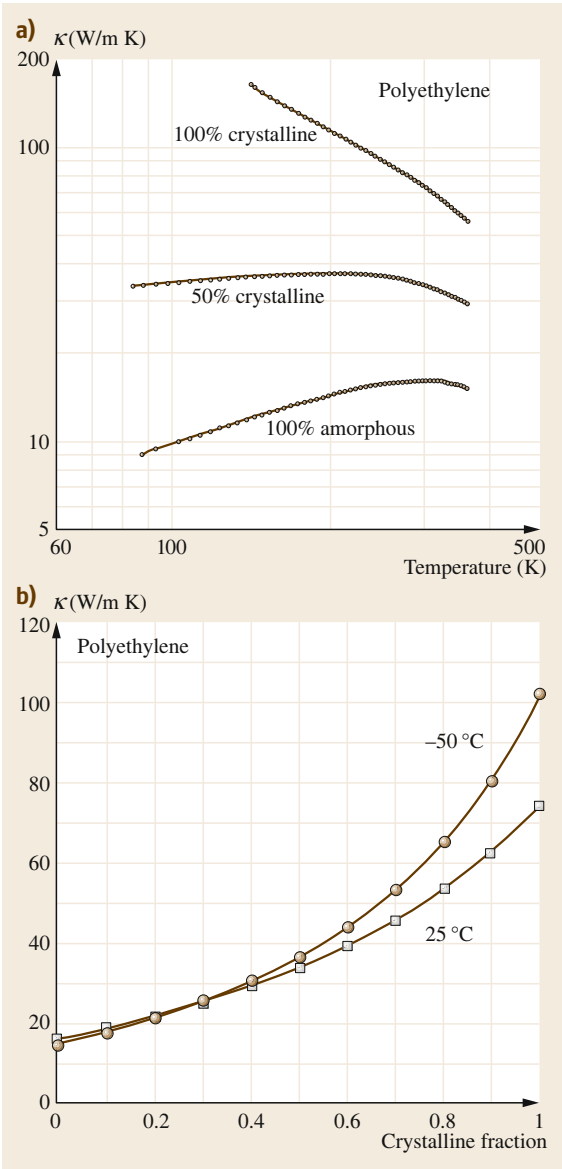


Fig. 19.8 **(a)** κ–T behavior for polyethylene for a fully crystalline, fully amorphous and 50% crystalline structures. **(b)** Dependence of κ on the crystallinity of polyethylene. Data extracted and replotted from [19.30]

tend to follow this type of behavior in which there are three regions of distinct behavior. At very low temperatures (I in Fig. 19.7), below 1–2 K, $\kappa \approx T^n$ where $n = 1.8$–2.0. In the intermediate temperature regime, from about 2 K to about 20–30 K, κ exhibits a plateau. At high temperatures, above 20–30 K, beyond the plateau, κ increases with T, with a decreasing slope.

Polymers

Crystalline polymers tend to follow the general predictions in Fig. 19.6 that at high temperatures κ decreases with the temperature. As the degree of crystallinity decreases, κ also decreases, and eventually for amorphous polymers κ is substantially smaller than the crystalline phase, and κ increases with increasing temperature. If ρ_c and ρ_a are the densities of a given polymer in the crystalline (c) and amorphous (a) phase, then various empirical rules have been proposed to relate the conductivities of c- and a-phases through the ratio ρ_c/ρ_a of their densities. For example, in one empirical rule [19.30]

$$\frac{\kappa_c}{\kappa_a} = 1 + 1.58\left[\frac{\rho_c}{\rho_a} - 1\right] .$$

This means that, if we know κ_c, and ρ_c/ρ_a, we can calculate κ_a. If we know the density ρ of the partially crystalline polymer or the volume fraction v_a of the amorphous phase, then we can use an appropriate mixture rule to find the overall κ. (A solved problem is given as an example in [19.10, Chap. 17].) Figure 19.8 shows the dependence of κ on T for three different degrees of crystallinity in the structure; clearly κ versus T is strongly influenced by the crystallinity of the structure.

Figure 19.9 shows the temperature dependence of κ when it has been normalized to its value $\kappa(T_g)$ at the glass-transition temperature T_g against the normalized temperature, T/T_g for a number of amorphous polymers. While there are marked differences between various polymers, there is nonetheless an overall trend in which κ increases up to T_g and thereafter decreases with T. The solid curve in Fig. 19.9 indicates an overall trend. Notice how similar the κ values are.

The overall or effective thermal conductivity of polymer blends can be usually estimated from the individual conductivities, the volume or weight fractions of the constituents, and the crystallinity of the structure; the problem can be quite complicated. A good example in which κ has been examined for a polymer blend as a function of blend composition has been given by *Agari* et al. [19.31], who also test various mixture rules.

The thermal conductivity of polymeric composite materials (as opposed to blends) can be estimated from various mixture rules [19.32, 33]. For example, if a matrix with κ_1 contains spherical particles with κ_2, then the composite has a thermal conductivity,

$$\kappa = \kappa_1 \frac{2\kappa_1 + \kappa_2 - 2v_2(\kappa_1 - \kappa_2)}{2\kappa_1 + \kappa_2 + v_2(\kappa_1 - \kappa_2)} , \qquad (19.21)$$

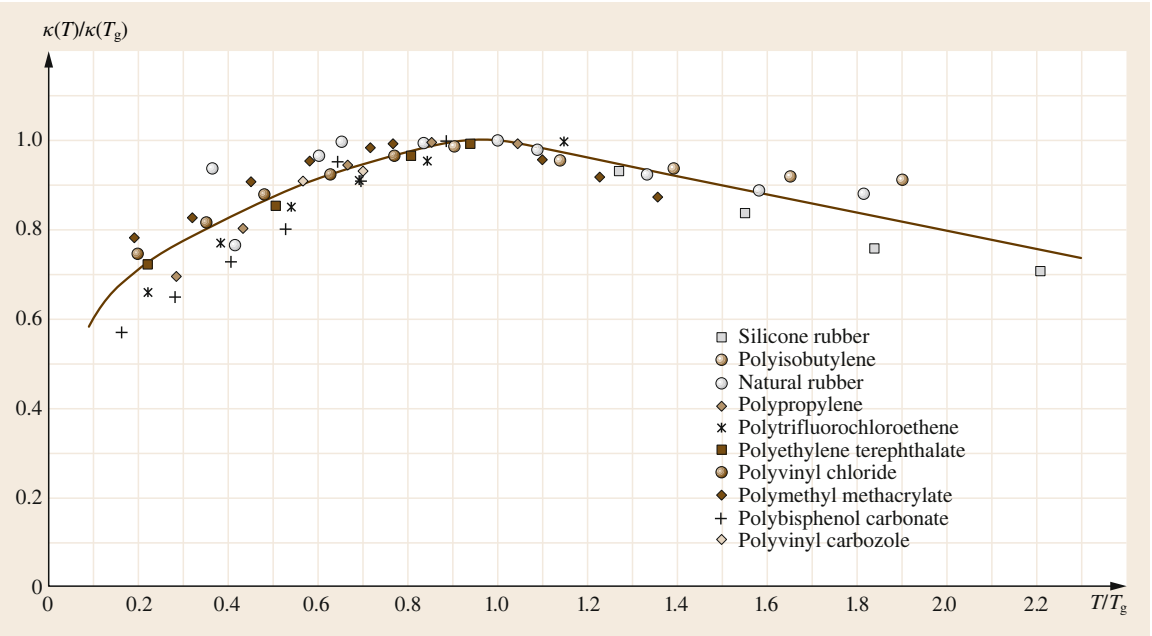


Fig. 19.9 Normalized thermal conductivity, $\kappa(T)/\kappa(T_g)$ versus normalized temperature, T/T_g, for various polymers. (After [19.10])

where v_2 is the volume fraction of the particles (with κ_2).

19.2.4 Thermal Conductivity of Metals

Although the heat capacity of metals is due the lattice vibrations, the thermal conductivity is due to the free conduction electrons. Both charge transport and heat transport are carried by the conduction electrons, and hence the electrical and thermal conductivity, σ and κ respectively, are related. Normally only electrons around the Fermi level contribute to conduction and energy transport. In general κ and σ can be written as

$$\kappa = \frac{1}{3} C_e v_F \ell_e \quad \text{and} \quad \sigma = \frac{1}{3} e^2 v_F \ell_e g(E_F) \,, \tag{19.22}$$

where C_e is the electronic heat capacity per unit volume, v_F is the Fermi velocity, $g(E_F)$ is the density of states at E_F (Fermi level), and ℓ_e is the electron mean free path. The scattering of electrons limits ℓ_e, which then limits κ and σ. If we then apply the free-electron theory to substitute for C_e and $g(E_F)$ then the ratio of κ to σT turns out to be a constant, which depends only on the constants k_B and e; ℓ_e cancels out in this ratio. A proper derivation involves solving the Boltzmann transport equation in the presence of a temperature gradient [19.34], and leads to the well-known *Wiedemann–Franz–Lorenz* (WFL) law

$$\frac{\kappa}{\sigma T} = C_{WFL} = \frac{\pi^2 k_B^2}{3e^2} = 2.44 \times 10^{-8} \frac{W\,\Omega}{K^2} \,, \tag{19.23}$$

where C_{WFL} is called the *Lorentz number*.

As shown in Fig. 19.10, the WFL law is well obeyed at room temperature for most pure metals and their alloys. The WFL law is normally well obeyed over a wide temperature range, and also by many liquid pure metals. Consequently the thermal conductivity κ of a metal or an alloy can be readily calculated given the conductivity or resistivity ρ. For example, we can easily calculate the resistivity of a solid solution alloy by using Matthiessen's and Nordheim's rules (as explained in this handbook by *Kasap* et al. in Chap. 2) and then simply calculate κ from (19.23). Figure 19.11 shows the typical dependence of κ on T for two pure metals (Cu and Al) and two alloys (brass and Al–14% Mg). Above the Debye temperature, for pure metals, $\sigma \propto 1/T$, so that the thermal conductivity remains almost constant, as in Fig. 19.11. For alloys, σ values are smaller than the corresponding pure metals and σ exhibits a weaker temperature dependence. Consequently κ is smaller and increases with the temperature, as shown in Fig. 19.11.

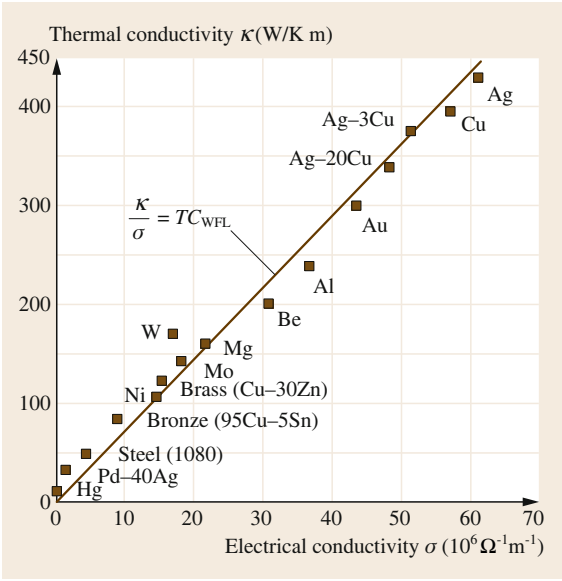


Fig. 19.10 Thermal conductivity versus electrical conductivity for various metals (both pure metals and alloys) at 20 °C. (After [19.5])

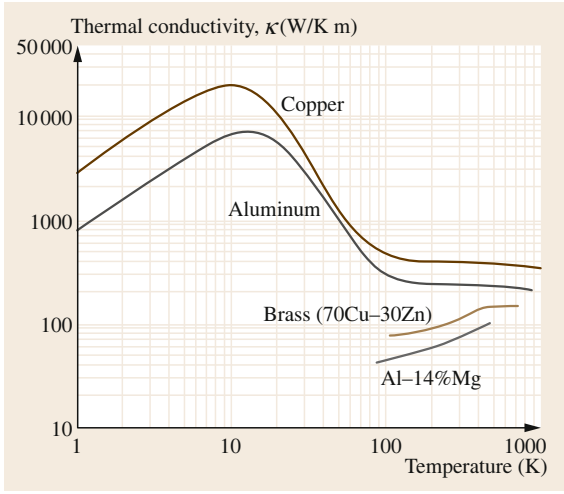


Fig. 19.11 Thermal conductivity versus temperature for two pure metals (Cu and Al) and two alloys (brass and Al–14% Mg). (After [19.5])

While for pure metals and many alloys κ is due to the conduction electrons, the phonon contribution can nonetheless become comparable to the electron contribution for those alloys that have substantial resistivity or at low temperatures. Glassy metals (or metallic glasses) usually have two or more elements alloyed; one of the elements is usually a nonmetal or a metalloid. Their resistivities tend to be higher than usual crystalline or polycrystalline alloys. Both conduction electrons and also phonons contribute to the thermal conductivity of these glassy metals. For example, for $Fe_{80}B_{20}$, $\rho = 1220\,\mathrm{n\Omega\,m}$ (compare with $17\,\mathrm{n\Omega\,m}$ for Cu), and the phonon contribution is about 15% at 273 K and 61% at 100 K [19.35].

19.3 Thermal Expansion

19.3.1 Grüneisen's Law and Anharmonicity

Nearly all materials expand as the temperature increases. The thermal coefficient of linear expansion, or simply, the thermal expansion coefficient α is defined as the fractional change in length per unit temperature increase, or

$$\alpha_\mathrm{L} = \frac{1}{L_0}\left(\frac{\mathrm{d}L}{\mathrm{d}T}\right)_{T_0}, \qquad (19.24)$$

where L_0 is the original length of the substance at temperature T_0, $\mathrm{d}L/\mathrm{d}T$ is the rate of change of length with temperature at the reference temperature T_0.

The principle of thermal expansion is illustrated in Fig. 19.12, which is the potential-energy curve $U(r)$ for two atoms separated by a distance r in a crystal. At temperature T_1, the atoms would be vibrating about their equilibrium positions between B and C; compressing (B) and stretching (C) the bond between them. The line BC corresponds to the total energy E of the pair of atoms. The average separation at T_1 is at A, halfway between B and C. The PE curve $U(r)$ is asymmetric, and it

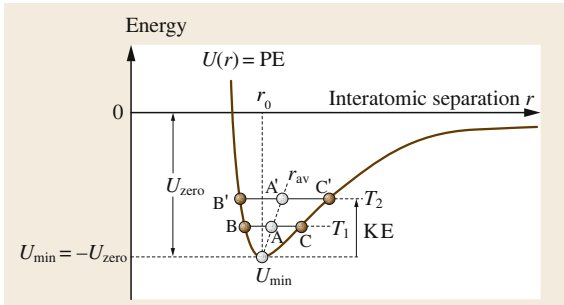


Fig. 19.12 Interatomic potential energy (PE) U curve, which has a minimum when the atoms in the solid attain the interatomic separation at $r = r_0$

Table 19.3 The Grüneisen parameter for some selected materials with different types of interatomic bonding. Only typical values are used; and for certain materials there can be significant variation in quoted data in the literature. Data extracted from [19.5]. (FCC is face-centered cubic and BCC is body-centered cubic crystal structures.)

Material	ρ (g/cm^{-3})	α_L ($\times 10^{-6}\,K^{-1}$)	K (GPa)	c_s (J/kg K)	γ
Iron (metallic, BCC)	7.9	12	170	450	1.7
Copper (metallic, FCC)	8.96	17	140	385	2.1
NaCl (ionic)	2.17	44	25	850	1.8
CsI (ionic)	4.51	48	13	201	2.1
Germanium (covalent)	5.32	6	77	322	0.81
Silicon (covalent)	2.32	2.6	99	703	0.47
Glass (covalent-ionic)	2.45	8	50	800	0.61
ZnSe (ionic/covalent)	5.27	7.4	62	350	0.75
Tellurium (covalent/van der Waals)	6.24	17	30	200	1.23
Polystyrene (van der Waals)	1.1	80	≈ 3	1300	0.50
Polyethylene terephthalate PET (van der Waals)	1.38	70	≈ 3	1200	0.38

is this asymmetry that leads to the phenomenon of thermal expansion. When the temperature increases from T_1 to T_2 the atoms vibrate between B' and C' and the average separation between the atoms also increases, from A to A', which is identified as thermal expansion. If the PE curve were symmetric, then there would be no thermal expansion. The extent of expansion (A to A') depends on the amount of increase from BC to $B'C'$ per degree of increase in the temperature. α must therefore also depend on the heat capacity. When the temperature increases by a small amount δT, the energy per atom increases by $(C_v \delta T)/N$, where C_v is the heat capacity per unit volume and N is the number of atoms per unit volume. If $C_v \delta T$ is large then the line $B'C'$ in Fig. 19.12 will be higher up on the energy curve and the average separation A' will therefore be larger. Thus, $\alpha \propto C_v$. Further, the average separation, point A, depends on how much the bonds are stretched and compressed. For large amounts of displacement from equilibrium, the average A will be greater as more asymmetry of the PE curve is used. Thus, smaller is the elastic modulus K, the greater is α_L; clearly $\alpha_L \propto C_v/K$.

If we were to expand $U(r)$ about its minimum value U_{min} at $r = r_0$, we would obtain the Taylor expansion to the cubic term as,

$$U(r) = U_{min} + \frac{1}{2}\beta(r-r_0)^2 - \frac{1}{3}g(r-r_0)^3 \,, \quad (19.25)$$

where β and g are coefficients related to second and third derivatives of U at r_0. The term $(r-r_0)$ is missing because we are expanding a series about U_{min} where $dU/dr = 0$. The U_{min} and the $\beta_2(r-r_0)^2$ term give a parabola about U_{min} which is a symmetric curve around r_0 and therefore does not lead to thermal expansion. It is the cubic term that gives the expansion because it leads to asymmetry. Thus, the amount of expansion α also depends on the amount of asymmetry with respect to symmetry that is β/g. Thus, α is proportional to both C_v/K and g/β. The third term in (19.25) is called *anharmonicity* in the interatomic PE curve. The ratio of g to β depends on the nature of the bond. A simplified analytical treatment gives α_L as,

$$\alpha_L \approx \frac{\gamma}{3}\frac{C_v}{K} \,; \quad \gamma = \frac{r_0 g}{3\beta} \,, \quad (19.26)$$

where γ is a constant called the *Grüneisen parameter*, which represents the relative asymmetry of the energy curve, g/β. γ is of the order of unity for many materials; experimentally, $\gamma = 0.1{-}2$. We can also write the Grüneisen law in terms of the molar heat capacity C_m (heat capacity per mole or the specific heat capacity c_s (heat capacity per unit mass)). If ρ is the density, and M_{at} is the atomic mass of the constituent atoms of the crystal, then

$$\alpha_L = \frac{\gamma}{3}\frac{\rho C_m}{M_{at} K} = \frac{\gamma}{3}\frac{\rho c_s}{K} \,. \quad (19.27)$$

Given the experimental values for α_L, K, ρ and c_s, the Grüneisen parameters have been calculated from (19.27) and are listed in Table 19.3. An interesting feature is that the experimental γ values, within a factor of 2–3, are about the same, at least to an order of magnitude. Equation (19.26) also indicates that the dependence of α_L on T should resemble the dependence of C_v on T, which is borne out by experiments.

19.3.2 Thermal Expansion Coefficient α_L

The thermal expansion coefficient normally depends on the temperature, $\alpha_L = \alpha_L(T)$, and typically increases

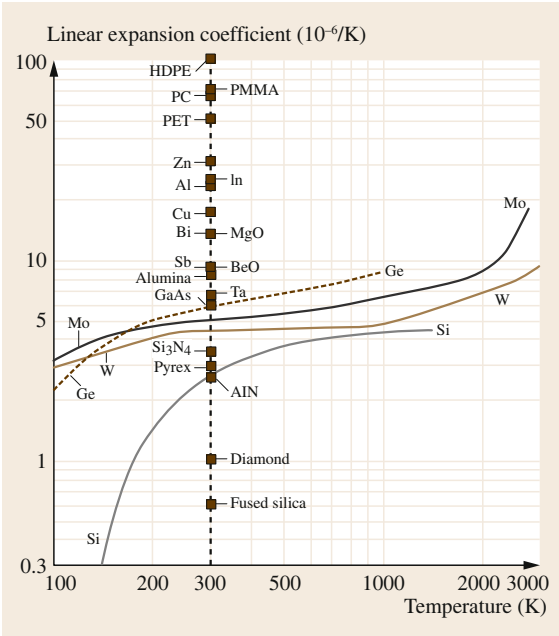


Fig. 19.13 α versus T for selected materials and comparison of α_L for a range of materials at 300 K. (After [19.5])

with increasing temperature, except at the lowest temperatures. Figure 19.13 shows α_L versus T for selected materials and also compares α_L for a wide range of materials at 300 K. In very general terms, except at very low (typically below 100 K) and very high temperatures (near the melting temperature), for most metals α_L does not depend strongly on the temperature; many engineers take α_L for a metal to be approximately temperature independent. Further, higher-melting-point temperature metals tend to have lower α_L.

To calculate the final length L from the original length L_0 that corresponds to the temperature increasing from T_0 to T, we have to integrate $\alpha_L(T)$ with respect to temperature from T_0 to T. If $\bar{\alpha}_L$ is the mean value for the expansion coefficient from T_0 to T, then

$$L = L_0[1 + \bar{\alpha}_L(T - T_0)] \text{ and} \tag{19.28}$$

$$\bar{\alpha}_L = \frac{1}{(T - T_0)} \int_{T_0}^{T} \alpha_L(T)\mathrm{d}T \tag{19.29}$$

For example, the expansion coefficient of silicon over the temperature range 120–1500 K is given by [19.36] as

$$\alpha_L = 3.725 \times 10^{-6} \left(1 - \mathrm{e}^{-5.88 \times 10^{-3}(T-124)}\right) + 5.548 \times 10^{-10} T\,, \tag{19.30}$$

where T is in Kelvin. At 20 °C, $\alpha_L = 2.51 \times 10^{-6}\,\mathrm{K}^{-1}$, whereas from 20 °C to 320 °C, $\bar{\alpha}_L = 3.35 \times 10^{-6}\,\mathrm{K}^{-1}$. Most $\alpha_L(T)$ expressions have exponential-type expressions as in (19.30) but they can always be expanded in terms of T or $(T - T_0)$ up to the third term, beyond which errors become insignificant.

19.4 Enthalpic Thermal Properties

19.4.1 Enthalpy, Heat Capacity and Phase Transformations

Consider the behavior of the enthalpy $H(T)$ versus temperature T of a typical material as it is cooled from a liquid state starting at O as shown in Fig. 19.14 (the sketch is idealized to highlight the fundamentals). For crystalline solids, at the melting or fusion temperature T_m, the liquid L solidifies to form a crystalline solid C. Some degree of undercooling is necessary to nucleate the crystals but we have ignored this fact in Fig. 19.14 so that T_m is the thermodynamic melting temperature where the Gibbs free energies of the liquid (L) and the crystal (C) are the same. The heat capacity $C_{liquid} = \mathrm{d}H_{liquid}/\mathrm{d}T$ for the liquid is greater than that for the crystal, $C_{crystal} = \mathrm{d}H_{crystal}/\mathrm{d}T$, as apparent from the H–T slopes in Fig. 19.14. The fusion enthalpy ΔH_f corresponds to the enthalpy change between the liquid and crystalline phases, L and C, in Fig. 19.14.

When a glass-forming liquid is cooled, it does not crystallize at T_m; its crystallization kinetics are too slow or negligible. Its enthalpy H continues to decrease along the liquid H versus T line with a slope C_{liquid} and it becomes a supercooled liquid. However, as the cooling proceeds, eventually, over a certain small temperature range around a temperature marked as T_g in Fig. 19.14, the atomic motions become too sluggish, and the viscosity becomes too high to follow the liquid-like behavior. The material cools along a solid-like H–T curve that is almost parallel to the crystalline or solid H–T line. This region where the slope or the heat capacity is similar to that of the solid or the crystal ($C_{crystal}$) is called the glass region, and the material is said to be in the *glassy state*. The intersection of the liquid-like (supercooled) and solid-like (glass) H–T lines defines

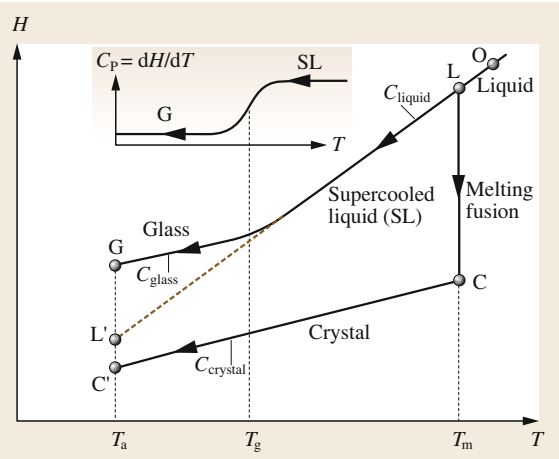


Fig. 19.14 Enthalpy H versus temperature T

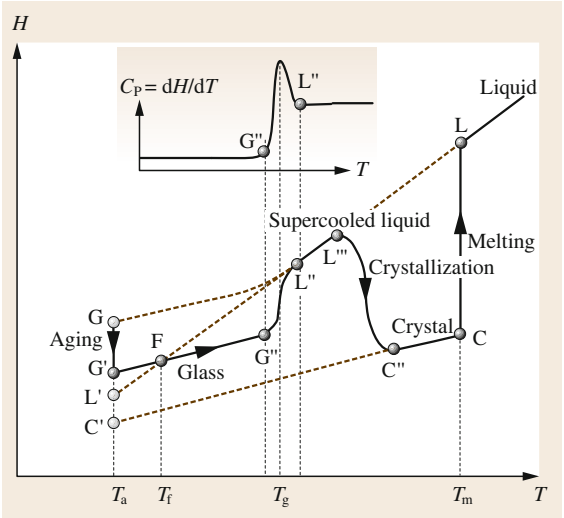


Fig. 19.15 Isothermal aging at T_a changes the glass enthalpy. When the glass is heated at a constant rate r, it undergoes glass transition at around T_g, and then ideally crystallization and then melting

a *glass-transition temperature* T_g as observed under cooling at a particular cooling rate q. T_g depends on the cooling rate (the slower the cooling, the lower the glass transition temperature), and obviously on the structure. C_P changes in a step-like fashion over a small interval around T_g during cooling as shown in Fig. 19.14. The shift in T_g with q can provide valuable information on the kinetics of the glass-transformation process as discussed below. The glass enthalpy is greater than the crystalline enthalpy due to the disorder in the glass structure.

It is apparent from Fig. 19.14 that, at least theoretically, if extremely slow cooling was applied, the enthalpy of the supercooled liquid could decrease at a pace suggested by the LL′ line in Fig. 19.14 without undergoing glass transition and eventually decrease below that of the crystal. This would, in consequence, mean that such massively undercooled liquid would have a negative entropy at 0 K, which is impossible. This scenario is called the *Kauzmann paradox* [19.37] and was, obviously, never truly observed in practice. There are two main theories of how to address the Kauzmann paradox – the first one states that the glass-transition kinetics would always intervene before enthalpy of the supercooled liquid would reach that of the crystal (i. e., no conditions exist under which supercooled liquid would be maintained to such low temperature). The second theory proposes the existence of a thermodynamic glass transition, which is overlapped by its kinetic manifestation. The question whether or not the glass transition is a true thermodynamic transition is still discussed in the literature [19.38]. For practical purposes it is, however, important to note that, regardless of the underlying thermodynamics, the experimentally observed glass transition is a kinetic phenomenon.

Suppose we stop the cooling at an ambient temperature T_a, corresponding to the glass state G. The glass enthalpy $H_{glass}(T_a)$ will relax via structural relaxations toward the enthalpy of the metastable liquid-like equilibrium state $H_{liquid}(T_a)$ at L′. This process is called aging or annealing. When we start heating a glass, its initial enthalpy can be anywhere between G and L′, depending on the thermal history (including aging) and the cooling rate.

When the sample at $H(T_a)$ at G′ is heated at a constant rate r, as shown in Fig. 19.15, the enthalpy of the system, as a result of long structural relaxation times at these low temperatures, follows the glass H–T curve until the atomic and molecular motions due to the increased temperature, are sufficiently rapid to overcome the kinetic barriers and allow the system to recover toward equilibrium $H_{liquid}(T)$; the transition is from G″ to L″. The sigmoidal-like change in $H(T)$ leads to the glass-transformation endotherm and the so-called overshoot in the C_P–T behavior observed for many glasses when they are heated through T_g. This endothermic heat, called the enthalpy of relaxation, depends on the duration of aging, and the structural relaxation processes, which in turn depend on the structure [19.39]. The observed glass-transition temperature T_g during heating depends not only on the heating rate r but also on the initial state $H(T_a)$ or G′. On the other hand, during cooling from the melt down to low temperatures, one observes only a change in the heat capacity. As we further heat the sample, the supercooled liquid state can crystallize, from L‴ to C″ and form a true solid, usu-

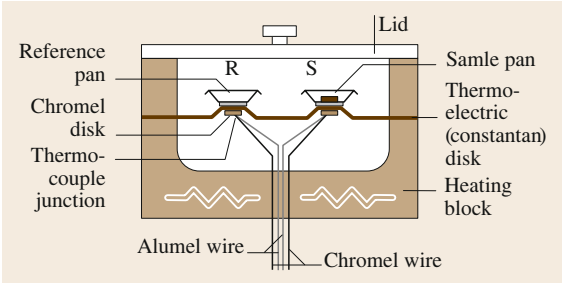


Fig. 19.16 Schematic illustration of heat-flux DSC cell. R and S refer to reference and sample pans

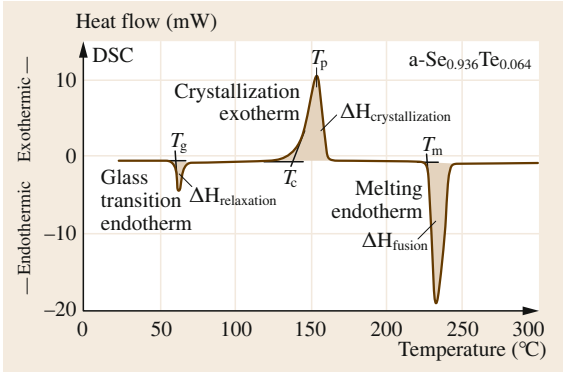


Fig. 19.17 Typical DSC signal for a chalcogenide $Se_{0.936}Te_{0.064}$ glass sample (well relaxed or aged). From left to right: the glass transition, crystallization and the melting temperatures, and the corresponding enthalpies of transitions for this material. $\Delta H_c = \Delta H_{crystallization}$; $\Delta H_f = \Delta H_{fusion}$

ally polycrystalline. This transformation from L‴ to C″ releases heat; it is exothermic. At T_m, the solid melts to form the liquid, C changes to L, which means that heat is absorbed.

It is apparent from the simple H–T behavior in Fig. 19.15 that there are at least several key parameters of interest in characterizing the thermal properties of materials: the heat capacity of the liquid, crystal, and glass phases, C_{liquid}, $C_{crystal}$, C_{glass}, their temperature dependences, melting or fusion temperature T_m, fusion enthalpy ΔH_f, crystallization onset temperature (L‴), exothermic enthalpy of crystallization (L‴ − C‴), T_g (which depends on various factors), and the endothermic glass-transformation enthalpy (L″ − G″). These properties are most conveniently examined using a *differential scanning calorimeter* (DSC).

19.4.2 Conventional Differential Scanning Calorimetry (DSC)

Principles

Differential scanning calorimetry (DSC) is a method that is extensively used to measure heat and temperatures of various transitions and has been recognized as a very useful tool for the interpretation of thermal events as discussed and reviewed by a number of authors (see, e.g., [19.40–46]). In essence, the DSC measures the net rate of heat flow $Q' = dQ/dT$, into a sample with respect to heat flow into a reference inert sample, as the sample temperature is ramped. Figure 19.16 shows a simplified schematic diagram of the heat-flux DSC cell. The constantan disk allows heat transfer to and from the reference and sample pans by thermal conduction. Due to symmetry, the thermal resistance of the heat paths from the heater to the reference and sample pans is approximately the same. The constantan disk has two slightly raised platforms on which the reference and sample pans are placed. Chromel disks fixed under these platforms form area thermocouple junctions with the constantan platforms. The constantan-chromel thermocouples are in series and read the temperature difference ΔT between the reference and sample pans. There are also alumel–chromel thermocouples attached to the chromel disks to measure the individual reference and sample temperatures.

The measured temperature difference ΔT corresponds to a differential heat flow $Q' = dQ/dt$ between the sample and the reference pans by virtue of Fourier's law of heat conduction

$$Q' = \Delta T/R\,, \tag{19.31}$$

where R is the effective thermal resistance of the heat flow path to the sample (or reference) through the constantan disk. In conventional thermal analysis, the sample temperature is either ramped linearly at a constant heating or cooling rate or kept constant (as in isothermal experiments).

A typical example of a heating scan performed for a glassy sample by means of a conventional DSC is shown in Fig. 19.17 where the thermogram evinces three distinct regions that correspond to the glass transition (T_g), crystallization and melting. T_g in DSC analysis is normally defined in terms of an onset temperature as in Fig. 19.17. The crystallization exotherm is normally used to obtain a crystallization onset temperature T_c and a peak crystallization rate temperature T_p that corresponds to the location of the peak crystallization. ΔH_c and fusion ΔH_f enthalpies are obtained by integrating the corresponding peaks as shown in Fig. 19.17 with an appropriate baseline. The base line represents the overall heat capacity of the sample, which changes during the transformation since the sample is gradually transformed into the new phase. Usually a simple sigmoidal, or even a straight line between the onset and the

end points, is sufficient for an accurate determination of the enthalpy.

Conventional DSC has been used to study and characterize a wide range of physical and chemical transformations. When the DSC is used in a heating or cooling scan, the temperature of the sample is ramped along $T = T_0 + rt$, where T_0 is the initial temperature. Consequently, the transformations in DSC are carried out under nonisothermal conditions, and well-known isothermal rate equations cannot be directly applied without some modification.

Glass Transition

Glass transition is one of the most intriguing phenomena in solid-state science. Despite extensive discussions in the literature (e.g., [19.47–52]), its nature is still not fully understood. Contrary to the simple thermally activated processes (crystallization, decomposition, etc.), the glass-transition kinetics exhibits a number of very complex features: hysteresis, nonlinearity, nonexponentiality, and thermorheological complexity [19.53]. Accordingly, special models had to be developed to properly account for these features – the *Tool–Narayanaswamy–Moynihan* (TNM) [19.54–56] and *Adam–Gibbs* (AG) [19.57] phenomenological models belong to the most successful and widespread. Both these models are based on the *Kohlrausch–Williams–Watts* (KWW) [19.58] stretch exponential relaxation function behavior combined with the concept of the so-called *reduced time integral*, which is explained below. For simplicity, we will consider the relaxation of the excess enthalpy $\Delta H(t)$ at any instant toward its equilibrium value, as shown in Fig 19.15. Within the relaxation description of glass transition, it is assumed that $\Delta H(t)$ at any time can be found from the relaxation function $F(t)$ given by the reduced time integral

$$F(t) = \exp\left[-\left(\int_0^t \frac{\mathrm{d}t}{\tau\,(T, T_\mathrm{f})}\right)^{\beta}\right], \qquad (19.32)$$

where the exponent β is a *nonexponentiality parameter*, which indicates the distribution of relaxation times associated with the particular structural cluster; ($\beta = 1$ corresponds to the fully exponential behavior, where all the clusters relax in a similar way and the overall relaxation response can be described by a single exponential function). Both TNM and AG models further utilize a quantity called the *fictive temperature* quantity T_f, which can be defined as the temperature of supercooled liquid with structure similar to that of the relaxing glass. Visualization of this definition can be made with the help of Fig. 19.15, where T_f can be determined as the intersection of the glass line passing through the starting enthalpic state G′ and the extended liquid H–T lines. It depends on the starting enthalpy G′ so that T_f is used as a convenient temperature parameter to identify the initial state of the amorphous structure at G′ (or at any other state reached, for example, by aging of or by application of a more complicated thermal history). The difference between the two models lies in the expression for the relaxation time τ, that is

$$\tau_\mathrm{TNM} = A \exp\left[x\frac{\Delta h^*}{RT} + (1-x)\frac{\Delta h^*}{RT_\mathrm{f}}\right], \qquad (19.33)$$

$$\tau_\mathrm{AG} = A \exp\left[\frac{B}{T\,(1 - T_2/T_\mathrm{f})}\right]. \qquad (19.34)$$

The following relaxation parameters can be identified for the TNM model: A is the pre-exponential constant, x is the nonlinearity parameter (which determines the relative contributions of temperature and structure to the relaxation process), and Δh^* is the apparent activation energy. B is a structure and material-dependent parameter discussed below. Notice that both expressions have T_f, which means that the relaxation times in (19.33) and (19.34) also depend on the initial state at G′. Although all these parameters are phenomenological and have no exact physical meaning, recently it was shown [19.59] that their qualitative interpretation can be used to explain the relaxation phenomena in terms of actual structural changes and movements at a molecular level. On the other hand, although the AG model is also denoted as phenomenological, its relaxation parameters have much clearer physical meaning: T_2 is the temperature at which the macroscopic configurational entropy extrapolates to zero (a quantity similar to the *Kauzmann temperature* introduced in Sect. 19.4.2) and B can be defined as

$$B = \frac{N_\mathrm{A}\Delta\mu S_\mathrm{c}^*}{k_\mathrm{B}\Delta C_\mathrm{P}\,(T_2)}, \qquad (19.35)$$

where k_B is the Boltzmann constant, $\Delta\mu$ is the potential barrier hindering structural rearrangement, S_c^* is the macroscopic configurational entropy, N_A is Avogadro's number, and ΔC_P is the T_g heat capacity change extrapolated to a given temperature (T_2).

In principle, both models can be accommodated within a curve-fitting procedure to the DSC data during which all the model parameters are determined; one of the most convenient algorithms can be found in [19.67]. In practice, in many engineering applications, it is usually sufficient to estimate the position of T_g for the given experimental conditions and/or material. Table 19.4 lists some of the most important empirical rules that can be used to estimate the position of T_g based on some

Table 19.4 Some selected examples of T_g dependences on various factors

Rule	Notation	Comment
$T_g \approx (2/3)T_m$	T_g and T_m in K. T_m = melting temperature of corresponding crystalline phase.	Kauzmman's empirical rule [19.37]. Most glass structures including many amorphous polymers [19.60]. Some highly symmetrical polymers with short repeat units follow $T_g \approx (1/2)T_m$ [19.10]
$\ln(q) \approx -\Delta h^*/RT_g + C$	q = cooling rate; Δh^* = activation energy (during colling); C = constant.	Dependence of T_g on the cooling rate. Δh^* may depend on the range of temperature accessed. Bartenev–Lukianov equation [19.61, 62]
$T_g \approx T_g(\infty) + C/M_n$	M_n = average molecular weight of polymer; C = constant; $T_g(\infty)$ is T_g for very large M_n	Dependence of T_g on the average molecular weight of a polymer [19.63, 64]. Tanaka's rule
$\ln(T_g) \approx 1.6Z + C$	Z = mean coordination number, C = constant (≈ 2.3)	Network glasses. Dependence of T_g on the mean coordination number. Neglects the heating rate dependence. [19.65]
$T_g(x) = T_g(0) - 626x$	T_g in K; x is atomic fraction in a : $(Na_2O + MgO)_x(Al_2O_3 + SiO_2)_{1-x}$ b : $(PbO)_x(SiO_2)_{1-x}$ c : $(Na_2O)_x(SiO_2)_{1-x}$ $T_g(0) = 1080$ K for a; 967 K for b; 895 K for c.	±5%. Silicate glasses [19.66]; x is network modifier. $0.01 < x < 0.6$

fundamental knowledge of the given material (composition, temperature of melting). Since T_g also depends on the heating or cooling rate, such rules should be used as an approximation; nonetheless, they are extremely useful in engineering as a guide to the selection and use of materials.

None of the two above-described approaches is, however, ideal in the sense that tedious time-consuming nonlinear optimizations of a large number of DSC measurements need to be performed to obtain a full description of the relaxation behavior by the two models (which may not be necessary). In certain cases, only a rough estimate of T_g is obtained, based on uncertain assumptions. Recently, a new approach has been introduced [19.68], extending the earlier works of *Kovacs* et al. [19.69] and *Hutchinson* et al. [19.70] in which a so-called constant ratio (CR) cyclic temperature history is introduced in the measurements in such a way that the ratio of the heating and cooling rates (r^+/r^-) is maintained constant. The sample is alternately cooled and heated through T_g at various rates while maintaining $(r^+/r^-) =$ constant. The shift in the peak temperature T_p, corresponding to the maximum of the relaxation overshoot in the heating DSC, is examined as a function of the heating rate r^+. The apparent activation energy Δh^* is then determined from these DSC scans through

$$\frac{\Delta h^*}{R} = -\left[\frac{\mathrm{d}\ln|r^+|}{\mathrm{d}\left(1/T_p\right)}\right]_{(r^-/r^+)=\text{const}}, \qquad (19.36)$$

where r^+ and r^- are the heating and cooling rates, respectively, as defined above. The peak temperature T_p can be replaced by the T_g midpoint introducing errors. Strength of this approach lies in the special design of the DSC experiment, which is performed as a series of alternating cooling and heating scans through T_g, where the cooling rates vary and heating rates are performed at the same rate as was that of the preceding cooling [19.68, 71]. The Δh^* value obtained in this way can be consequently used to calculate the position of T_g for a broad range of experimental conditions. Furthermore, this technique has been extensively tested for all types of structural relaxation behavior [19.71] and has been shown to be highly accurate and relatively insensitive to most data-distortive effects encountered in DSC experiments, including thermal lag, imprecise heating/cooling rates, or incomplete achievement of a true glassy state during cooling.

Very good example of how the results of conventional DSC enthalpy relaxation measurements correlate with the relaxation parameters obtained by means of other commonly used experimental techniques can be shown for the case of amorphous As_2Se_3, which is one of the best glass-formers from the family of chalcogenide glasses and was one of the first extensively studied inorganic glasses with respect to the glass transition kinetics. In [19.72], the enthalpy relaxation of As_2Se_3 glass was studied by means of conventional and temperature-modulated DSC – the TMDSC technique will be described in detail later in Sect. 19.5. For the conventional DSC, the apparent activation en-

Table 19.5 Activation energy (Δh^*) from DSC cooling rate (q) scans (inflection and onset T_g data), from TMDSC modulation frequency (ω) scans (inflection and onset T_g data) [19.72], and Δh_η calculated from *Henderson* and *Ast* viscosity data [19.73] using an average for T_g value that represents the T_g range of interest

	$\ln(q)$ versus $1/T_g$ (inflection T_g)	$\ln(q)$ versus $1/T_g$ (onset T_g)	$\ln(\omega)$ versus $1/T_g$ (inflection T_g)	$\ln(\omega)$ versus $1/T_g$ (onset T_g)
Activation energy (Δh^*)	407.4 kJ mol^{-1}	322.9 kJ mol^{-1}	243.5 kJ mol^{-1}	231.1 kJ mol^{-1}
Activation energy (Δh_η) of viscosity	385.2 kJ mol^{-1}	317.7 kJ mol^{-1}	287.5 kJ mol^{-1}	265.3 kJ mol^{-1}
T_g average	447.8 K	463.6 K	472.9 K	481 K
Difference between Δh^* and Δh_η	5.5%	1.6%	18%	15%

ergies were evaluated from the dependence of T_g on the cooling rate (the methodology was originally developed by *Moynihan* et al. [19.74]); in the case of the TMDSC technique, the activation energies were evaluated from the dependence of T_g on angular frequency of temperature modulation. The resulting Δh^* values were then compared with the activation energies of viscous flow Δh_η determined in the same temperature ranges (viscosity data by *Henderson* and *Ast* [19.73] were used). All determined Δh^* and Δh_η values are listed in Table 19.5 together with the average T_g values for the given combination of used experimental technique (DSC/TMDSC) and T_g evaluation method (onset/inflection). It is apparent that the activation energy decreases as the glass transition (indicated by the mean T_g value) increases, due to different measurement conditions. Nonetheless, good agreement was found between the Δh^* and Δh_η values within the respective temperature ranges, confirming that both the enthalpy relaxation and viscosity processes are closely interconnected and exhibit similar apparent activation energies. In addition, it was also found [19.75] that the volume relaxation measured by means of thermomechanical analysis provides comparable values of apparent activation energy ($\Delta h^*_{TMA} = 319\,\mathrm{kJ\,mol^{-1}}$), which are in excellent agreement with the viscosity data obtained in the given temperature range ($\Delta h_\eta = 316\,\mathrm{kJ\,mol^{-1}}$). The above example clearly demonstrates that the glass transition behavior is universal and manifests similarly across a large variety of physical properties.

Crystallization

The crystallization-phase transformation takes place in many amorphous and glassy materials. This process can be followed directly by microscopy or indirectly by macroscopic methods, such as DSC, x-ray diffraction, electrical conductivity measurements, etc. Microscopic techniques are time consuming and often also experimentally difficult. Therefore, it is not surprising that they are frequently replaced by macroscopic methods and especially by DSC. Despite the high universality of DSC, there are also some shortcomings. Two most important ones are related to the detection limit and instrumental time constant. The sensitivity of modern DSC instruments is very high. However, for a slow process associated with rather small heat evolution, the signal-to-noise ratio might be well below the detection limit and the uncertainty of baseline subtraction may cause substantial errors. On the other hand, for very fast crystallization where the rate constant becomes comparable to the instrumental time constant, a significant part of the signal might be distorted due to transient effects.

For crystallization in the glassy state or in glass-forming supercooled liquids, the overall rate of transformation usually involves the nucleation of the new phase followed by its growth. Theoretical description of nucleation-growth processes is based on the classical nucleation theory that has been developed some 80 years ago. The growth proceeds until the whole or substantial part of the sample has been transformed. Consequently, the crystallization process is a typical sigmoidal time-transformation curve. The shape of this typical curve has been described in 1939 by *Johnson* and *Mehl* [19.76] and *Avrami* [19.77–79]. In general, the equation for volume fraction α transformed as a function of time under isothermal conditions can be written in the following form

$$\alpha = 1 - \exp\left[-(Kt)^m\right] , \qquad (19.37)$$

where the parameter m is known as the Avrami exponent and K is a constant. For some simple cases of crystallization, it is possible to find a characteristic value of the kinetic exponent m whose value depends on nucleation and crystal growth mechanism. A detailed summary of possible m values has been given by *Christian* [19.80]. Equation (19.37) and its differential form expressed by (19.39) and (19.40) are usually referred to as the *Johnson–Mehl–Avrami (JMA) equation*. The parameter K in (19.37) is defined as an effective overall crystallization rate that depends on both the nucleation and growth rates. It should be emphasized that the va-

lidity of the JMA equation is based on the following assumptions [19.81, 82]:

1. Isothermal crystallization conditions
2. Homogeneous nucleation or heterogeneous nucleation at randomly dispersed second-phase particles
3. Growth rate of new crystalline phase is controlled by temperature and it is independent of time
4. Low anisotropy of growing crystals.

Henderson [19.81, 82] has shown that the validity of the JMA equation expressed by (19.37), (19.39), and (19.40), can be extended to nonisothermal conditions if the entire nucleation process takes place during the early stages of the transformation, and it becomes negligible afterward. The crystallization rate is then defined only by temperature and does not depend on the previous thermal history. Similar kinetic equations for nonisothermal crystal growth from preexisting nuclei have also been developed by *Ozawa* [19.83].

It is known that JMA equation must be modified for certain types of isothermal crystallization processes. Examples of such situations are given by systems where finite size effects are of importance, where multiple crystallization mechanisms occur, where inert inclusions are present in the matrix, and where the second-phase particles are anisotropic. *Shepilov* and *Baik* [19.84] have examined the behavior of the Avrami exponent in three dimensions, using computer simulations of the crystallization of randomly oriented ellipsoidal crystals. They found that shielding effects (such as blocking due to anisotropic growth) tend to reduce the value of the Avrami exponent as the crystallization process progresses. However, their analyses were limited to the case of site saturation (prenucleation). *Weinberg* [19.85] examined the behavior of Avrami exponent for anisotropic particle formation in systems exhibiting continuous nucleation and site saturation concluding that for the case studied, blocking effect (stronger impingement) causes a reduction in the Avrami exponent, but the unblocked values are reestablished at infinite times. These features should be taken into account when attempting to determine reaction mechanisms from crystallization data from which anisotropic particles form. The relative reduction in the Avrami exponent as a function of $\alpha(t)$ is larger for the prenucleated samples due to more efficient blocking which occurs in this case. More mechanisms that cause deviations from (19.37) were reported by *Starink* [19.86].

The conditions of validity of the JMA equation concerning the nucleation process and crystal growth morphology are essential for practical application of this equation in the analysis of DSC data. Nevertheless, there are some additional assumptions inherent to any calorimetric measurement that should also be fulfilled. These assumptions can be formulated as follows [19.87]:

1. It is assumed that the rate of the crystallization process $\mathrm{d}\alpha/\mathrm{d}t$ is proportional to the net rate of heat flow, Q', normalized per sample mass

$$\left(\frac{\mathrm{d}\alpha}{\mathrm{d}t}\right) = \frac{1}{\Delta H_\mathrm{c}} Q' \,, \tag{19.38}$$

   where ΔH_c is the total enthalpy change associated with the crystallization process. The fraction transformed, α, can easily be obtained by partial integration of isothermal or nonisothermal crystallization curve. The total peak area then corresponds to ΔH_c.
2. The rate of the crystallization process can be expressed as a product of temperature-dependent crystallization rate constant $K(T)$ and a function $f(\alpha)$ that depends on the fraction transformed,

$$\left(\frac{\mathrm{d}\alpha}{\mathrm{d}t}\right) = K(T) f(\alpha) \,. \tag{19.39}$$

By differentiation of (19.37) with respect to time, we obtain the $f(\alpha)$ function that corresponds to the JMA nucleation-growth equation,

$$f(\alpha) = m\,(1-\alpha)\,[-\ln\,(1-\alpha)]^{1-1/m} \,. \tag{19.40a}$$

The crystallization rate constant in (19.37) is usually assumed to follow an Arrhenius-type temperature dependence, that is, $K(T) = A\exp(-E_\mathrm{a}/RT)$, where A is pre-exponential factor and E_a is an apparent activation energy describing the overall crystallization process. This assumption is appropriate when the nucleation and growth processes vary in an Arrhenian way with temperature, which is fulfilled over a sufficiently narrow range of temperature (such as in the range of DSC crystallization peaks). If the nucleation frequency is negligible over the limited temperature range then the apparent activation energy is approximately equal to activation energy of crystal growth [19.88]. The value of E_a can be determined from the shift of the crystallization peak temperature T_p with the heating rate r. The plot of $\ln(r/T_\mathrm{p}^2)$ versus $1/T_\mathrm{p}$ is called a *Kissinger plot* [19.89, 90], and gives the activation energy E_a. Thus,

$$\ln\left(\frac{r}{T_\mathrm{p}^2}\right) = -\frac{E_\mathrm{a}}{RT_\mathrm{p}} + \text{constant} \,. \tag{19.40b}$$

Combining (19.38) and (19.39), we can write for the heat flow rate corresponding to the crystallization

process

$$Q' = \Delta H_c A \exp\left(-E_a/RT\right) f(\alpha)\,, \tag{19.41}$$

where the $f(\alpha)$ function following the JMA model is expressed by (19.40a). This function should be invariable with respect to the experimental conditions such as sample mass and heating rate (nonisothermal conditions) or temperature (isothermal conditions).

Any direct comparison of the isothermal $Q'(t)$ and nonisothermal $Q'(T)$ DSC crystallization curves is complicated by the fact that they strongly depend on the temperature and heating rate, respectively. Moreover, the shape and symmetry of these curves that correspond to the same kinetic model may be quite different in isothermal and nonisothermal conditions. However, there are two functions $y(\alpha)$ and $z(\alpha)$ that can easily be obtained by a simple transformation of DSC data [19.87]. In isothermal conditions, these functions are defined as

$$y\left(\alpha\right) = Q' \tag{19.42a}$$

$$z\left(\alpha\right) = Q' t\,. \tag{19.42b}$$

In nonisothermal conditions these functions are defined as follows

$$y\left(\alpha\right) = Q' \exp\left(E_a/RT\right)\,, \tag{19.43a}$$

$$z\left(\alpha\right) = Q' T^2\,. \tag{19.43b}$$

For practical reasons, the $y(\alpha)$ and $z(\alpha)$ functions are normalized within the (0, 1) range. From (19.42a) and (19.42b), it is evident that both the $y(\alpha)$ and $z(\alpha)$ functions in isothermal conditions directly follow from the experimental data. Similarly, the $z(\alpha)$ function in nonisothermal conditions is obtained by a simple transformation of experimental data. However, the value of apparent activation energy E_a is needed to calculate the $y(\alpha)$ function in nonisothermal conditions.

The validity of the assumptions (1) and (2) can easily be verified by checking the invariance of the $y(\alpha)$ and $z(\alpha)$ plots with respect to procedural variables such as heating rate (nonisothermal conditions) and temperature (isothermal conditions). These functions should be identical for both isothermal and nonisothermal DSC data provided that the kinetic model $f(\alpha)$ does not change. In case that there are considerable differences or pronounced influence of the procedural parameters affecting the shape of the $y(\alpha)$ and $z(\alpha)$ functions then, probably, these assumptions are not fulfilled. Such behavior can be a consequence of many different factors. One possibility is that the baseline has not been drawn correctly due to substantial change in the sample heat capacity during the measurement. Another possible explanation is that the measured data corresponds to a complex process (parallel or consecutive processes, branching, etc.) featuring different reaction scheme than expressed by (19.42). Thermal inertia effects caused by lower thermal contact between the sample and temperature sensor or low thermal conductivity of amorphous materials can also play an important role [19.91].

The $y(\alpha)$ and $z(\alpha)$ plots exhibit maxima at α_y^* and α_z^*, respectively. The maximum of the $y(\alpha)$ function for the JMA model depends on the value of the kinetic exponent. For $m \leq 1$, it is $\alpha_y^* = 0$ and for $m > 1$ it can be expressed as

$$\alpha_y^* = 1 - \exp\left(m^{-1} - 1\right)\,. \tag{19.44}$$

This equation can be used to estimate the value of the kinetic exponent. The maximum of the $z(\alpha)$ function α_z^* that is a characteristic constant for the JMA model: $\alpha_z^* = 0.632$. Therefore, it can be used as a simple test of the applicability of this model [19.87]. Taking into account a typical error limit of α_z^*, determined from calorimetric data (± 0.02) then the maximum of the $z(\alpha)$ function can be expected within the interval $0.61 \leq \alpha_z^* \leq 0.65$ for the JMA model. Higher or lower values indicate increasing complexity and the JMA model can hardly be applied. This can be caused, for example, by the non-negligible influence of heterogeneous nucleation. However, a complex behavior can also be observed when the temperature distribution within the sample is nonuniform due to the heat liberated by crystallization process at the crystal–liquid interface [19.82]. Such local overheating can lead to an autocatalytic behavior of the crystallization kinetics that has been observed in several noncrystalline systems. In this case, the crystallization can be described by means of an empirical two-parameter autocatalytic (AC) model [19.91]

$$f(\alpha) = \alpha^M \left(1-\alpha\right)^N\,, \tag{19.45}$$

where the parameters M and N define relative contributions of acceleratory and decelerating parts of the crystallization process, respectively. It has been shown [19.92] that (19.45) is meaningful only for $M < 1$. The maxima of the $y(\alpha)$ and $z(\alpha)$ functions depend on the value of the kinetic exponents M and N. The maximum α_y^* of the $y(\alpha)$ function can be expressed as [19.87]

$$\alpha_y^* = \frac{M}{M+N}\,. \tag{19.46}$$

However, the maximum α_z^* of the $z(\alpha)$ function cannot be expressed in an analytical form and has to be obtained by a numerical solution [19.87].

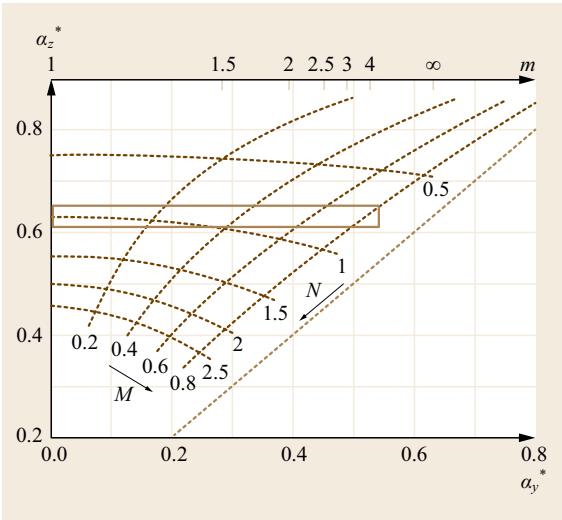


Fig. 19.18 The α_y^* versus α_z^* diagram for the JMA and AC model [19.91]. Numbers mark the values of the kinetic exponents M and N for the AC model and the Avrami exponent m. The area marked by full lines corresponds to the applicability range of the JMA model. The broken line shows the theoretical limit for $\alpha_y^* = \alpha_z^*$

Figure 19.18 shows the α_y^* versus α_z^* diagram for the JMA and AC models. The dotted lines were calculated for a constant values of the parameters M and N. The whole area of this diagram, restricted by a broken line defined as $\alpha_y^* = \alpha_z^*$ (i. e., theoretical limit [19.24, 87] for any $f(\alpha)$ function defined by (19.40) corresponds to the AC model. The JMA model is restricted to the small area where the condition of $0.61 \leq \alpha_z^* \leq 0.65$ is fulfilled. Therefore, the JMA model is a special case of the two-parameter AC model; it is a plausible description for the nucleation-growth processes in non-crystalline solids. A similar conclusion was also found by the shape analysis of DSC curve for the JMA and AC models [19.93]. The increasing value of the kinetic exponent M indicates a more important role of the crystallized phase on the overall kinetics. It also seems that a higher value of the kinetic exponent $N > 1$ means increasing complexity. However, the temptation to associate parameters M and N with a particular crystallization mechanism should be avoided, and too much physical significance should not be attached to the numerical values of what is essentially a phenomenological convenience. In particular, this advice applies to processes exhibiting a complex behavior. Any meaningful conclusions on the mechanism of the crystallization process should always be based on direct experimental evidence, such as microscopic observation of the crystal growth [19.91].

Probably the most popular method for analysis of isothermal data is an inspection of the linearity of the Avrami plot; that is, the dependence of $\ln[-\ln(1-\alpha)]$ as a function of logarithm of time. The kinetic exponent m is obtained from the slope of this plot [19.87]

$$\frac{\mathrm{d}\ln[-\ln(1-\alpha)]}{\mathrm{d}\ln t} = m\,. \tag{19.47}$$

A similar method known as the *Coats–Redfern–Sestak plot* [19.94, 95] has been developed for nonisothermal DSC data. The slope of $\ln[-\ln(1-x)]$ versus $1/T$ dependence obtained from a single scan DSC experiment then provides the product of m and the activation energy E_a,

$$\frac{\mathrm{d}\ln[-\ln(1-\alpha)]}{\mathrm{d}\,(1/T)} \cong -\frac{mE_\mathrm{a}}{R}\,. \tag{19.48}$$

The value of the kinetic exponent m can then be calculated if E_a is known. The latter can be extracted from multiscan DSC measurements by using (19.40b), so in principle, one can determine both m and E_a. However, it is well known that a double logarithmic function, in general, is not very sensitive to subtle changes to its argument. One can expect that the plots of $\ln[-\ln(1-\alpha)]$ versus $\ln t$ or $1/T$ may appear linear even in the case that the JMA model is not fulfilled. Consequently the value of the parameter m estimated in this way may not necessarily correspond to the true m that describes the crystallization kinetics. Therefore, it seems that better way to estimate the Avrami exponent is (19.44) [19.87].

19.5 Temperature-Modulated DSC (TMDSC)

19.5.1 TMDSC Principles

In the early 1990s, a greatly enhanced version of the DSC method called temperature-modulated differential scanning calorimetry (MDSC) was introduced by the efforts of *Reading* and coworkers [19.96–98]. The MDSC method incorporates not only the ability of conventional DSC but it also provides significant and distinct advantages over traditional DSC. The benefits of the MDSC technique have been documented in several recent papers, and include the following: separation of complex transitions, e.g., glass transition, into eas-

ily interpreted components; measurement of heat flow and heat capacity in a single experiment; ability to determine more accurately the initial crystallinity of the studied material; increased sensitivity for the detection of weak transitions; increased resolution without the loss of sensitivity; measurement of thermal conductivity [19.99, 100]. One of the most important benefits is the separation of complex transitions such as the glass transition into more easily interpreted components. Recent applications of MDSC to glasses has shown that it can be very useful for the interpretation of thermal properties, such as the heat capacity, in relation to the structure as, for example, in the case of chalcogenide glasses (e.g., [19.101–103]).

The MDSC that is currently commercialized by TA Instruments uses a conventional heat-flux DSC cell whose heating block temperature is sinusoidally modulated. In MDSC, the sample temperature is modulated sinusoidally about a constant ramp so that the temperature T at time t is,

$$T = T_0 + rt + A\sin(\omega t)\,, \tag{19.49}$$

where T_0 is the initial (or starting) temperature, r is the heating rate (which may also be a cooling ramp q), A is the amplitude of the temperature modulation, $\omega = 2\pi/P$ is the angular frequency of modulation and P is the modulation period. It should be emphasized that (19.49) is a simplified statement of the fact that the cell has reached a steady-state operation and that the initial temperature transients have died out. The resulting instantaneous heating rate, $\mathrm{d}T/\mathrm{d}t$, therefore varies sinusoidally about the average heating rate r, and is given by

$$\frac{\mathrm{d}T}{\mathrm{d}t} = r + A\omega\sin(\omega t)\,. \tag{19.50}$$

At any time, the apparatus measures the sample temperature and the amplitude of the instantaneous heat flow (by measuring ΔT) and then, by carrying out a suitable Fourier deconvolution of the measured quantities, it determines two quantities (which have been termed by TA Instruments):

1. Reversing heat flow (RHF).
2. Nonreversing heat flow (NHF).

Fourier transforms are made on one full cycle of temperature variation, which means that the average quantities refer to moving averages. The average heat flow, which corresponds to the average heating rate (r), is called the total heat flow (HF). Total heat flow is the only quantity that is available and hence it is the only quantity that is always measured in conventional DSC experiments. MDSC determines the heat capacity using the magnitudes of heat flow and heating rate obtained by averaging over one full temperature cycle. If triangular brackets are used for averages over one period P, then $\langle Q'\rangle$ is the average heat flow per temperature cycle. The heat capacity per cycle is then calculated from

$$mC_\mathrm{P} = \frac{\langle\text{Heat flow}\rangle}{\langle\text{Heating rate}\rangle}\,, \tag{19.51}$$

where m is the mass of the sample. This C_P has been called the *reversing heat capacity*, though Schawe has defined as the complex heat capacity [19.104]. The reversing in this context refers to a heat flow that is reversing over the time scale of the modulation period. Furthermore, it is assumed that C_P is constant, that is, it does not change with time or temperature over the modulation period. The reversing heat flow is then obtained by

$$\mathrm{RHF} = C_\mathrm{P}\left\langle\frac{\mathrm{d}T}{\mathrm{d}t}\right\rangle\,. \tag{19.52}$$

The nonreversing heat flow (NHF) is the difference between the total heat flow and the reversing heat flow and represents heat flow due to a kinetically hindered process such as crystallization. There are a few subtle issues in the above condensed qualitative explanation. First is that the method requires several temperature cycles during a phase transition to obtain RHF and NHF components, which sets certain requirements on r, A and ω. The second is that the phase difference between the heat flow and the heating rate oscillations is assumed to be small as it would be the case through a glass-transition region or crystallization; but not through a melting process. There are a number of useful discussions and reviews on the MDSC technique and its applications in the literature [19.105–109].

19.5.2 TMDSC Applications

There has been considerable scientific interest in applying TMDSC measurements to the study of glass-transformation kinetics in glasses and polymers (e.g., [19.101, 102, 110–114]). The interpretation of TMDSC measurements in the glass-transition region has been discussed and reviewed by *Hutchinson* and *Montserrat* [19.115–118]. The reversing heat flow (RHF) through the T_g region exhibits a step-like change and represents the change in the heat capacity. The hysteresis effects associated with thermal history seem to be less important in the RHF but present in the NHF. The measurement of T_g from the RHF in TMDSC

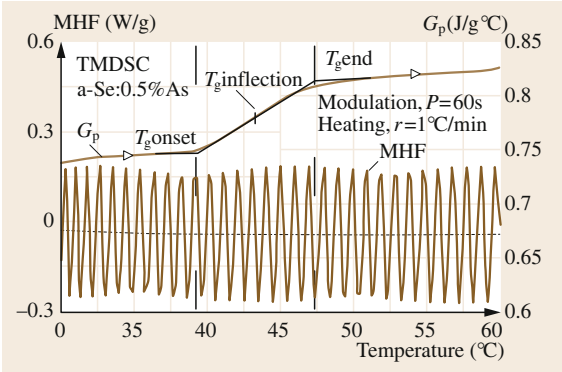


Fig. 19.19 A typical TMDSC scan

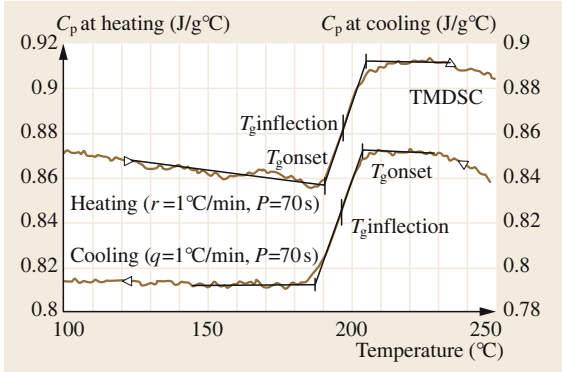


Fig. 19.20 Typical TMDSC heating and cooling scans on As_2Se_3 glass

experiments shows only a weak dependence on glass aging and thermal history [19.119–121], which is a distinct advantage of this technique. The interpretation of the NHF has been more difficult but it is believed that it provides a qualitative indication of the enthalpy loss during the annealing period below T_g [19.105] though more research is needed to clarify its interpretation. Figure 19.19 shows a typical MDSC result through the glass-transition region of $Se_{99.5}As_{0.5}$ glass.

It is important to emphasize that ideally the underlying heating rate in TMDSC experiments should be as small as possible. In this way we can separate the conventional DSC experiment, which also takes place during TMDSC measurements, from the dynamic, frequency-controlled TMDSC experiment. The oscillation amplitude A in the TMDSC must be properly chosen so that the C_P measurements do not depend on A; typically $A = \pm 1.0\,^{\circ}\mathrm{C}$ [19.72]. The oscillation period P should be chosen to ensure that there are at least four full modulations within the half width of the temperature transition, that is, a minimum of eight oscillations over the T_g range, as can be seen in the example in Fig. 19.19 which has a particularly narrow glass-transition range and required a $P = 60\,\mathrm{s}$. The difference between the average specific capacity, as determined from the conventional DSC heat flow, and the complex heat capacity, as determined from TMDSC RHF, in the T_g region is substantial [19.117]. The influence of experimental conditions on TMDSC measurements have been well discussed in the literature [19.104].

The glass transition T_g as observed in DSC scans is well known to depend on thermal history. The classical well-known example is the fact that the T_g value determined by DSC under a heating scan at a given rate r is different from that determined under a cooling scan at a rate $q = r$. Further, T_g determined in a DSC heating scan also depends on aging. The case for As_2Se_3 is well documented [19.122]. Figure 19.20 shows an example in which the T_g values for vitreous As_2Se_3 determined in TMDSC heating and cooling scans are approximately the same for a given modulation period P, within experimental errors. T_g depends only on the modulation frequency $\omega\,(= 2\pi/P)$ provided that the underlying heating or cooling scan is sufficiently slow, usually less than $1\,^{\circ}\mathrm{C/min}$. Other examples are also available in the literature for both glasses and polymers [19.123, 124]. The T_g measurement in the TMDSC experiments represents the relaxation processes associated with the temperature modulation, and not the underlying heating rate as in the conventional DSC. For example, T_g in TMDSC represents a temperature where the relaxation time τ is comparable with the modulation period P. Above T_g, $\tau \ll P$ and below T_g, $\tau \gg P$, and in the T_g region $\tau \approx P$. As P (or ω) is varied, from experiment to experiment, different values of τ satisfy $\tau \approx P$ and hence T_g depends on ω. It is possible to examine the shift in T_g with ω, and then plot $\ln\omega$ versus $1/T_g$ to extract the activation energy of the relaxation time τ, as shown for As_2Se_3 [19.72]. T_g that is observed in conventional DSC experiments is associated with the devitrification (on heating) and vitrification (on cooling) of the structure, whereas T_g in TMDSC experiments represents a dynamic glass-transition temperature; it is important to arrange the experimental conditions so that the two processes are separated in the measurement [19.125].

While a single phenomenological relaxation time provides a convenient description of the overall glass-transformation behavior, a more complete description necessarily involves incorporating a distribution of relaxation times, as discussed in [19.118].

19.5.3 Tzero Technology

It is well known that there is a thermal lag between the sample and temperature-control sensors in the DSC and

most DSCs require calibration at the heating rate to be used. Alternatively, calibration at multiple rates is required to correct for the heating rate dependence of the thermal lag. Without this calibration, there is considerable error in the temperature data in changing from one heating rate to another. This problem is especially acute in cooling experiments due to the thermal lag being in the opposite direction to that of heating, thereby doubling the lag effect. The lag problem has been solved in a recent DSC technique called the *Tzero technology*, as developed by TA Instruments [19.126, 127]. With the Q-series DSC system, an additional thermocouple on the DSC sensor allows a complete calibration of the DSC cell and compensation for all thermal lags, including those caused by the encapsulated sample. This means that, for all practical purposes, the DSC can be calibrated at any reasonable heating rate, and the DSC temperature data will be correct within a few tenths of a Celsius degree for data taken at other heating or cooling rates. Inasmuch as advanced Tzero compensates for the effect of pan thermal mass and coupling, it is possible to calibrate using one pan type and then use another pan type without incurring substantial errors. In summary, since the Tzero technology uses more information in the DSC measurement, it is more accurate under a wider range of conditions compared with the ordinary DSC without calibration under those specific conditions.

The thermal lag error is proportional to heat flow, heating rate, and to the mass of the sample/pan system. Hence, this error becomes greatest with fast scanning rates, large sample masses, massive sample pans, or sample specimens with an especially high heat capacity. The thermal lag error is also proportional to the thermal resistance between the sample and sensors so it is made worse by using pans made of poor thermal conductivity or pans making poor thermal contact. However, even in ordinary polymer samples, using optimally coupled aluminum pans, the error produced could be more than two Celsius degrees because of poor thermal conductivity; for other samples it could be several times larger. In summary, the Tzero technology has enabled better DSC experiments to be carried out.

Acknowledgments. The authors thank Dr. Dancho Tonchev for many helpful discussions on thermal properties, and his invaluable contributions to the first edition of this chapter.

References

19.1 W. Martienssen, H. Warlimont (Eds.): *Springer Handbook of Condensed Matter and Materials Data* (Springer, Berlin, Heidelberg, New York 2005)
19.2 O. Madelung: *Semiconductors: Data Handbook*, 3rd edn. (Springer, Berlin, Heidelberg, New York 2004)
19.3 S. Adachi: *Properties of Group IV, III–V and II–VI Semiconductors* (Wiley, Chichester 2005)
19.4 P. Debye: Ann. Phys. **39**, 789 (1912)
19.5 S.O. Kasap: *Principles of Electronic Materials and Devices*, 4th edn. (McGraw–Hill, Dubuque 2017)
19.6 S. Elliott: *The Physics and Chemistry of Solids* (Wiley, Chichester 1998)
19.7 R.B. Stephens: Phys. Rev. B **8**, 2896 (1973)
19.8 J. De Launay: In: *Solid State Physics*, Vol. 2, ed. by F. Seitz, D. Turnbull (Academic, New York 1956)
19.9 K. Ichikawa: J. Phys. C **18**, 4631 (1985)
19.10 D.W. Van Krevelen, P.J. Hoftyzer: *Properties of Polymer* (Elsevier, Amsterdam 1976)
19.11 M. Pyda, E. Nowak-Pyda, J. Mays, B. Wunderlich: J. Polymer Sci. B **42**, 4401 (2004)
19.12 B. Wunderlich: Thermochim. Acta **300**, 43 (1997)
19.13 D.E. Sharp, L.B. Ginther: J. Am. Ceram. Soc. **34**, 260 (1951)
19.14 S. Inaba, S. Oda: J. Non-Cryst. Solids **325**, 258 (2003)
19.15 S.A. Khalimovskaya-Churkina, A.I. Priven: Glass Phys. Chem. **26**, 531 (2000)
19.16 S. Rudtsch: Thermochim. Acta **382**, 17 (2002)
19.17 M.J. O'Neill: Anal. Chem. **38**, 1331 (1966)
19.18 Y.P. Joshi, G.S. Verma: Phys. Rev. B **1**, 750 (1970)
19.19 C.J. Glassbrenner, G. Slack: Phys. Rev. **134**, A1058 (1964)
19.20 M.G. Holland: *The Proceedings of the 7th Int. Conf. Phys. Semicond., Paris* (Dunond, Paris 1964)
19.21 C.M. Bhandari, C.M. Rowe: *Thermal Conduction in Semiconductors* (Wiley, New Delhi 1988)
19.22 R.C. Zeller, R.O. Pohl: Phys. Rev. B **4**, 2029 (1971)
19.23 C. Kittel: Phys. Rev. **75**, 972 (1949)
19.24 C. Kittel: *Introduction to Solid State Physics*, 8th edn. (Wiley, New York 2005)
19.25 P.B. Allen, J.L. Feldman: Phys. Rev. Lett. **62**, 645 (1989)
19.26 P.B. Allen, J.L. Feldman: Phys. Rev. B **48**, 12581 (1993)
19.27 A. Jagannathan, R. Orbach, O. Entin-Wohlman: Phys. Rev. B **30**, 13465 (1989)
19.28 C. Oligschleger, J.C. Schön: Phys. Rev. B **59**, 4125 (1999)
19.29 M.P. Zaitlin, A.C. Anderson: Phys. Rev. Lett. **33**, 1158 (1974)
19.30 K. Eiermann: Kolloid-Z. Z. Polymere **201**, 3 (1965)
19.31 Y. Agari, A. Ueda, Y. Omura, S. Nagai: Polymer **38**, 801 (1997)
19.32 B. Weidenfeller, M. Höfer, F. Schilling: Composites A **33**, 1041 (2002)
19.33 B. Weidenfeller, M. Höfer, F.R. Schilling: Composites A **35**, 423 (2004)

19.34 R. Bube: *Electronic Properties of Crystalline Solids: An Introduction to Fundamentals* (Academic, New York 1974)
19.35 A. Jezowski, J. Mucha, G. Pompe: J. Phys. D **20**, 1500 (1987)
19.36 Y. Okada, Y. Tokumaru: J. Appl. Phys. **56**, 314 (1984)
19.37 W. Kauzmann: Chem. Rev. **43**, 219 (1948)
19.38 T.M. Nieuwenhuizen: J. Chem. Phys. **115**, 8083 (2001)
19.39 J.M. Hutchinson, P. Kumar: Thermochim. Acta **391**, 197 (2002)
19.40 R.C. Mackenzie: Thermochim. Acta **28**, 1 (1979)
19.41 W.W. Wedlandt: *Thermal Analysis*, 3rd edn. (Wiley, New York 1986) p. 3
19.42 B. Wunderlich: *Thermal Analysis* (Academic, New York 1990)
19.43 E.F. Palermo, J. Chiu: Thermochim. Acta **14**, 1 (1976)
19.44 S. Sarig, J. Fuchs: Thermochim. Acta **148**, 325 (1989)
19.45 W.Y. Lin, K.K. Mishra, E. Mori, K. Rajeshwar: Anal. Chem. **62**, 821 (1990)
19.46 T. Ozawa: Thermochim. Acta **355**, 35 (2000)
19.47 J. Wong, C.A. Angell: *Glass, Structure by Spectroscopy* (Marcel Dekker, New York 1976)
19.48 J. Zaryzycki: *Glasses and the Vitreous State* (Cambridge Univ. Press, Cambridge 1991)
19.49 J. Jäckle: Rep. Prog. Phys. **49**, 171 (1986)
19.50 C.A. Angell: J. Res. Natl. Inst. Stand. Technol. **102**, 171 (1997)
19.51 C.A. Angell, B.E. Richards, V. Velikov: J. Phys. Cond. Matter **11**, A75 (1999)
19.52 I. Gutzow, B. Petroff: J. Non-Cryst. Solids **345**, 528 (2004)
19.53 G.W. Scherer: *Relaxation in Glass and Composites* (Wiley, New York 1986)
19.54 A.Q. Tool: J. Am. Ceram. Soc. **29**, 240 (1946)
19.55 O.S. Naraynaswamy: J. Am. Ceram. Soc. **54**, 491 (1971)
19.56 C.T. Moynihan, A.J. Easteal, M.A. DeBolt, J. Tucker: J. Am. Ceram. Soc. **59**, 12 (1976)
19.57 G. Adam, J.H. Gibbs: J. Chem. Phys. **43**, 139 (1965)
19.58 G. Williams, D.C. Watts: Trans. Faraday Soc. **66**, 80 (1970)
19.59 R. Svoboda: Acta Materialia **61**, 4534 (2013)
19.60 R.F. Boyer: J. Appl. Phys. **25**, 825 (1954)
19.61 H.N. Ritland: J. Am. Ceram. Soc. **37**, 370 (1954)
19.62 G.M. Bartenev, I.A. Lukyanov: Zh. Fiz. Khim. **29**, 1486 (1955)
19.63 T.G. Fox, P.J. Flory: J. Polym. Sci. **14**, 315 (1954)
19.64 T.G. Fox, S. Loshaek: J. Polym. Sci. **15**, 371 (1955)
19.65 K. Tanaka: Solid State Commun. **54**, 867 (1985)
19.66 I. Avramov, T. Vassilev, I. Penkov: J. Non-Cryst. Solids **351**, 472 (2005)
19.67 I.M. Hodge, A.R. Berens: Macromolecules **15**, 762 (1982)
19.68 R. Svoboda: J. Therm. Anal. Calorim. **121**, 895 (2015)
19.69 A.J. Kovacs, J.J. Aklonis, J.M. Hutchinson, A.R. Ramos: J. Polym. Sci. **17**, 1097 (1979)
19.70 J.M. Hutchinson, M. Ruddy, M.R. Wilson: Polymer **29**, 152 (1988)
19.71 R. Svoboda: Eur. Polym. J. **59**, 180 (2014)
19.72 S.O. Kasap, D. Tonchev: J. Mater. Res. **16**, 2399 (2001)
19.73 D.W. Henderson, D.G. Ast: J. Non-Cryst. Solids **64**, 43 (1984)
19.74 C.T. Moynihan, A.J. Easteal, M.A. DeBolt, J. Tucker: J. Am. Ceram. Soc. **59**, 12 (1976)
19.75 J. Málek, J. Shánělová: J. Non-Cryst. Solids **351**, 3458 (2005)
19.76 W.A. Johnson, R.F. Mehl: Trans, Am. Inst. Min. Metall. Eng. **135**, 416 (1939)
19.77 M. Avrami: J. Chem. Phys. **7**, 1103 (1939)
19.78 M. Avrami: J. Chem. Phys. **8**, 211 (1940)
19.79 M. Avrami: J. Chem. Phys. **9**, 177 (1941)
19.80 J.W. Christian: *The Theory of Transformations in Metals and Alloys*, 2nd edn. (Pergamon Press, New York 1975)
19.81 D.W. Henderson: J. Therm. Anal. **15**, 325 (1979)
19.82 D.W. Henderson: J. Non-Cryst. Solids **30**, 301 (1979)
19.83 T. Ozawa: Bull. Chem. Soc. Jpn. **57**, 639 (1984)
19.84 M.P. Shepilov, D.S. Baik: J. Non-Cryst. Solids **171**, 141 (1994)
19.85 M.C. Weinberg, D.P. Birnie: J. Non-Cryst. Solids **202**, 290 (1996)
19.86 M.J. Starink: J. Mater. Sci. **36**, 4433 (2001)
19.87 J. Málek: Thermochim. Acta **355**, 239 (2000)
19.88 H. Yinnon, D.R. Uhlmann: J. Non-Cryst. Solids **54**, 253 (1983)
19.89 H.E. Kissinger: J. Res. Natl. Bur. Stand. **57**, 217 (1956)
19.90 H.E. Kissinger: Anal. Chem. **29**, 1702 (1957)
19.91 J. Málek, T. Mitsuhashi, J.M. Criado: J. Mater. Res. **16**, 1862 (2001)
19.92 J. Málek, J.M. Criado, J. Šesták, J. Militký: Thermochim. Acta **153**, 429 (1989)
19.93 J. Málek: Thermochim. Acta **222**, 105 (1993)
19.94 A.W. Coats, J.P. Redfern: Nature **201**, 68 (1964)
19.95 J. Sestak: Thermochim. Acta **3**, 150 (1971)
19.96 M. Reading, D. Elliott, V.L. Hill: J. Therm. Anal. **40**, 949 (1993)
19.97 M. Reading: Trends Polym. Sci. **1**, 248 (1993)
19.98 M. Reading, A. Luget, R. Wilson: Thermochim. Acta **238**, 295 (1994)
19.99 E. Verdonck, K. Schaap, L.C. Thomas: Int. J. Pharm. **192**, 3 (1999)
19.100 C.M.A. Lopes, M.I. Felisberti: Polym. Test. **23**, 637 (2004)
19.101 T. Wagner, S.O. Kasap: Philos. Mag. **74**, 667 (1996)
19.102 T. Wagner, M. Frumar, S.O. Kasap: J. Non-Cryst. Solids **256**, 160 (1999)
19.103 P. Boolchand, D.G. Georgiev, M. Micoulaut: J. Optoelectron. Adv. Mater. **4**, 823 (2002)
19.104 J.E.K. Schawe: Thermochim. Acta **271**, 127 (1996)
19.105 K.J. Jones, I. Kinshott, M. Reading, A.A. Lacey, C. Nikopoulos, H.M. Pollosk: Thermochim. Acta **305**, 187 (1997)
19.106 Z. Jiang, C.T. Imrie, J.M. Hutchinson: Thermochim. Acta **315**, 1 (1998)
19.107 B. Wunderlich: Thermochim. Acta **355**, 43 (2000)
19.108 H. Huth, M. Beiner, S. Weyer, M. Merzlyakov, C. Schick, E. Donth: Thermochim. Acta **377**, 113 (2001)

19.109 Z. Jiang, C.T. Imrie, J.M. Hutchinson: Thermochim. Acta **387**, 75 (2002)
19.110 T. Wagner, S.O. Kasap, K. Maeda: J. Mater. Res. **12**, 1892 (1997)
19.111 I. Okazaki, B. Wunderlich: J. Polym. Sci. **34**, 2941 (1996)
19.112 L. Thomas, A. Boller, I. Okazaki, B. Wunderlich: Thermochim. Acta **291**, 85 (1997)
19.113 L. Thomas: NATAS Notes (North American Thermal Analysis Society, Sacramento) **26**, 48 (1995)
19.114 B. Hassel: NATAS Notes (North American Thermal Analysis Society, Sacramento) **26**, 54 (1995)
19.115 J.M. Hutchinson, S. Montserrat: J. Therm. Anal. **47**, 103 (1996)
19.116 J.M. Hutchinson, S. Montserrat: Thermochim. Acta **305**, 257 (1997)
19.117 J.M. Hutchinson: Thermochim. Acta **324**, 165 (1998)
19.118 J.M. Hutchinson, S. Montserrat: J. Therm. Anal. **377**, 63 (2001)
19.119 A. Boller, C. Schick, B. Wunderlich: Thermochim. Acta **266**, 97 (1995)
19.120 J.M. Hutchinson, A.B. Tong, Z. Jiang: Thermochim. Acta **335**, 27 (1999)
19.121 D. Tonchev, S.O. Kasap: Mater. Sci. Eng. **A328**, 62 (2002)
19.122 S.O. Kasap, S. Yannacopoulos: Phys. Chem. Glasses **31**, 71 (1990)
19.123 P. Kamasa, M. Pyda, A. Buzin, B. Wunderlich: Thermochim. Acta **396**, 109 (2003)
19.124 D. Tonchev, S.O. Kasap: Thermal characterization of glasses and polymers by temperature modulated differential scanning calorimetry: Glass transition temperature. In: *High Performance Structures and Materials II*, ed. by C.A. Brebbia, W.P. De Wilde (WIT, Southampton 2004) p. 223
19.125 S. Weyer, M. Merzlyakov, C. Schick: Thermochim. Acta **377**, 85 (2001)
19.126 L.E. Waguespack, R.L. Blaine: Design of a new DSC cell with Tzero technology, Proc. 29th North Am. Therm. Anal. Soc., St. Louis, ed. by K.J. Kociba (NATAS, Sacramento 2001) p. 722
19.127 R.L. Danley: Thermochim. Acta **395**, 201 (2003)

20. Electrical Characterization of Semiconductor Materials and Devices

M. Jamal Deen, Fabien Pascal

Semiconductor materials and devices continue to occupy a preeminent technological position due to their importance when building integrated electronic systems used in a wide range of applications from computers, cell-phones, personal digital assistants, digital cameras and electronic entertainment systems, to electronic instrumentation for medical diagnositics and environmental monitoring. Key ingredients of this technological dominance have been the rapid advances made in the quality and processing of materials – semiconductors, conductors and dielectrics – which have given metal oxide semiconductor device technology its important characteristics of negligible standby power dissipation, good input–output isolation, surface potential control and reliable operation. However, when assessing material quality and device reliability, it is important to have fast, nondestructive, accurate and easy-to-use electrical characterization techniques available, so that important parameters such as carrier doping density, type and mobility of carriers, interface quality, oxide trap density, semiconductor bulk defect density, contact and other parasitic resistances and oxide electrical integrity can be determined. This chapter describes some of the more widely employed and popular techniques that are used to determine these important parameters. The techniques presented in this chapter range in both complexity and test structure requirements from simple current–voltage measurements to more sophisticated low-frequency noise, charge pumping and deep-level transient spectroscopy techniques.

20.1 **Resistivity** 454
20.1.1 Bulk Resistivity 454
20.1.2 Contact Resistivity 459
20.2 **Hall Effect** 462
20.2.1 Physical Principles 462
20.2.2 Hall Scattering Factor 463
20.3 **Capacitance–Voltage Measurements** 464
20.3.1 Average Doping Density by Maximum–Minimum High-Frequency Capacitance Method 464
20.3.2 Doping Profile by High-Frequency and High–Low Frequency Capacitance Methods 466
20.3.3 Density of Interface States 467
20.4 **Current–Voltage Measurements** 470
20.4.1 *I–V* Measurements on a Simple Diode 470
20.4.2 *I–V* Measurements on a Simple MOSFET 470
20.4.3 Floating Gate Measurements 471
20.5 **Charge Pumping** 472
20.6 **Low-Frequency Noise** 473
20.6.1 Introduction 473
20.6.2 Noise from the Interfacial Oxide Layer 474
20.6.3 Impedance Considerations During Noise Measurement 475
20.7 **Deep-Level Transient Spectroscopy** 477
References 479

The continued evolution of semiconductor devices to smaller dimensions in order to improve performance – speed, functionality, integration density and reduced cost – requires layers or films of semiconductors, insulators and metals with increasingly high quality that are well-characterized and that can be deposited and patterned to very high precision. However, it is not always the case that improvements in the quality of materials have kept pace with the evolution of integrated circuit down-scaling. An important aspect of assessing the material quality and device reliability is the development and use of fast, nondestructive and accurate electrical characterization techniques to determine important parameters such as carrier doping density, type and mo-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_20

bility of carriers, interface quality, oxide trap density, semiconductor bulk defect density, contact and other parasitic resistances and oxide electrical integrity. This chapter will discuss several techniques that are used to determine these important parameters. However, it is not an extensive compilation of the electrical techniques currently used by the research and development community; rather, it presents a discussion of some of the more widely used and popular ones [20.1–4].

An important aspect of electrical characterization is the availability of appropriate test components [20.1–4]. In this chapter, we concentrate on discussing techniques that use standard test devices and structures. In addition, we will use the metal–oxide–semiconductor field-effect transistor (MOSFET) whenever possible because they are widely available on test chips. This is also motivated by the fact that MOSFETs continue to dominate the semiconductor industry for a wide range of applications from memories and microprocessors to signal and imaging processing systems [20.5]. A key reason for this dominance is the excellent quality of the silicon wafers and the silicon–silicon dioxide interface, both of which play critical roles in the performance and reliability of the device. For example, if the interface has many defects or interface states, or it is rough, then the device's carrier mobility decreases, low-frequency noise increases and its performance and reliability degrades. In particular, it is not only the interface that is important, but also the quality of the oxide; good-quality oxide prevents currents from flowing between the gate and substrate electrodes through the gate oxide. Both interface and oxide quality allows for excellent isolation between the input and output terminals of the MOSFETs, causing it to behave as an almost ideal switch. Therefore, it is important to have good experimental tools to study the interface properties and the quality of the gate dielectric.

Electrical characterization of semiconductors and the semiconductor–dielectric interface is important for a variety of reasons. For example, the defects at and in the interfacial oxide layer in silicon–silicon dioxide (Si–SiO_2) systems and in the bulk semiconductor play critical roles in their low-frequency noise, independent of whether the device is surface-controlled such as a MOSFET, or a bulk transport device such as a polysilicon emitter bipolar junction transistor (PE BJT). These defects can affect the charge transfer efficiency in charge coupled devices (CCDs), p–n photodiodes or complementary metal-oxide-semiconductor (CMOS) imagers, and can be the initiation point of catastrophic failure of oxides. Interface and bulk states can act as scattering centers to reduce the mobility in MOSFETs, thus affecting their performance parameters such as switching speed, transconductance and noise. This chapter is devoted to the electrical characterization of semiconductors, insulators and interfaces. In the first part (Sects. 20.1 and 20.2), the basic electrical properties of materials (such as resistivity, concentration and mobility of carriers) are studied. The main measurement techniques used to determine these electrical parameters are presented. Due to its increasing importance in modern ultrasmall geometry devices, electrical contacts are also studied. All of the characterization techniques presented in this first part are associated with specially designed test structures. In the second part (Sects. 20.4–20.7), we use active components such as capacitors, diodes and transistors (mainly MOSFETs) in order to determine more specific electrical parameters such as traps, oxide quality and noise level that are associated with material or devices. Of course this involves specific measurement techniques that are often more sophisticated than those discussed in the previous two sections.

20.1 Resistivity

Resistivity is one of the most important electrical parameters of semiconductors [20.1–4]. First, we present the basic physical relations concerning the bulk resistivity. The main electrical measurement techniques are then described: the two oldest ones that are still relevant today – the *four-point-probe* technique and the *van der Pauw* technique – and then the *spreading resistance* technique. Second, because it is closely linked with bulk resistivity measurement techniques and it is increasingly important in modern ultrasmall geometry devices, contact resistivity will be presented. Special attention will be given to *Kelvin contact resistance* (*KCR*) *measurement* and the *transmission line measurement* (*TLM*) techniques.

20.1.1 Bulk Resistivity

Physical Approach, Background and Basics

The bulk resistivity ρ is an intrinsic electrical property related to carrier drift in materials such as metals and semiconductors [20.6]. From a macroscopic point of view, the resistivity ρ can be viewed as the normalization of the bulk resistance (R) by its geometrical dimensions – the cross-sectional area ($A = Wt$) through

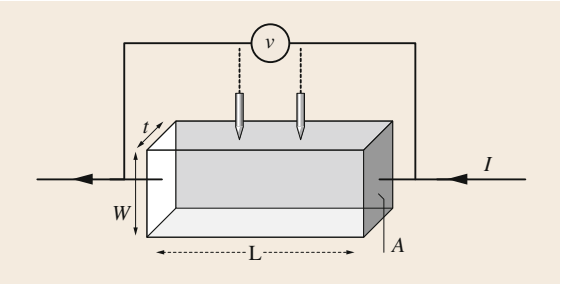


Fig. 20.1 Bulk resistance and its geometrical dimensions

which the current flows, and the distance between the two ideal contacts L, as shown in Fig. 20.1. The resistivity is given by

$$\rho = \frac{RA}{L} \quad \text{in } \Omega\,\text{m or commonly } \Omega\,\text{cm}. \tag{20.1}$$

For thin semiconductor layers, the sheet resistivity ρ_s is often used instead of the bulk resistivity ρ. The sheet resistivity ρ_s is the bulk resistivity divided by the sample's thickness t. This normalized parameter is related to the resistance of a square of side L. For this particular geometry in Fig. 20.1, since $A = Wt$, then $\rho_s = R_\square$, the sheet resistance. The unit of sheet resistance is Ω/square or $\Omega/\square$. The parameter $R_\square$ is convenient for integrated circuit designers because it allows them to quickly design the geometry for a specific value of resistance using very thin implanted or diffused semiconductor regions or polycrystalline layers. Resistivity (or its inverse, the conductivity σ in $\Omega^{-1}\,\text{cm}^{-1}$ or S/cm) and its variation with temperature is often used to classify material into metals, semiconductors and insulators.

Since different semiconductors can have the same resistivity, and also different values of resistivity can be found for a given semiconductor, depending on how it is processed for example, then resistivity is not a fundamental material parameter. From solid state theory, in the case of homogeneous semiconductor materials, the resistivity expresses the proportionality between the applied electric field E and the drift current density J; that is, $J = (1/\rho)\,E$. It can be defined by the microscopic relation

$$\rho = \frac{1}{q(n\mu_n + p\mu_p)}\,, \tag{20.2}$$

where q is the electronic charge, n and p are the free electron and hole concentrations, and μ_n and μ_p are the electron and hole drift mobilities, respectively. In this way, the resistivity is related to fundamental semiconductor parameters: the number of free carriers, and their ability to move in the lattice when an electric field is applied.

In n-type or donor (N_D)-doped, or p-type or acceptor (N_A)-doped semiconductors, the free carrier densities are determined by the ionized impurities (N_D or $N_A \gg$ the intrinsic carrier concentration n_i), then (20.2) can be simplified to

$$\rho \approx \frac{1}{qn\mu_n}\,, \text{ for an n-type semiconductor}\,, \tag{20.3}$$

$$\text{and } \rho \approx \frac{1}{qp\mu_p}\,, \text{ for a p-type semiconductor}\,. \tag{20.4}$$

In the following sections, only single-type semiconductors will be studied. This corresponds to most semiconductor materials used in electronic and optoelectronic devices because either $N_D \gg N_A$ or $N_A \gg N_D$ in a typical semiconductor layer.

Measurement Techniques

The simplest way to determine bulk resistivity is to measure the voltage drop along a uniform semiconductor bar through which a direct current (DC) I flows, as shown in Fig. 20.1. Thus, the measured resistance and knowledge of the geometrical dimensions can lead to an estimate for the bulk resistivity according to (20.1). Unfortunately the measured resistance (R_{mea}) includes the unexpected contact resistance ($2R_c$), which can be significant for small-geometry samples because R_c is strongly dependent on the metal-semiconductor structure. Therefore, special processing technologies are used to minimize the influence of R_c (Sect. 20.1.2). Now, the measured resistance is expressed as

$$R_{mea} = R + 2R_c\,. \tag{20.5}$$

If probes are used instead of large metal-semiconductor contacts, then the spreading resistance (R_{sp}) under the two probes must also be added, as shown in Fig. 20.2. In this case, (20.5) becomes

$$R_{mea} = R + 2R_{sp} + 2R_c\,, \tag{20.6}$$

where R_{sp} for a cylindrical contact of radius r, and for a semi-infinite sample, it can be expressed by

$$R_{sp} = \frac{\rho}{4r}\,. \tag{20.7}$$

For a hemispherical contact of radius r, R_{sp} is given by

$$R_{sp} = \frac{\rho}{2\pi r}\,. \tag{20.8}$$

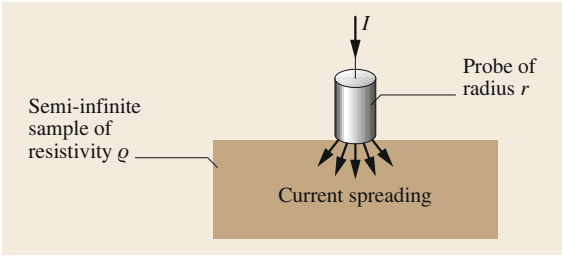


Fig. 20.2 Spreading resistance associated with a probe contact

In both cases, it is very difficult to provide a direct measurement or an accurate model of the contact resistance. So determining the bulk resistivity by this approach is not recommended, except when the spreading resistance is the dominant term in (20.6) and when (20.7) or (20.8) can be applied. In this case, the resistivity is determined by the *spreading resistance* technique measurement. Nevertheless, despite the lack of accuracy of the two contact techniques, it can be sufficient for monitoring some process steps and it is often used in the semiconductor industry as a process monitor.

Four-Point Probe Technique. In order to eliminate or at least minimize the contact contribution to the measured resistance value, techniques based on separate current injection and voltage drop measurements have been developed. First, the two-probe technique can be used, as reported in Fig. 20.1. This measurement is very simple, but it is affected by several parameters: lateral contact geometry, probe spacing, and minority carrier injection near the lateral contacts. The main disadvantage of this technique is the need for lateral contacts. This requirement is overcome with the four-point probe technique, where two probes are used for current injection and the other two probes are used to measure the voltage drop. The more usual probe geometry configuration is when the four probes are placed in a line, as shown in Fig. 20.3.

The voltage at probe 2, V_2, induced by the current flowing from probe 1 to probe 4 is given by:

$$V_2 = \frac{\rho I}{2\pi}\left(\frac{1}{s_1} - \frac{1}{s_2 + s_3}\right). \tag{20.9a}$$

The voltage at probe 3 is

$$V_3 = \frac{\rho I}{2\pi}\left(\frac{1}{s_1 + s_2} - \frac{1}{s_3}\right). \tag{20.9b}$$

Then, by measuring $V = V_2 - V_3$, the voltage drop between probes 2 and 3, and the current I through probes 1 and 4, the resistivity can be determined using (20.9a)

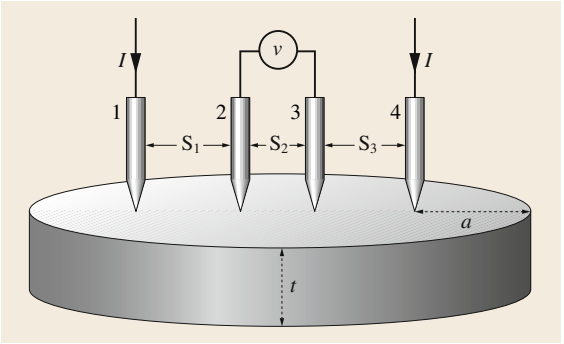


Fig. 20.3 Linear four-point probe configuration. The sample thickness is t and a is the distance from the edge or boundary of the sample

and (20.9b) as

$$\rho = \frac{2\pi V/I}{\left(\frac{1}{s_1} + \frac{1}{s_2} - \frac{1}{s_2+s_3} - \frac{1}{s_1+s_2}\right)} \tag{20.10}$$

Thus, a direct measurement of the resistivity can be made using a high-impedance voltmeter and a current source. When the probe spacings are equal ($s_1 = s_2 = s_3 = s$), which is the most practical case, then (20.10) becomes

$$\rho = 2\pi s\frac{V}{I}. \tag{20.11}$$

Equations (20.10) and (20.11) are valid only for semi-infinite samples; that is, when both t and the sample surface are very large ($\to \infty$), and the probes' locations must be far from any boundary. Because these relations can be applied only to large ingots, then in many cases a correction factor f must be introduced in order to take into account the finite thickness and surface of the sample and its boundary effects. Further, for epitaxial layers, f must also consider the nature of the substrate – whether it is a conductor or an insulator. Thus, (20.11) becomes

$$\rho = 2\pi s\frac{V}{I}f. \tag{20.12}$$

For a thin semiconductor wafer or thin semiconducting layer deposited on an insulating substrate, and for the condition $t < s/2$, which represents most practical cases because the probe spacing s is usually on the order of a millimeter, then the correction factor due to the thickness is

$$f = \frac{(t/s)}{2\ln 2} \quad \text{so that} \quad \rho = 4.532t\frac{V}{I}. \tag{20.13}$$

The noninfinite sample surface must be corrected if the ratio of the wafer diameter to the probe spacing is not

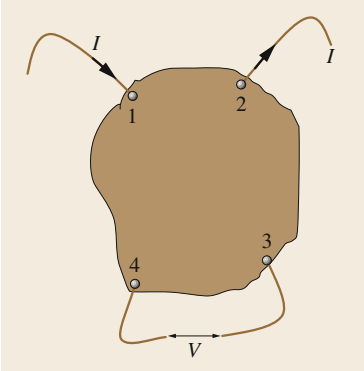


Fig. 20.4 van der Pauw method for an arbitrarily shaped sample

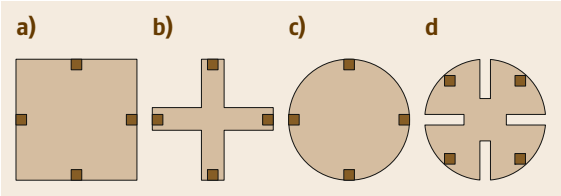


Fig. 20.5a–d Symmetrical van der Pauw structures: **(a)** square, **(b)** Greek cross, **(c)** circle and **(d)** cloverleaf

greater than 40, otherwise a correction factor of less than unity has to be introduced [20.3].

If the probe header is too close to any boundary, then (20.13) is no longer valid and another correction factor must be introduced. This correction factor is close to 1 until the ratio a/s is greater than 2, where a is the distance from the edge of the sample that is shown in Fig. 20.3. A study of various 8 inch n- and p-type silicon wafers have demonstrated that the edge exclusion limit is 5 mm [20.7].

In the case of a different arrangement of probes, for instance a square array or when a different measurement configuration of the four-point collinear probes is used, such as current injection between probes 1 and 3, other specific correction factors are required. Here, rather than detail all the different correction factors, complementary information can be found in [20.4, Chap. 4] and [20.3, Chap. 1].

Taking into account the appropriate correction factors as well as some specific material parameters such as hardness or surface oxidation, it is possible to map the resistivities of different types of semiconductor wafers or deposited semiconductor layers with an accuracy better than 1% over a large range of resistivity values using commercial equipment and appropriate computational techniques.

Van der Pauw Technique. Based on the same basic principle of separating the current injection and voltage measurement, the *van der Pauw* [20.8] measurement technique allows for the determination of resistivity on a sample of arbitrary shape using four small contacts placed on the periphery, as shown in Fig. 20.4. Then, the resistivity of a uniform sample of thickness t is given by

$$\rho = \frac{\pi t}{\ln 2}\frac{(R_\mathrm{A}+R_\mathrm{B})}{2} f \,. \tag{20.14}$$

Here, R_A and R_B are resistances measured by injecting current on two adjacent contacts and by measuring the voltage drop on the two remaining ones. With the notation in Fig. 20.4, one can define

$$R_\mathrm{A} = \frac{V_3 - V_4}{I_{1,2}} \,, \quad R_\mathrm{B} = \frac{V_4 - V_1}{I_{2,3}} \,,$$

f is a correction factor that is

$$\text{a function of the ratio} \quad R_f = \frac{R_\mathrm{A}}{R_\mathrm{B}} \,. \tag{20.15}$$

with R_f obtained from

$$\frac{(R_f - 1)}{(R_f + 1)} = \frac{f}{\ln 2}\operatorname{arccosh}\left(\frac{\exp(\ln 2/f)}{2}\right). \tag{20.16}$$

In the case of samples with symmetrical geometries, and when the contacts are also symmetrical, as shown in Fig. 20.5, then $R_\mathrm{A} = R_\mathrm{B}$, $R_f = 1$ and $f = 1$, and (20.14) becomes

$$\rho = \frac{\pi t}{\ln 2} R_\mathrm{A} = 4.532 t R_\mathrm{A} \,. \tag{20.17}$$

In order to minimize errors caused by the finite dimensions of the contacts (since ideally the contact area should be zero) and the finite thickness of the sample, then the distance between the contacts must be larger than both the diameter and the thickness of the contact. Also, the cloverleaf configuration in Fig. 20.5d is recommended to prevent contact misalignment, but this configuration requires a more complicated patterning technology.

The main advantage of the van der Pauw technique compared to the four-point probe technique is its use of a smaller area for the test structure. Therefore, this measurement technique is often used in integrated circuit technology. Also, because of its simple structure, the Greek cross configuration in Fig. 20.5b is widely used (experimental results obtained on SiGeC epitaxial layers are reported in Fig. 20.6 as an example). However, when narrows arms are used, current crowding at the corners may have a significant influence, and in this case a different Greek cross layout can be considered to reduce this current crowding effect [20.9].

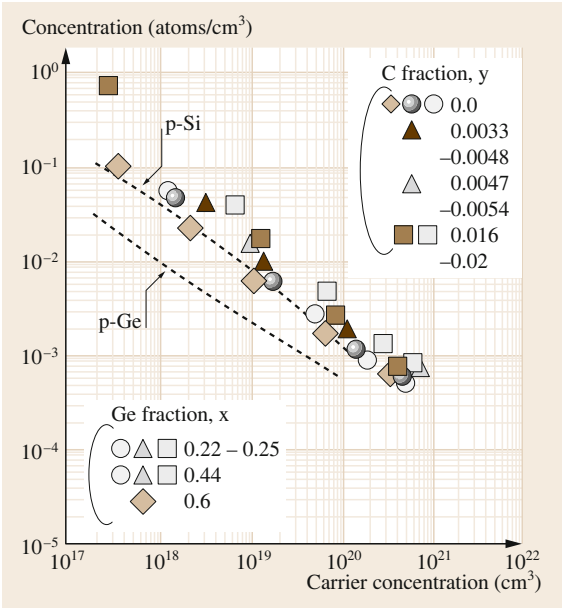


Fig. 20.6 Resistivity versus carrier concentration in $Si_{1-x-y}Ge_xC_y$ films obtained using a cloverleaf van der Pauw structure. (After [20.10])

Spreading Resistance Technique. The spreading resistance technique is based on the modeling of current spreading from a probe tip or a small metallic contact and flowing into a bulk semiconductor, as shown previously in Fig. 20.2. Equations (20.7) and (20.8) presented above are for cylindrical probes and hemispherical probes, respectively. Basically, the principle of this method is opposite to the previous four-contact techniques where the separation of the current injection from the measured voltage drop was used to avoid the spreading resistance. Here, the spreading resistance is expected to be the dominant term in (20.6). Only two contacts are needed: two closely aligned probes, a small top contact probe or a metallic contact and a large bottom contact. In the first case, surface mapping can be performed, but the main use of this compact probe configuration is for resistivity profiling using a bevelled sample [20.3]. The second configuration has been used to measure the substrate resistivity of silicon integrated circuits where simple test structures – for example the square top contact of $25\,\mu m \times 25\,\mu m$ and $50\,\mu m \times 50\,\mu m$ shown in [20.12] – have been included on a test chip.

More recently, semiconductor resistivity has been nanocharacterized using scanning spreading resistance microscopy (SSRM) with a standard atomic force microscope (AFM) of lateral resolution of 10 to 20 nm. A SSRM image of a $0.5\,\mu m$ nMOSFET is given in Fig. 20.7a [20.11]. The resistance is low in the highly doped regions (dark) and high in the lower doped regions (bright): source, drain, gate and well regions are clearly observed in Fig. 20.7a or the resistance profile in Fig. 20.7b. With such a high resolution, scanning the lateral and vertical diffusion of dopants in active regions of submicron transistors is possible. An example is shown in Fig. 20.7c where the extra implantations (halo and lightly doped drain (LDD)) in a $0.35\,\mu m$ nMOSFET process are clearly visible and result in a change of L_{eff} from 295 nm without extra implantations to 229 nm with the extra implantations.

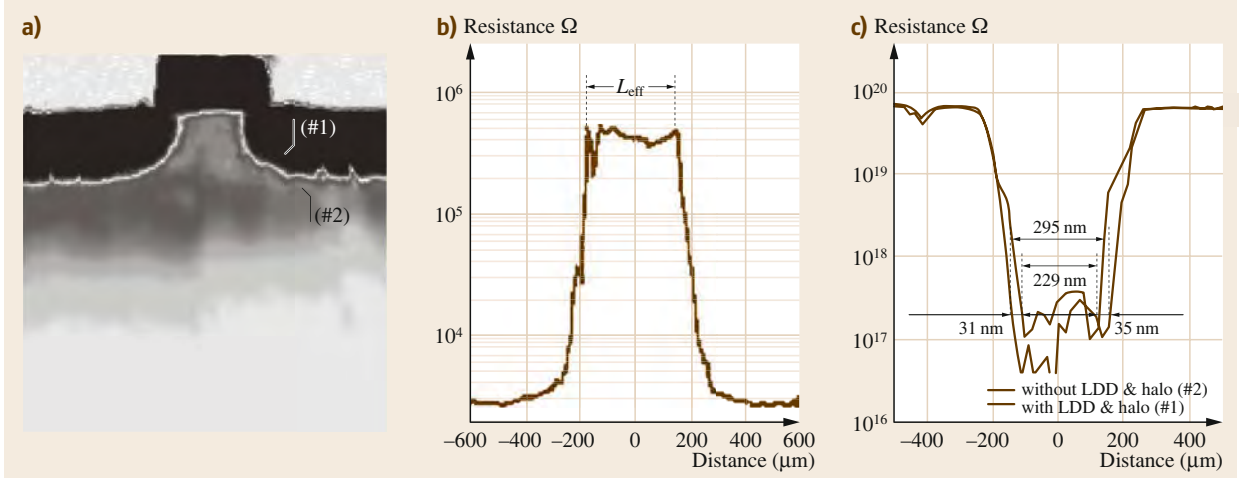


Fig. 20.7 **(a)** SSRM resistance image (scan size: $1.5\,\mu m \times 1.5\,\mu m$) of a $0.5\,\mu m$ nMOSFET; **(b)** lateral section taken 10 nm under the gate oxide of the same transistor; **(c)** lateral carrier concentration profiles measured with SSRM 10 nm under the gate oxide for two $0.35\,\mu m$ nMOSFETs (with and without halo and LDD process). (After [20.11])

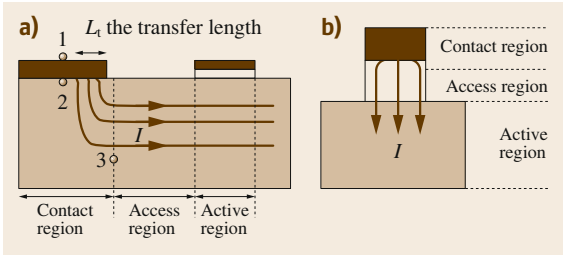


Fig. 20.8 (a) Horizontal contact and (b) vertical contact. *Black* indicates the metallic conductor, *white* the semiconductor material or an insulator

20.1.2 Contact Resistivity

The contact resistance of an active device and interconnection becomes larger as the dimensions are scaled down. As a consequence, the performance of single transistors as well as integrated circuits can be seriously limited by increasing *RC* time constants and power consumption. This is of major interest for the semiconductor industry, as reported by the International Technology Roadmap for Semiconductors, ITRS 2001 [20.5], and in [20.13].

Contact Resistance Elements

Basically, the contact resistance R_c is the resistance localized from a contact pad, a probe or from the metallization process to an active region. However, it does not include all of the access resistances between these two regions, as shown in Fig. 20.8a for a horizontal contact and Fig. 20.8b for a vertical contact.

Starting from the contact pad (Figs. 20.8a and 20.9), the contact resistance includes the resistance of the metal R_m, the interfacial metal-semiconductor resistance R_i, and the resistance associated with the semiconductor just below the contact in the contact region R_{sc}. Thus, the contact resistance can be expressed as

$$R_c = R_m + R_i + R_{sc} \,. \tag{20.18}$$

The last component R_{sc} cannot be accurately defined because the boundary between the contact and access regions is very difficult to determine due to (for example) interdiffusion of metal and semiconductor atoms, and because the current flow into this region is not homogeneous due to current spreading and lateral or vertical current crowding at the periphery of the contact. The relative importance of each component of R_c is strongly dependent on different parameters of the process itself – annealing temperature, doping density and the geometry used (lateral or vertical).

When comparing different contact technologies and different contact areas, the most convenient parameter

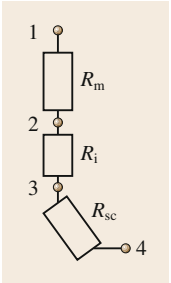


Fig. 20.9 Different components of the contact resistance

to use is the contact resistivity ρ_c, which is referred to as the specific contact resistance in $\Omega\,cm^2$, and ρ_c is given by

$$\rho_c = R_c A_{ceff} \,, \tag{20.19}$$

where A_{ceff} is the effective contact area; that is, the current injection area. The concept of an effective contact area can be approximated by the contact geometry in the case of a vertical contact in Fig. 20.8b. However, A_{ceff} is more difficult to specify for a lateral contact, where a transfer length L_T, representing the length where the current flow transfers from the contact into the semiconductor just underneath, must be introduced, as shown in Fig. 20.8a. L_T is defined as the length over which the voltage drops to e^{-1} of its value at the beginning of the contact [20.3], and is given by

$$L_T = \sqrt{\frac{\rho_c}{\rho_{sc}}} \,, \tag{20.20}$$

where ρ_{sc} is the sheet resistivity of the semiconductor below the contact.

Because of its various components, it is difficult to accurately model the contact resistivity. Nevertheless, a theoretical approach to the interfacial resistivity (see R_i in (20.18)), ρ_i, can be determined from the well-known Schottky theory of metal–semiconductor contacts. The interfacial resistivity ρ_i is defined by

$$\rho_i = \left.\frac{\partial V}{\partial J}\right|_{V=0} \,. \tag{20.21}$$

This metal–semiconductor structure is equivalent to an abrupt p-n junction. According to the Schottky theory (for more details see [20.14, Chap. 5]), the J–V characteristic of a metal–semiconductor contact in the case of a low-doped semiconductor is given by

$$J = A^* T^2 \exp\left(-\frac{q\phi_B}{kT}\right)\left[\exp\left(\frac{qV}{kT}\right) - 1\right], \tag{20.22}$$

where A^* is Richardson's constant, and T the absolute temperature. ϕ_B is the barrier height formed at

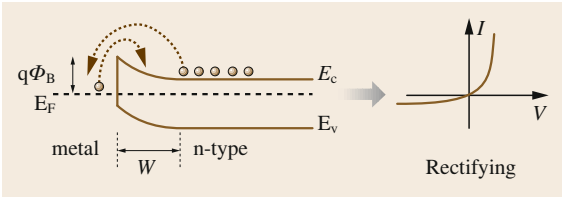


Fig. 20.10 Energy-band diagram of an n-type semiconductor–metal contact and related rectifying contact. W is the width of the depletion layer

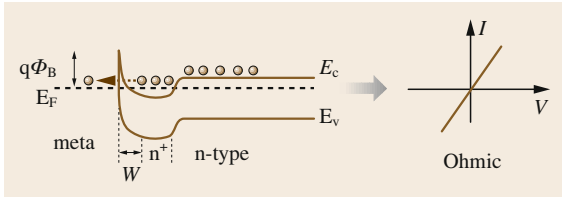


Fig. 20.11 Energy-band diagram of an n^+-n semiconductor–metal structure and related ohmic contact

the metal–semiconductor interface – the difference between the vacuum level and the Fermi level of the metal and of the semiconductor materials respectively, and ϕ_B is given by

$$\phi_B = \phi_M - \chi \tag{20.23}$$

where ϕ_M is the metal work function and χ the semiconductor electron affinity.

The energy band diagram for a low-doped n-type semiconductor–metal contact is shown in Fig. 20.10. In this case, the current transport is dominated by the thermionic emission current, resulting in a rectifying contact.

Thus, when the conduction mechanism is controlled by the thermionic emission (TE), the interfacial resistivity in (20.21) is simply obtained from the derivative of (20.22), and $\rho_{i,TE}$ is

$$\rho_{i,TE} = \frac{k}{qA^*T} \exp\left(\frac{q\phi_B}{kT}\right) \tag{20.24}$$

Due to the presence of surface states, the barrier height ϕ_B is positive and weakly dependent on the metal–semiconductor material. ϕ_B is $\approx 2E_g/3$ for an n-type semiconductor and $\approx E_g/3$ for a p-type semiconductor. Therefore, high values of interfacial resistivity $\rho_{i,TE}$ are usually obtained except when narrow bandgap semiconductors are used.

The way to fabricate ohmic contacts with low contact resistivity values is to process the metal on a heavily doped semiconductor layer. In this case, the depletion width decreases ($W \approx N_D^{-1/2}$) and the probability of carrier tunneling through the barrier increases. Thus, the conduction mechanism is dominated by tunneling, as shown in Fig. 20.11.

The electron tunneling current is expressed as

$$J_{tun} \approx \exp\left(\frac{q\phi_B}{E_{00}}\right), \tag{20.25}$$

where

$$E_{00} = \frac{q\hbar}{2}\sqrt{\frac{N_D}{\varepsilon_s m_n^*}}, \tag{20.26}$$

ε_s is the permittivity of the semiconductor and m_n^* is the effective mass of the electron.

From (20.21), (20.25) and (20.26), the interfacial resistivity $\rho_{i,T}$ is found to be

$$\rho_{i,T} \propto \frac{2\sqrt{\varepsilon_s m_n^*}}{\hbar} \frac{\phi_B}{\sqrt{N_D}}. \tag{20.27}$$

Comparing $\rho_{i,TE}$ from (20.24) to $\rho_{i,T}$ from (20.27), we see that a highly doped layer can significantly reduce the interfacial resistivity. For $N_D \geq 10^{19}\,cm^{-3}$, the tunneling process dominates the interfacial resistivity, while for $N_D \leq 10^{17}\,cm^{-3}$, the thermionic emission current is dominant.

As most semiconductors such as Si, SiGe, GaAs, InP are of relatively wide bandgap, the deposition of a heavily doped layer before the metallization is commonly used in order to form a tunneling contact. For compound semiconductor manufacturing processes, the contact layer is generally formed from the same semiconductor material, or at least from the same material as the substrate. For silicon and related materials such as SiGe alloys or polysilicon, silicidation techniques are commonly used to make the contact layer with very thin silicide layers such as $CoSi_2$ or $TiSi_2$ layers.

Measurement Techniques

As mentioned above, it is difficult to accurately model the contact resistance, so direct measurements of the contact resistance or of the contact resistivity are of great importance. The two main test structures used to determine contact characteristics will now be discussed: the *cross Kelvin resistor* (*CKR*) *test structure* and the *transmission line model* (*TLM*) structure.

Kelvin Test Structure. The Kelvin test structure, also referred to as the cross Kelvin resistor (CKR) test structure, is shown in Fig. 20.12. The contact resistance R_c is determined from the potential drop in the contact window (V_{34}) when a current I is forced through the contact window from contact pad 1 to pad 2, and R_c is

$$R_c = \frac{V_{34}}{I}. \tag{20.28}$$

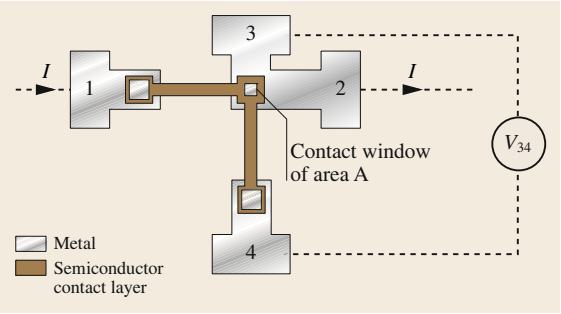


Fig. 20.12 Cross Kelvin resistor test structure

Therefore, a measure of R_c and knowledge of the contact area A allows for direct extraction of the contact resistivity ρ_c, given by

$$\rho_c = R_c A \,. \tag{20.29}$$

This basic approach is not valid when parasitic effects are present. One of the main problems is current crowding around the contact. In order to extract accurate values for the contact resistivity using Kelvin test structures, it is necessary to take into account the two-dimensional current-crowding effect. This is achieved using the results from numerical simulations [20.15]. Nevertheless, the development of ohmic contacts with very low values of contact resistivities require complex technology with different materials and usually with several interfaces. In this case, a large discrepancy between the extracted and the measured contact parameters can be found [20.15, 16]. To improve the accuracy, three-dimensional models are now used to take into account the different interfacial and vertical parasitic effects [20.17].

Transmission Line Model Test Structures. The transmission line model test structure (TLM) consists of depositing a metal grid pattern of unequal spacing L_i between the contacts. This leads to a scaled planar resistor structure. Each resistor changes only by its distance L_i between two adjacent contacts, as shown in Fig. 20.13, and it can be expressed by

$$R_i = \frac{\rho_s L_i}{W} + 2R_c \tag{20.30}$$

Then, by plotting the measured resistances as a function of the contact spacing L_i, and according to (20.30), the layer sheet resistivity ρ_s and the contact resistance R_c can be deduced from the slope and from the intercept at $L_i = 0$ respectively, as shown in Fig. 20.14

$$\text{Slope} = \frac{\rho_s}{W}\,;\; R_i(\text{intercept}) = 2R_c\,;$$
$$\text{and}\;\; |L_i(\text{intercept})| = 2L_T \tag{20.31}$$

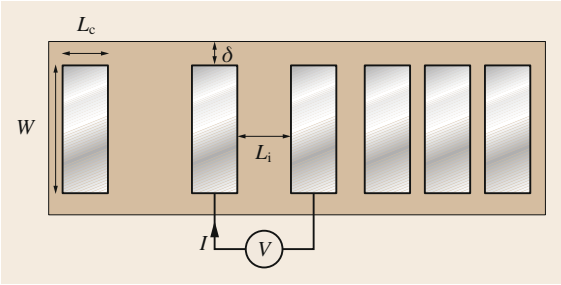


Fig. 20.13 Transmission line model (TLM) test structure

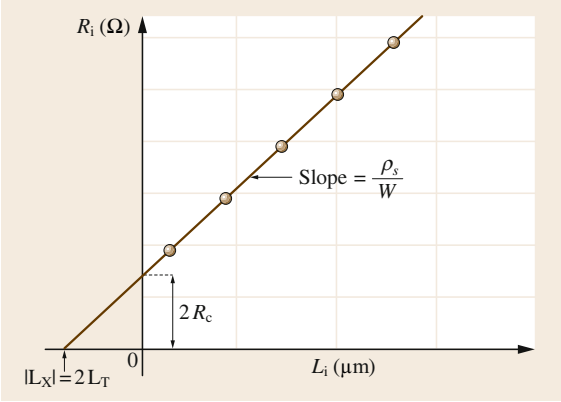


Fig. 20.14 Determination of the sheet resistivity and characterization of the contact using a TLM test structure

As discussed in Sect. 20.1.2, the most suitable parameter for characterizing a contact is its contact resistivity (ρ_c) or the specific contact resistance ($R_c A_{ceff}$), given by

$$\rho_c = R_c A_{ceff} = R_c W L_T \tag{20.32}$$

As shown in Fig. 20.8a, for a planar resistor, the effective contact area requires the notion of the transfer length L_T. According to (20.20), and assuming that the sheet resistance under the contact ρ_{sc} is equal to the sheet resistance between the contacts ρ_s, then L_T can be expressed by (20.20).

Therefore, the substitution of R_c into (20.32) in (20.30) leads to

$$R_i = \frac{\rho_s}{W} L_i + \frac{\rho_s}{W} 2L_T \tag{20.33}$$

Now, extrapolation to $R_i = 0$ allows us to determine the value of L_T. The main advantage of the *TLM* method is its ability to give two main electrical parameters, the resistivity of the semiconductor contact layer ρ_s and the contact resistance R_c. However, this is done at the expense of a questionable assumption that the sheet resistance under the contact must be equal to the sheet resistance between the contacts. More on this technique can be found in [20.3].

20.2 Hall Effect

As mentioned before, the resistivity of a semiconductor is not a fundamental material parameter. One can consider the carrier density (n or p) or the carrier mobility (μ_n or μ_p) to be fundamental or microscopic parameters. For a semiconductor material, the resistivity is related to these two parameters (density and mobility) by (20.2). The strength of the Hall effect is to directly determine the sheet carrier density by measuring the voltage generated transversely to the current flow direction in a semiconductor sample when a magnetic field is applied perpendicularly, as shown in Fig. 20.15a. Together with a resistivity measurement technique such as the four-point probe or the van der Pauw technique, Hall measurements can be used to determine the mobility of a semiconductor sample.

In modern semiconductor components and circuits, knowledge of these two fundamental parameters n/p and μ_n/μ_p is critical. Currently, Hall effect measurements are one of the most commonly used characterization tools in the semiconductor industry and research laboratories. This is not just because of the parameters that can be extracted for use in device modeling or materials characterization, but also because of the quantum Hall effect (QHE) in condensed matter physics [20.18]. Moreover, in the applied electronics domain, one should note the development of different sensors based on the physical principle of the Hall effect, such as commercial CMOS Hall sensors.

As is very often the case, the development of a characterization technique is related to its cost, simplicity of implementation and ease of use. Since these practical characteristics are satisfied even when specially shaped samples are required, then the Hall effect measurement technique has become a very popular method of characterizing materials.

In this section, we will first present the physical principle of the Hall effect. Then we will show how it can be used to determine the carrier density and mobility. Finally, the influence of the Hall scattering factor will be presented, followed by some practical issues about the implementation of the Hall effect method.

20.2.1 Physical Principles

The Hall effect was discovered by *Hall* in 1879 [20.19] during an experiment on current transport in a thin metal strip. A small voltage was generated transversely when a magnetic field was applied perpendicularly to the conductor.

The basic principle of this Hall phenomenon is the deviation of some carriers from the current line due to the Lorentz force induced by the presence of a transverse magnetic field. As a consequence, a voltage drop V_H is induced transversely to the current flow. This is shown in Fig. 20.15a for a p-type bar-shaped semiconductor, where a constant current flow I_x in the x-direction and a magnetic field in the z-direction results in a Lorentz force on the holes. If both holes and electrons are present, they deviate towards the same direction. Thus, the directions of electrical and magnetic fields must be accurately specified.

The Lorentz force is given by the vector relation

$$F_L = q(\boldsymbol{v} \times \boldsymbol{B}) = -q v_x B_z \,, \tag{20.34}$$

where v_x is the carrier velocity in the x-direction. Assuming a homogeneous p-type semiconductor

$$v_x = \frac{I}{qtWp} \,. \tag{20.35}$$

As a consequence, an excess surface electrical charge appears on one side of the sample, and this gives rise to an electric field in the y-direction E_y. When the magnetic force F_L is balanced by the electric force F_{EL}, then the Hall voltage V_H is established, and from a balance between F_L and F_{EL}, we get

$$F = F_L + F_{EL} = -q v_x B_z + q E_y = 0 \,, \tag{20.36}$$

$$\text{so } E_y = \frac{BI}{qtWp} \,. \tag{20.37}$$

Also, the Hall voltage V_H is given by

$$V_H = V_y = E_y W = \frac{BI}{qtp} \,. \tag{20.38}$$

So if the magnetic field B and the current I are known, then the measurement of the Hall voltage gives the hole sheet concentration p_s from

$$p_s = pt = \frac{BI}{qV_H} \,. \tag{20.39}$$

If the conducting layer thickness t is known, then the bulk hole concentration can be determined (see (20.41)) and expressed as a function of the Hall coefficient R_H, defined as

$$R_H = \frac{tV_H}{BI} \tag{20.40}$$

$$\text{and } p = \frac{1}{qR_H} \tag{20.41}$$

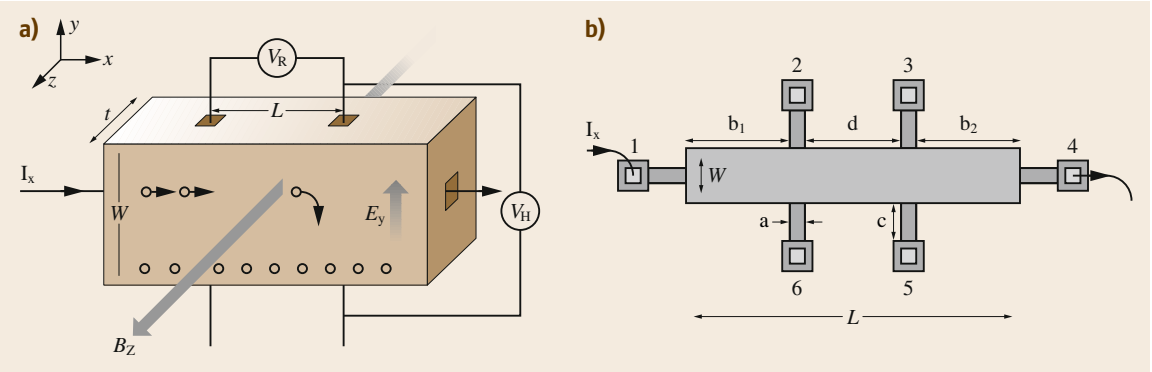


Fig. 20.15 **(a)** Representation of the Hall effect in an p-type bar-shaped semiconductor. **(b)** Practical sample geometry: a six-terminal Hall-bar geometry

Using the same approach for an n-type homogeneous semiconductor material leads to

$$R_H = -\frac{tV_H}{BI}\,, \tag{20.42}$$

$$\text{and } n = -\frac{1}{qR_H} \tag{20.43}$$

Now, if the bulk resistivity ρ is known or can be measured at the same time using a known sample such as a Hall bar or van der Pauw structure geometry in zero magnetic field, then the carrier drift mobility can be obtained from

$$\mu = \frac{|R_H|}{\rho} \tag{20.44}$$

There are two main sample geometries commonly used in Hall effect measurements in order to determine either the carrier sheet density or the carrier concentration if the sample thickness is known, and the mobility. The first one is the van der Pauw structure presented in Sect. 20.1.1. The second one is the Hall bar structure shown in Fig. 20.15b, where the Hall voltage is measured between contacts 2 and 5, and the resistivity is measured using the four-point probes technique presented in Sect. 20.1.1 (contacts 1, 2, 3 and 4). Additional information about the shapes and sizes of Hall structures can be found in [20.3, 4, 20].

Whatever the geometry used for Hall measurements, one of the most important issues is related to the offset voltage induced by the nonsymmetric positions of the contact. This problem, and also those due to spurious voltages, can be controlled by two sets of measurements, one for a magnetic field in on direction and another for a magnetic field in the opposite direction.

The Hall effect has also been investigated on specific structures, and an interesting example can be found in reference [20.21], where a Hall bar structure was combined with a double-gate n-silicon-on-insulator (SOI) MOSFET. This was done in order to understand the mobility behavior in ultra-thin devices and to validate the classical drift mobility extraction method based on current–voltage measurements.

In the Hall effect experiment, the measurement of the Hall coefficient R_H leads to the direct determination of the carrier concentration and mobility. Moreover, the sign of R_H can be used to determine the type of conductivity of the semiconductor sample. If various types of carriers are present, then the expression for R_H becomes more complex and approximations in the limit of low and high magnetic field are necessary [20.3, Chap. 8].

We have so far discussed the Hall effect on a uniformly doped substrate or single semiconductor layer deposited on an insulating or semi-insulating substrate. In the case of a semiconductor layer deposited on a semiconducting substrate of opposite doping type, Hall effect measurements can be performed if the space charge region can act as an insulator. In the case of multilayers, the problem is more difficult, but an approximation for transport experiments has been developed for two-layer structures [20.22] and applied to different metal–semiconductor field-effect-transistor (MESFET) structures, for instance [20.23].

20.2.2 Hall Scattering Factor

The relations presented above are based on an energy-independent scattering mechanism. With this assumption made, the Hall carrier concentration and the Hall mobility are equal to the carrier concentration and the carrier drift mobility. When this assumption is no longer valid, these electrical parameters are different and the Hall scattering factor r_H must be taken into account. In this case (20.41), (20.43) and (20.44) must be modified

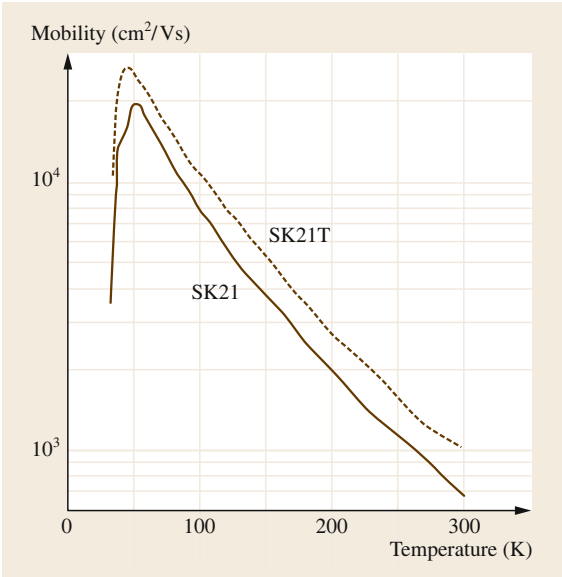


Fig. 20.16 Hall mobility as a function of temperature on two SOI films. (After [20.24])

as follows

$$p_{\mathrm{H}} = \frac{r_{\mathrm{H}}}{qR_{\mathrm{H}}} = r_{\mathrm{H}} p \,, \tag{20.45}$$

$$n_{\mathrm{H}} = -\frac{r_{\mathrm{H}}}{qR_{\mathrm{H}}} = r_{\mathrm{H}} n \,, \tag{20.46}$$

and

$$\mu_{\mathrm{H}} = r_{\mathrm{H}} \frac{|R_{\mathrm{H}}|}{\rho} = r_{\mathrm{H}} \mu \,. \tag{20.47}$$

The Hall scattering factor [20.25] is related to the energy dependence of the mean free time between carrier collisions $\tau(E)$, and r_{H} is given by

$$r_{\mathrm{H}} = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} \,. \tag{20.48}$$

According to theory [20.3], the Hall scattering factor tends to unity in the limit of high magnetic field. Therefore, r_{H} at low magnetic fields can be determined by measuring the Hall coefficient in the limit of both high and low magnetic fields [20.25] using

$$r_{\mathrm{H}} = \frac{R_{\mathrm{H}}(B)}{R_{\mathrm{H}}(\infty)} \tag{20.49}$$

Depending on the scattering mechanism involved (lattice, ionized or neutral impurity, electron, or phonon scattering), r_{H} is found to vary between 0.6 and 2 [20.26]. However, due to valence band distortion effects, values as low as 0.26 have been found in strained p-type SiGe epilayers [20.27]. Therefore, the Hall carrier concentration and especially the Hall mobility must be distinguished from carrier concentration and carrier drift mobility.

As the different scattering mechanisms have different temperature (T) dependences, then the Hall mobility as function of temperature is often used to separate the different scattering processes. An example is given in Fig. 20.16 for silicon-on-insulator (SOI) films [20.24]. The increase in the mobility between 4 and 45 K, which is given by $\mu \propto T^{2.95}$, is related to the ionized donor scattering mechanism. The decrease in mobility between 46 and 120 K given by $\mu \propto T^{-1.55}$ is associated with lattice scattering. However, after 150 K, the rapid decrease in mobility observed, where $\mu \propto T^{-2.37}$, suggests that other scattering mechanisms as well as the lattice scattering mechanism, such as electron or phonon scattering, must be taken into account.

20.3 Capacitance–Voltage Measurements

Capacitance–voltage (C–V) measurements are normally made on metal-oxide semiconductor (MOS) or metal-semiconductor (MS) structures in order to determine important physical and defect information about the insulator and semiconductor materials. For example, high-frequency (HF) and low-frequency (LF) or quasi-static C–V measurements in these structures are used to determine process and material parameters – insulator thickness, doping concentration and profile, density of interface states, oxide charge density, and work function or barrier height. In this section, we describe various C–V measurements and how they can be used to provide process parameters as well as valuable information about the quality of the materials. A typical C–V curve for a MOS capacitor with an n-type semiconductor is shown in Fig. 20.17. For a MOS capacitor with a p-type substrate, the C–V curve be similar to that in Fig. 20.17, but reflected about the y-axis.

20.3.1 Average Doping Density by Maximum–Minimum High-Frequency Capacitance Method

The maximum–minimum high-frequency (HF) capacitance method uses the HF capacitance under strong accumulation (C_{OX}) and strong inversion ($C_{\mathrm{HF,min}}$) to

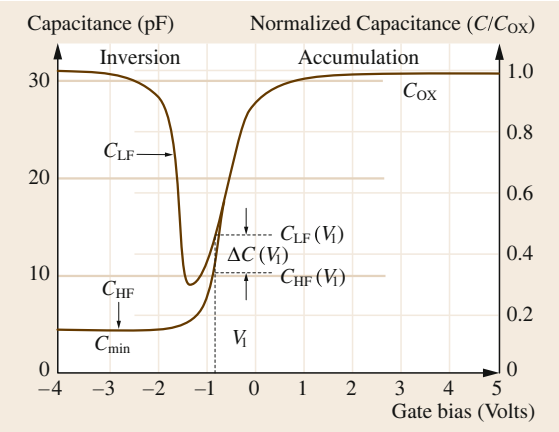


Fig. 20.17 Typical C–V curve for a MOS capacitor on an n-type substrate. (After [20.28])

determine the average doping density [20.29, pp. 406–408]. Note that under strong inversion and at high frequencies, the interface trap capacitance is negligible ($C_{it} \approx 0$). Under strong inversion, the depletion width (w_{max}) is a maximum and so the high frequency capacitance per unit area $C_{HF,min}$ is a minimum, since the minority carriers cannot respond to the high-frequency signal. Since the inversion layer is very thin compared to the depletion layer, then

$$w_{max} = \varepsilon_{Si}\left(\frac{1}{C_{HF,min}} - \frac{1}{C_{OX}}\right), \tag{20.50}$$

where ε_{Si} is the permittivity of silicon and C_{OX} is the gate oxide capacitance per unit area.

At the conditions for w_{max}, the band bending ψ_{max} is a maximum, and it is

$$\begin{aligned}\psi_{max} &= 2\phi_B + \frac{kT}{q}\ln\left(2\frac{q}{kT}\phi_B - 1\right)\\ &= 2\frac{kT}{q}\left\{\ln\left(\frac{n}{n_i}\right) + \frac{1}{2}\ln\left[2\ln\left(\frac{n}{n_i}\right) - 1\right]\right\},\end{aligned} \tag{20.51}$$

where $\phi_B = (k_B T/q)\ln(n/n_i)$ is the shift of the Fermi level from the intrinsic Fermi level $\phi_i = (E_c - E_v)/2q$ in the bulk of the silicon in the MOS structure due to the doping concentration n, and n_i is the thermally generated carrier concentration in silicon. For a uniformly doped sample,

$$w_{max}^2 = \frac{2\varepsilon_{Si}\psi_{max}}{qn} \tag{20.52}$$

and from (20.50) and (20.52), a relation between the doping concentration n and the measured capacitance

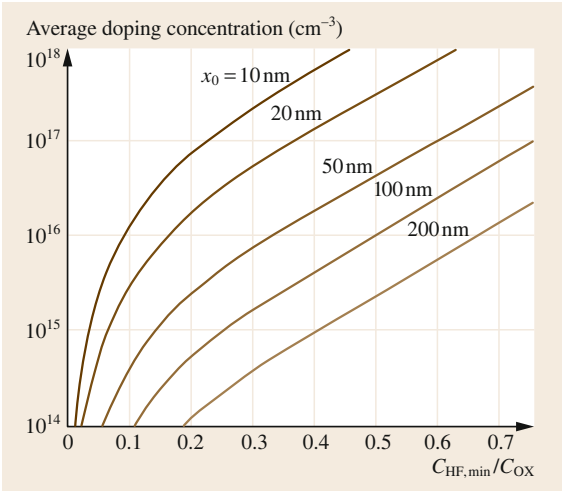


Fig. 20.18 Doping concentration n as function of $C_{HF,min}/C_{OX}$ with oxide thickness, based on (20.53). (After [20.29])

can be established [20.29, p. 407] as

$$\begin{aligned}&\frac{n}{\ln\left(\frac{n}{n_i}\right) + \frac{1}{2}\ln\left[2\ln\left(\frac{n}{n_i}\right) - 1\right]}\\ &= \frac{4kT}{q^2\varepsilon_{Si}}\frac{C_{OX}^2}{\left(\frac{C_{OX}}{C_{HF,min}} - 1\right)^2}.\end{aligned} \tag{20.53}$$

Equation (20.53) is a transcendental equation in average doping concentration n that can be solved numerically by iteration. Figure 20.18 shows the solutions as function of $C_{HF,min}/C_{OX}$ with oxide thickness, and this can be used to obtain the average doping n graphically. Equation (20.53) can be further simplified by neglecting the term $0.5\ln[2\ln(n/n_i) - 1]$, and assuming $C_{OX} = C_{HF,max}$ [20.30]. Also, an approximation of (20.53) for the average doping concentration n in unit cm^{-3} is obtained in [20.4] and [20.31] for silicon MOS structures at room temperature, and this is given by

$$\begin{aligned}\log_{10}(n) = {} & 30.38759 + 1.68278\\ & \times \log_{10}(C_{DM} - 0.03177)\\ & \times [\log_{10}(C_{DM})]^2,\end{aligned} \tag{20.54}$$

where the depletion capacitance (per cm^2 of area) C_{DM} is defined as

$$C_{DM} = \frac{C_{HF,min}C_{OX}}{C_{OX} - C_{HF,min}}, \tag{20.55}$$

where all capacitances are in units of F/cm^2.

20.3.2 Doping Profile by High–Frequency and High–Low Frequency Capacitance Methods

The doping profile in the depletion layer can be obtained [20.29, Sect. 9.4] by assuming that the depletion capacitance per unit area C_D and the oxide capacitance per unit area C_{OX} are connected in series; that is, that the measured high frequency capacitance C_{HF} is given by

$$\frac{1}{C_{HF}} = \frac{1}{C_D} + \frac{1}{C_{OX}} \Rightarrow \frac{1}{C_D} = \frac{1}{C_{HF}} - \frac{1}{C_{OX}} \,. \quad (20.56)$$

For a particular gate biasing V_G of the MOS structure, the depletion thickness $w(V_G)$ is obtained from C_D as

$$w(V_G) = \varepsilon_{Si}\left(\frac{1}{C_{HF}} - \frac{1}{C_{OX}}\right). \quad (20.57)$$

The doping concentration $n(V_G)$ is given by the slope of the $(1/C_{HF})^2$ versus V_G characteristic, given by

$$n(w) = \frac{-2}{q\varepsilon_{Si}\frac{\partial}{\partial V_G}\left(\frac{1}{C_{HF}^2}\right)} \,. \quad (20.58)$$

Note that a plot of $1/C_{HF}^2$ versus V_G (Fig. 20.19) can yield important information about the doping profile. The average n is related to the reciprocal of the slope in the linear part of the $1/C_{HF}^2$ versus V_G curve, and the intercept with V_G at a value of $1/C_{OX}^2$ is equal to the flat-band voltage V_{FB} caused by the fixed surface charge Q_{SS} and the gate–semiconductor work function ψ_{MS} [20.3, 30].

Equation (20.58) does not take into account the impact of interface traps, which cause the C–V curve to stretch. The traps are slow and do not respond to the high frequency of the test signal, but they do follow the changes in the gate bias. Therefore, ∂V_G must be replaced with ∂V_{G0} in (20.58), with ∂V_{G0} representing the case when no interface traps are present.

The value of ∂V_{G0} can be obtained by comparing high- and low-frequency (quasi-static) C–V curves for a MOS structure at the same gate biases V_G. Therefore, the ratio $\partial V_{G0}/\partial V_G$ can be found at any gate bias V_G, since the band-bending is the same for both HF and LF capacitances. In [20.29, Sect. 9.4], it is shown that

$$\frac{\partial V_{G0}}{\partial V_G} = \frac{C_{OX} + C_D}{C_{OX} + C_D + C_{IT}} = \frac{1 - C_{LF}/C_{OX}}{1 - C_{HF}/C_{OX}} \,. \quad (20.59)$$

In this case (20.58) is modified to

$$n(w) = \frac{-2}{q\varepsilon_{Si}\frac{\partial}{\partial V_G}\left(\frac{1}{C_{HF}^2}\right)} \frac{1 - C_{LF}/C_{OX}}{1 - C_{HF}/C_{OX}} \quad (20.60)$$

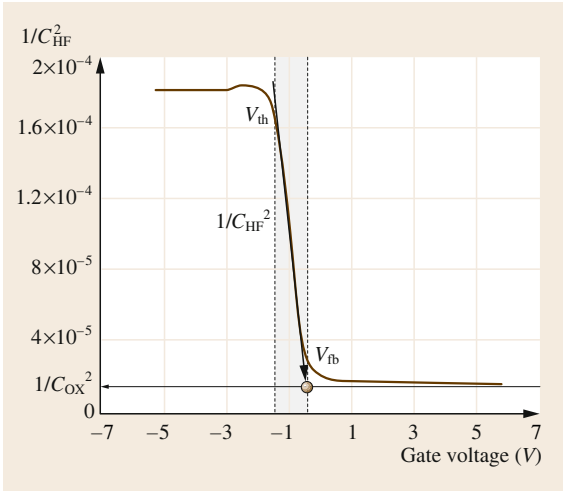


Fig. 20.19 A $1/C_{HF}^2$ versus V_G plot [20.30]. The slope of the *fitted arrow line* is proportional to the average doping, and the *arrow* points to the flat-band voltage V_{FB}, obtained at the V intercept with $1/C_{OX}^2$, shown with the *second horizontal arrow*

as originally proposed in [20.32] and illustrated in Fig. 20.20. As seen from Fig. 20.20, the stretching of the C–V curves due to the interface states induced by stress in Fig. 20.20a causes a disparity in the doping profile in Fig. 20.20b if only the high frequency capacitance is used. The disparity is well-suppressed in Fig. 20.20c by the high–low frequency capacitance measurement, taking into account the stretching of the C–V characteristics. Provided that the depletion layer capacitance is measured at a high frequency, the depletion layer width w is still obtained by (20.57).

Note that the maximum depth w_{max} (20.52) and the resolution Δ_w of the doping profile by means of C–V measurements is limited by the maximum band-bending ψ_{max} and the extrinsic Debye length λ_{Debye}, given by (20.51) and (20.61), respectively, and λ_{Debye} is

$$\lambda_{Debye} = \sqrt{\frac{\varepsilon_{Si}kT}{q^2 n}} \,. \quad (20.61)$$

The doping profile obtained in this way is reliable for depths w of between $3\lambda_{Debye}$ and $w_{max}/2$, when the MOS structure is in depletion and weak inversion, but not in accumulation. That is, $C_{LF} < 0.7C_{OX}$ as a simple rule. As illustrated in Fig. 20.21, the range of w values between $3\lambda_{Debye}$ and equilibrium, obtained via quasi-static C–V measurements, cover about half-a-decade. With proper corrections, the lower distance decreases

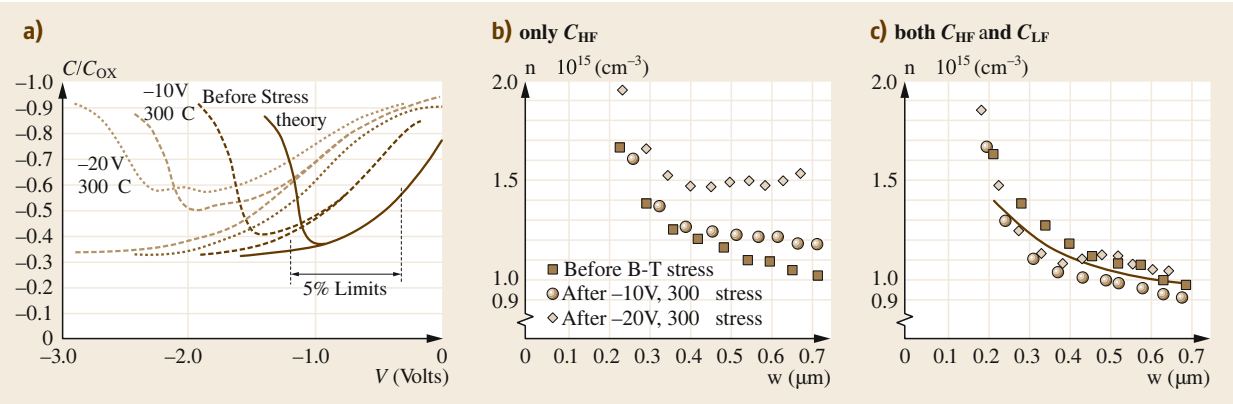


Fig. 20.20a–c *C–V* curves and doping profiles of a MOS structure with 145 nm oxide and uniform doping of 10^{15} cm^{-3} before and after bias temperature stress [20.32]. **(a)** Theoretical and (stretched) measured *C–V* curves. **(b)** Doping profile deduced from only C_{HF} (see (20.58)); **(c)** Doping profile deduced from both C_{HF} and C_{LF} (see (20.60)). (After [20.32])

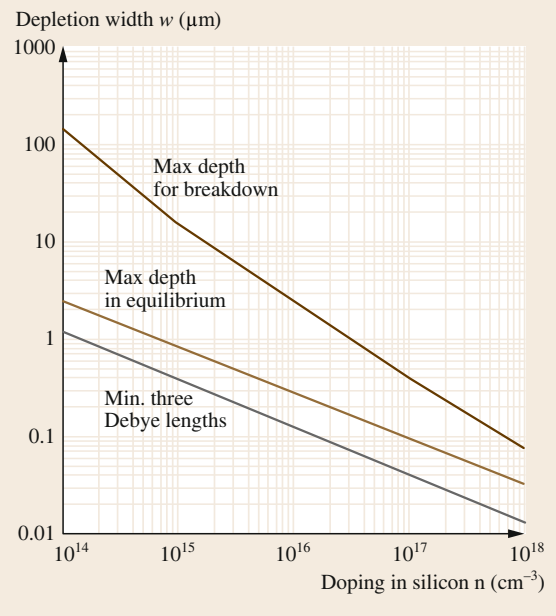


Fig. 20.21 Limitations on the depth achievable when profiling the doping of silicon MOS structures via *C–V* measurements at room temperature. (After [20.4, p. 86])

to one Debye length [20.30]. Using nonequilibrium (transient) *C–V* measurements in deep depletion, the profiling can be extended to higher distances by about an order of magnitude, but further limitations can appear due to the high-frequency response of the interface charge, measurement errors, avalanche breakdown in deep depletion, or charge tunneling in highly doped substrates and thin oxides. More details are presented in [20.29].

20.3.3 Density of Interface States

Interface traps change their charge state depending on whether they are filled or empty. Because interface trap occupancy varies with the slow gate bias, stretching of the *C–V* curves occurs, as illustrated in Fig. 20.20. A quantitative treatment of this *stretch-out* can be obtained from Gauss' law as

$$C_{OX}(V_G - \psi_S) = -Q_S - Q_{IT} = -Q_T \,, \qquad (20.62)$$

where Q_S and Q_{IT} are the surface and interface trap charges (per unit area), which are both dependent on the surface band-bending ψ_S, $Q_T = Q_S + Q_{IT}$ is the total charge in the MOS structure, C_{OX} is the gate capacitance (per unit area), and V_G is the bias applied at the gate of the MOS structure. For simplicity, the (gate metal)-to-(semiconductor bulk) potential ψ_{MS} is omitted in (20.62), but in a real structure the constant ψ_{MS} must be subtracted from V_G. As follows from (20.62), small changes ∂V_G in gate bias cause changes $\partial\psi_S$ in the surface potential bending, and the surface C_S and interface trap C_{IT} capacitances (both per unit area) can represent Q_S and Q_{IT}, given by

$$C_{OX}\partial V_G = (C_{OX} + C_S + C_{IT})\partial\psi_S \,. \qquad (20.63)$$

C_S and C_{IT} are in parallel and in series with the C_{OX}, respectively. Therefore, the measured low-frequency capacitance C_{LF} (per unit area) of the MOS structure becomes

$$C_{LF} = \frac{\partial Q_T}{\partial V_G} = \frac{\partial Q_T}{\partial\psi_S}\frac{\partial\psi_S}{\partial V_G} = \frac{C_{OX}\,(C_S + C_{IT})}{C_{OX} + C_S + C_{IT}} \,. \qquad (20.64)$$

Equation (20.64) shows that stretch-out in the C–V curve can arise due to a non-zero value of C_{IT}, which deviates from the ideal case of $C_{IT} = 0$.

According to [20.29, p. 142], D_{IT} is the density of interface states per unit area (cm^2) and per unit energy (1 eV) in units of $cm^{-2}\,eV^{-1}$. Since the occupancy of the interface states has a Fermi–Dirac distribution, then upon integrating over the silicon band-gap, the relation between C_{IT} and D_{IT} is

$$C_{IT}(\psi_S) = qD_{IT}(\phi_B + \psi_S)\,, \tag{20.65}$$

where $\phi_B = (k_BT/q)\ln(n/n_i)$ is the shift of the Fermi level from the intrinsic level $\phi_i = (E_c - E_v)/2q$ in the silicon bulk of the MOS structure due to the doping concentration n, and n_i is the thermally generated carrier concentration in silicon. Since the derivative of the Fermi–Dirac distribution is a sharply peaking function, then $C_{IT}(\psi_S)$ at particular ψ_S probes $D_{IT}(\phi_B + \psi_S)$ over a narrow energy range of k_BT/q, in which D_{IT} can be assumed to be constant and zero outside this interval. Thus, varying the gate bias V_G, and therefore ψ_S, (20.65) can be used to obtain the density of states D_{IT} at a particular energy shift $q(\phi_B + \psi_S)$ from the silicon intrinsic (mid-gap) energy E_i.

It is evident from (20.64) and (20.65) that the experimental values for D_{IT} can be obtained only when C_{IT}, and ψ_S are determined from C–V measurements. The simplest way to determine ϕ_B is to get the average doping density n using the maximum–minimum high-frequency capacitance method (see (20.53) and Fig. 20.18), or to use the values of n from doping profiles at $0.9\,w_{max}$ – see (20.59) [20.30]. Either the high-frequency or the low-frequency $C-V$ measurement can be used to obtain C_{IT}, but it is necessary to calculate C_S as function of ψ_S, which makes it difficult to process the experimental data.

The most suitable technique for experimentally determining D_{IT} is the combined high–low frequency capacitance method [20.29, Sect. 8.2.4, p. 332]. The interface traps respond to the measurement of low–frequency capacitance C_{LF}, whereas they do not respond to the measurement of the high-frequency measurement C_{HF}. Therefore, C_{IT} can be obtained from measurements by *subtracting* C_{HF} from C_{LF}, given by

$$C_{IT} = \left(\frac{1}{C_{LF}} - \frac{1}{C_{OX}}\right)^{-1} - \left(\frac{1}{C_{HF}} + \frac{1}{C_{OX}}\right)^{-1}. \tag{20.66}$$

Denoting $\Delta C = C_{LF} - C_{HF}$, the substitution of (20.66) into (20.65) provides a direct estimate of D_{IT} from C–V measurements (see also [20.3, p. 371]) as

$$D_{IT} = \frac{\Delta C}{q}\left(1 - \frac{C_{LF}}{C_{OX}}\right)^{-1}\left(1 - \frac{C_{HF}}{C_{OX}}\right)^{-1} \tag{20.67}$$

Note that the combined high–low frequency capacitance method provides C_{IT} and D_{IT} as function of gate bias V_G. However, if D_{IT} needs to be plotted as a function of the position in the energy band-gap, the surface band-bending ψ_S must also be determined as function of gate bias V_G, as follows from (20.65).

There are several ways to obtain the relation between ψ_S and V_G. One way is to create a theoretical plot of C_{HF} versus ψ_S and then, for any choice of C_{HF}, a pair of values for ψ_S and V_G is found [20.29, p. 327]. This method is relatively simple if the doping concentration n in the silicon is uniform and known, because the high-frequency silicon surface capacitance C_S under depletion and accumulation is a simple function of the band-bending ψ_S, and the flat-band capacitance C_{FB} [20.29, pp. 84, 97, 164] is given by

$$C_{FB} = \frac{\varepsilon_{Si}}{\lambda_{Debye}} = \sqrt{\frac{\varepsilon_{Si}q^2n}{k_BT}} \tag{20.68}$$

$$C_S(\psi_S) \approx \begin{cases} \dfrac{C_{FB}}{\sqrt{2}}\dfrac{\exp\left(\frac{q\psi_S}{k_BT}\right)-1}{\sqrt{\exp\left(\frac{q\psi_S}{k_BT}\right)-\frac{q\psi_S}{k_BT}-1}}\,, \\ \qquad \psi_S > 0 \text{ in accumulation} \\ C_{FB},\ \psi_S = 0 \text{ at flat band}, \\ \dfrac{C_{FB}}{\sqrt{2}}\dfrac{1-\exp\left(\frac{q\psi_S}{k_BT}\right)}{\sqrt{\exp\left(\frac{q\psi_S}{k_BT}\right)-\frac{q\psi_S}{k_BT}-1}}\,, \\ \qquad \psi_S < 0 \text{ in depletion}\,. \end{cases} \tag{20.69}$$

Since C_S is in series with C_{OX}, then the theoretical C_{HF} is obtained as a function of the band-bending ψ_S by

$$\frac{1}{C_{HF}(\psi_S)} = \frac{1}{C_{OX}} + \frac{1}{C_S(\psi_S)} \tag{20.70}$$

For a uniformly doped silicon with SiO_2 as the insulator, the ratio $C_{HF}(V_{FB})/C_{OX}$ at gate bias for flat-band conditions is given [20.3, p. 349] by

$$\frac{C_{HF}(V_{FB} \text{ or } \psi_S = 0)}{C_{OX}} = \frac{1}{1 + \frac{136\sqrt{T/300}}{t_{ox}\sqrt{n}}}\,, \tag{20.71}$$

where t_{ox} is the oxide thickness (cm), n is the doping (cm^{-3}), and the T is the temperature (K).

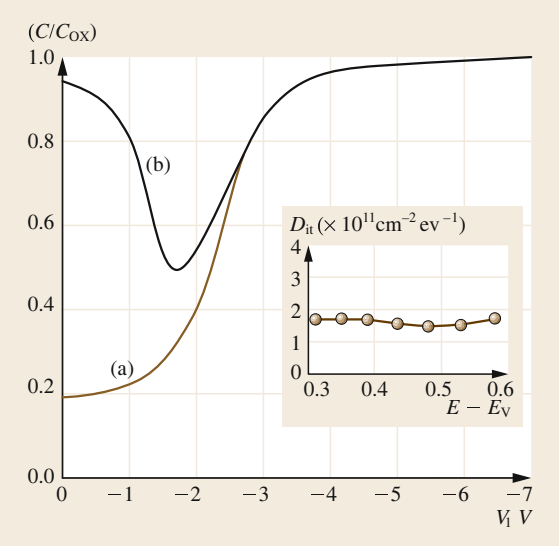


Fig. 20.22 Results from the combined high–low frequency capacitance method [20.33]. *(a)* High-frequency C–V curve; *(b)* low-frequency C–V curve. The energy profile for the density of interface states D_{IT} is shown in the *inset*, as calculated by (20.65), (20.67) and (20.72)

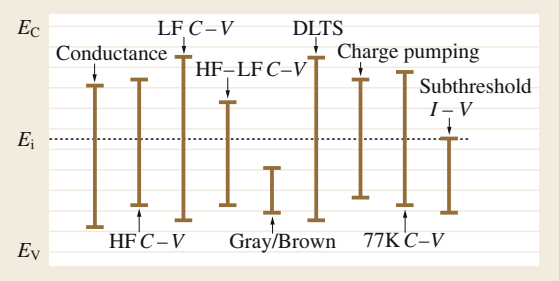


Fig. 20.23 Energy ranges in the silicon band-gap of a p-type substrate over which the density of interface traps can be determined using various measurement methods and characterization techniques. (After [20.3, p. 104])

It was demonstrated in [20.29] that the method of using a theoretical plot to obtain the relation between ψ_S and V_G works well in the case of uniformly doped silicon even if only high-frequency C–V measurements are used to obtain the density of states; that is, the $1/C_{HF}^2$ versus V_G plot is almost a straight line. However, with substrates that are not uniformly doped, the method is inconvenient because the corrections in (20.68) and (20.69) are difficult to implement. Therefore, in practice, a method based on low-frequency C–V measurement is preferred [20.30].

Low-frequency C–V measurement was first used to obtain the relation between ψ_S and V_G [20.34]. This method is based on the integration of (20.66) from an initial gate bias V_{G0}, arbitrarily chosen either under strong accumulation or strong inversion, to the desired V_G at which the band-bending $\psi_S(V_G)$ is to be obtained. Since C_{IT} is part of (20.66), then the low-frequency C–V curve [20.29, p. 93] is integrated as

$$\psi_S(V_G) = \psi_{S0} + \int_{V_{G0}}^{V_G} \left(1 + \frac{C_{LF}(V_G)}{C_{OX}}\right) \mathrm{d}V_G \,. \qquad (20.72)$$

The value of ψ_{S0} is selected such that $\psi_S(V_{FB}) = 0$ when integrating from V_{G0} to the flat-band gate voltage V_{FB}. In this case, V_{FB} is usually obtained beforehand from the point of V-intercept with $1/C_{OX}^2$ when extrapolating the linear part of the $1/C_{HF}^2$ versus V_G curve toward the V_G axis (Fig. 20.19). After determining ψ_{S0}, (20.72) provides the relation between ψ_S and V_G. Thus, the density of states D_{IT} obtained from (20.67) as function of V_G using the combined high–low frequency capacitance method can be plotted against the position of D_{IT} in the silicon band gap, as given by (20.65). High and low frequency C–V measurements can therefore be used to plot the data, as illustrated in the insert of Fig. 20.22.

Overall, many different techniques are used to determine the density of states D_{IT} (Fig. 20.23). For some of these techniques, the ability to sense the energy position of D_{IT} in the band-gap of silicon is summarized in [20.3]. Most of them use C–V measurements, but others are based on I–V measurements taken during the subthreshold operation of MOS transistors, deep-level transient spectroscopy (DLTS), charge pumping (CP) in a three-terminal MOS structure, cryogenic temperature measurements, and so on. Each technique has its strengths and weaknesses, which are discussed in [20.3, 35].

In the methods discussed above, it has been assumed that the gate bias V_G varies slowly with time, 20–50 mV/s, and that the MOS structure is in equilibrium; that is, the minority carriers are generated and the inversion layer is readily formed in the MOS structure when V_G is above the threshold. However, the time constant for minority carrier generation is high in silicon (≈ 0.1 s or more), and it is possible to use nonequilibrium high-frequency C–V measurements to further analyze the properties of the MOS structure. Some applications of these methods are presented later.

20.4 Current–Voltage Measurements

20.4.1 *I–V* Measurements on a Simple Diode

Current–voltage measurements of mainstream semiconductor devices are perhaps the simplest and most routine measurements performed, and they can provide valuable information about the quality of materials used. For example, if we consider the *I–V* characteristics of a p-n diode structure, the source–substrate or drain–substrate junctions can provide useful information on the quality of the junction, such as whether defects are present (they give rise to generation–recombination currents or large parasitic resistances for the contacts at the source, drain or substrate terminals).

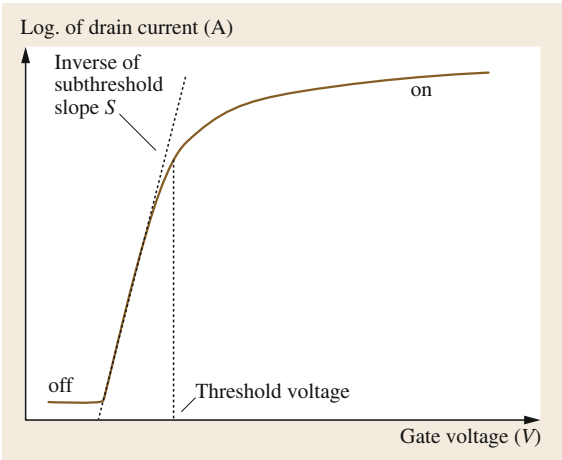


Fig. 20.24 Typical subthreshold characteristics of a MOSFET. The interface state density can be extracted from *S*

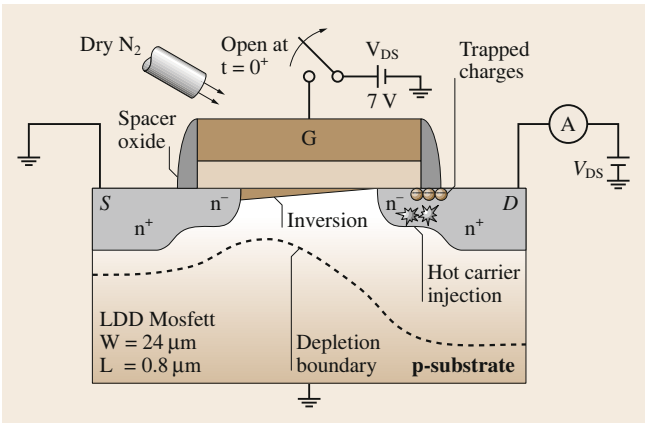


Fig. 20.25 Schematic diagram showing the set-up used for floating gate measurements. The area where charge is trapped after hot electrons are applied is shown

This is easily seen from the current–voltage relation given by the sum of the diffusion (I_{DIFF}) and recombination (I_{GR}) currents

$$I_D = I_{DIFF} + I_{GR} = I_{D0}\left[\exp\left(\frac{eV_D}{nk_BT}\right) - 1\right] + I_{GR0}\exp\left(\frac{eV_D}{2k_BT}\right) \tag{20.73}$$

where I_{D0} and I_{GRO} are the zero-bias diffusion and recombination currents respectively, n is an ideality factor (typically 1), and V_D is the voltage across the intrinsic diode, which is given by

$$V_D = V_{applied} - I_D R_{parasitic}\,. \tag{20.74}$$

For (20.73), a plot of $\ln(I_D)$ versus V_D allows us to separate out the diffusion and the recombination current components. From (20.73) and (20.74), we can also use the diode's *I–V* characteristics to determine the parasitic resistance in series with the intrinsic diode, as described in detail in [20.36]. In most cases, this $R_{parasitic}$ is dominated by the contact resistance.

20.4.2 *I–V* Measurements on a Simple MOSFET

Simple current–voltage measurements – drain current versus gate voltage (I_{DS}–V_{GS}), and I_{DS} versus drain voltage (V_{DS}) – are routinely taken on MOSFETs in order to study their electrical characteristics; however, these can also be used to obtain useful information on the quality of the semiconductor, contacts, oxide and semiconductor–oxide interface. For example, the $I_{DS} - V_{GS}$ characteristics at very small V_{DS} biases (linear region of operation) for a set of test transistors of fixed channel width and different channel lengths is often used to extract parameters such as the threshold voltage (V_T), the transconductance (g_m), the intrinsic mobility (μ_o) and the mobility degradation coefficients θ_0 and η, the parasitic source (R_S) and the drain resistances (R_D) in series with the intrinsic channel, the channel length reduction ΔL, the output conductance (g_{DS}) and the subthreshold slope (S) [20.37].

These parameters are required for modeling and they directly impact the device's performance. However, some of these parameters can also be used to assess the quality of the silicon–silicon dioxide (Si–SiO_2) interface [20.38, 40]. For example, interface states at the Si–SiO_2 interface can change the threshold voltage,

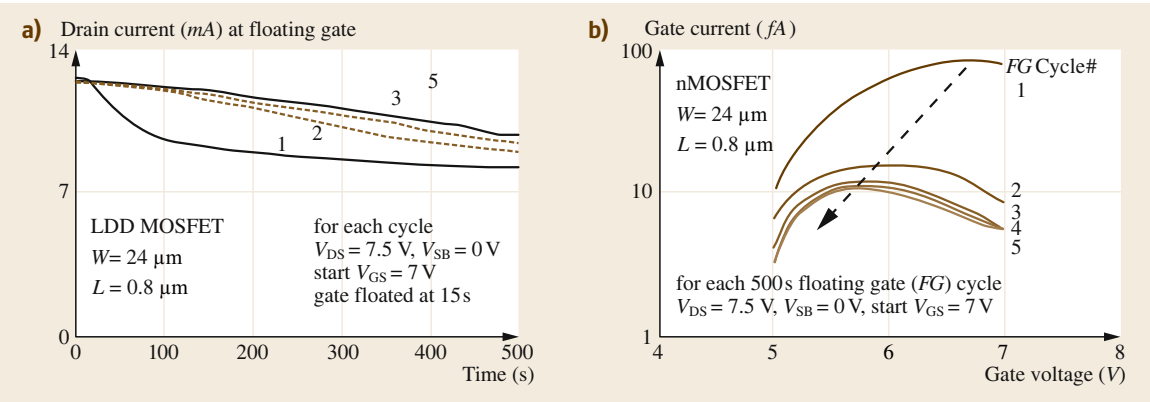


Fig. 20.26 **(a)** Evolution of the drain current over five floating gate cycles. Biasing conditions were chosen to maximize hot-electron gate currents. Note that the maximum drop in drain current occurs after the first floating gate cycle. **(b)** Extracted gate currents using (20.78) and the measurements in **(a)**. As with the drain current, the maximum change in gate current occurs between floating gate cycles 1 and 2. The shift in the peak of the gate current is explained in [20.38, 39]

the subthreshold slope and the mobility, all of which will impact on the drain–source (I_{DS}) current flowing through the device. Here, we look at one parameter in more detail – the subthreshold slope S in mV (of V_{GS})/decade (of I_{DS}).

First, the interface trap density (D_{IT}) can be determined from a semi-log plot of $I_{DS} - V_{GS}$ characteristics at very low drain biases, as shown in Fig. 20.24. We start with the expression for the subthreshold slope

$$S = \frac{k_B T}{q} \ln(10) \left(1 + \frac{C_D + C_{IT}}{C_{OX}}\right) \tag{20.75}$$

in which

$$C_D = \frac{q\varepsilon_{Si} N_A}{\sqrt{2\psi + |V_B| - \frac{k_B T}{q}}} \tag{20.76}$$

D_{IT} can then be calculated from

$$C_{IT} = qD_{IT} \tag{20.77}$$

once C_{IT} is determined from (20.75). In fact, a recent comparison in [20.41] of the interface trap densities extracted from capacitance, subthreshold and charge pumping measurements produced similar results, demonstrating that simple and fast I–V measurements based on the subthreshold technique can provide useful information on the Si–SiO_2 quality.

20.4.3 Floating Gate Measurements

The floating gate technique is another simple I–V measurement in which the evolution of the drain current I_{DS} is monitored after the gate bias has been removed. It has been shown to be particularly useful when monitoring early-mode hot-carrier activity in MOS transistors [20.39, 41]. In this measurement, we first check the oxide quality by biasing the transistor in the strong linear region (very low V_{DS} and a V_{GS} well above V_T), and then lift the gate voltage probe so that $V_{GS} = 0$ V and measure the evolution of I_{DS} with time. For a high-quality gate and spacer oxide, I_{DS} remains constant for a long time, indicating that there is negligible carrier injection across the gate oxide through Fowler–Nordheim tunneling or other leakage mechanisms.

A second precaution is to have a dry or inert gas such as nitrogen flowing over the chip to reduce the possibility of other leakage mechanisms such as that from water vapor. The measurement set-up for this experiment is shown in Fig. 20.25. The evolution of the floating gate current over several cycles with the gate voltage applied and then removed is shown in Fig. 20.26. The biasing voltages and time at which the gate is floated are also given in Fig. 20.26a. For this experiment, a biasing condition of $V_{GS} \approx V_{DS}$ was chosen for a high-impact ionization-induced gate current, but a lower-than-maximum electron injection situation was used for the initial biasing condition.

From the evolution of I_{DS} and the $I_{DS} - V_{GS}$ characteristics of a virgin (not intentionally stressed) transistor at the same V_{DS} as the floating gate measurements, and from measurements of the total capacitance associated with the gate (C_G), the gate current (I_G) evolution after each floating gate cycle can be determined using

$$I(V_{GS}) = C_G \frac{dV_{GS}}{dt} . \tag{20.78}$$

This $I_G - V_{GS}$ evolution is shown in Fig. 20.26b. An ancillary benefit of the floating gate technique

is that very small gate currents (in the fA range or even smaller) can be easily determined by measuring much larger drain currents using, for example, a semiconductor parameter analyzer. The reason for this is that the gate current is not directly measured in this technique – it is determined from $I_{DS} - V_{GS}$ and (20.78). Also, the change in the floating gate current after the first few cycles can be used to monitor for early mode failure after statistical evaluation.

20.5 Charge Pumping

Charge pumping (CP) is another electrical technique that is well suited to studying semiconductor–insulator interfaces in MOSFETs [20.42–47]. There are several versions of the CP technique: spatial profiling CP [20.43–47], energy profiling CP [20.48], and, more recently, new CP techniques [20.49] that permit the determination of both interface states (N_{IT}) and oxide traps (N_{OT}) away from the interface and inside the oxide. The charge pumping technique is more complicated than either of the I–V or floating gate methods. However, it is a very powerful technique for assessing interface quality and it works well even with very small transistor geometries and very thin gate oxides, where tunneling can be a problem.

The charge pumping technique was first used in 1969 [20.50] to measure the interface traps at Si−SiO_2 interface. Since then, there have been numerous publications with enhancements, refinements and applications of the technique to a variety of semiconductor–insulator interfacial studies. In the basic charge pumping experiment, the gate of an NMOST (for example) is pulsed from a low value (V_L) when the device is in accumulation to a high value (V_H) when the device is in inversion. This results in the filling of traps between $E_{F,ACC}$ (corresponding to V_L) and $E_{F,INV}$ (corresponding to V_H) with holes and electrons, respectively. When pulsing the gate between accumulation at V_L and inversion at V_H, a current flows due to the repetitive recombination at the interface traps of minority carriers from the source and drain junctions with majority carriers from the substrate. This current is termed the charge pumping current, and it was found to be proportional to the frequency of the gate pulse, the gate area and the interface state density. Its sensitivity is better than $10^9\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$. In the traditional CP experiment shown in Fig. 20.27, but with $\Delta V_S = |V_D - V_S| = 0\,\mathrm{V}$, the gate G is connected to a pulse generator, a reverse bias V_R or no bias is applied to both sources S and drain D terminals, and the charge pumping current flowing in the substrate terminal, I_{CP}, is measured. To generate a typical charge pumping curve (as shown in the top part of Fig. 20.28), the base level of the pulse is varied, taking the transistor from below flat-band to above surface inversion conditions, as shown in the bottom part of Fig. 20.28.

In the traditional CP experiment, the charge pumping current I_{CP} is given by

$$I_{CP} = qA_{GATE}\overline{D_{IT}}\Delta E\,, \tag{20.79}$$

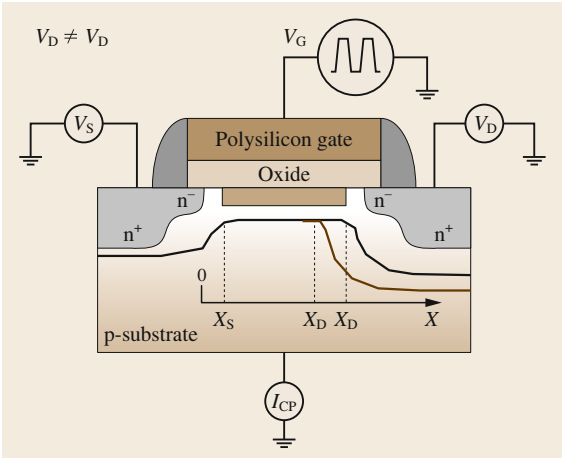


Fig. 20.27 Example of spatial profiling charge-pumping set-up used when the source and drain biases are slightly different

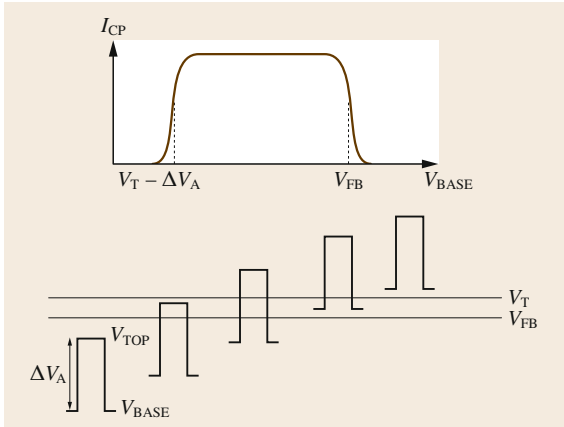


Fig. 20.28 Demonstration of how the CP curve is generated by varying the base level of the pulse so that the entire pulse is between V_{FB} and V_T

where

$$\Delta E = (E_{\mathrm{F,INV}} - E_{\mathrm{F,ACC}})\,. \tag{20.80}$$

This expression assumes that the electrical and physical channel lengths are the same. However, for short channel devices, this assumption results in an error. Therefore, a more accurate expression is

$$I_{\mathrm{CP}} = qf\,W \int_{x_s}^{x_d} N_{\mathrm{IT}}(x)(q\Delta\psi_{\mathrm{SO}}) \tag{20.81}$$

$$x_{\mathrm{d}} = L - \sqrt{\frac{2\varepsilon_{\mathrm{Si}}}{qN}}\left(\sqrt{V_{\mathrm{D}} + \psi_{\mathrm{S}}} - \sqrt{\psi_{\mathrm{S}}}\right). \tag{20.82}$$

The interface state density at the edge of the drain depletion region ($N_{\mathrm{IT}}(x_{\mathrm{d}})$) is given by

$$N_{\mathrm{IT}}(x_{\mathrm{d}}) = \frac{1}{qf\,Wq\Delta\psi_{\mathrm{S0}}}\left(\frac{\mathrm{d}x_{\mathrm{s}}}{\mathrm{d}V_{\mathrm{S}}}\right)^{-1}\left(\frac{\mathrm{d}I_{\mathrm{CP}}}{\mathrm{d}V_{\mathrm{S}}}\right)_{V_{\mathrm{D}}=\mathrm{constant}} \tag{20.83}$$

When performing spatial profiling CP experiments, some precautions are required. The first is that a voltage difference ΔV_{S} between V_{D} and V_{S} that is too small results in a difference in I_{CP} that is too small as well, and hence a large error in $N_{\mathrm{IT}}(x)$, as indicated from (20.83). On the other hand, values of ΔV_{S} that are too large result in a I_{D} that is too high and hence more substrate current I_{B}. This current I_{B} can interfere with I_{CP} if ΔV_{S} is large or if L is very short, resulting in a large error in $N_{\mathrm{IT}}(x)$. The range 50–100 mV for ΔV_{S} seems to be a good compromise for the devices investigated in [20.43, 44]. Experimental results for spatial profiling CP measurements indicate that $N_{\mathrm{IT}}(x)$ peaks near S/D edges. However, after normal mode stress, $N_{\mathrm{IT}}(x)$ only increases near D. This is shown in Fig. 20.29. More details about charge pumping can be found in a recent review [20.51].

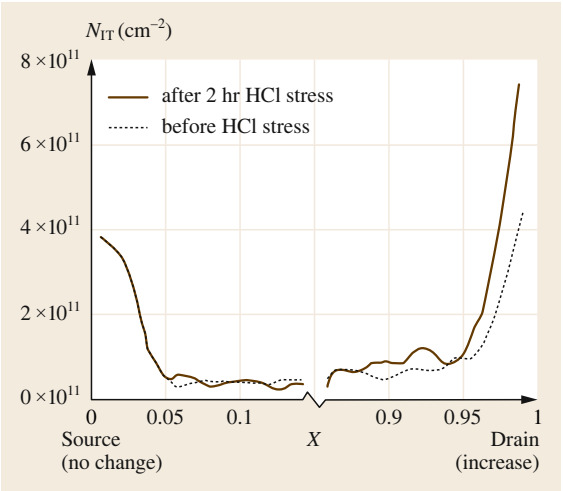


Fig. 20.29 Spatial interface state distribution over the channel in a 1 μm-long device. The stress was applied for 2 h at $V_{\mathrm{DS}} = 5\,\mathrm{V}$ and $V_{\mathrm{GS}} = 2.4\,\mathrm{V}$

20.6 Low-Frequency Noise

20.6.1 Introduction

Low-frequency noise (LFN) spectroscopy requires very good experimental skills in the use of low-noise instrumentation as well as grounding and shielding techniques. Other special considerations are also required, which are discussed later. Although it is time-consuming to perform, it has been widely used to probe microscopic electrical transport in semiconductors and metals. LFN is very sensitive to defects in materials and devices, and large differences in LFN characteristics can be observed in devices with identical electrical current–voltage characteristics. This is mainly because electrical I–V measurements only probe the average or macroscopic transport in devices and so are not as sensitive to defects as LFN. Due to its sensitivity to defects, traps or generation–recombination centers, LFN has been proposed as a good tool for predicting device reliability. For example, LFN has been used to predict the reliabilities of metal films [20.52], and has been used in processing steps that produce photodetectors with better performance [20.53, 54]. LFN noise is sensitive to both bulk and surface defects or contaminants of a material.

Using low-frequency noise spectroscopy and biasing the transistor in saturation, we can spatially profile the defect density near the drain and source terminals for devices in normal and reverse modes of operation [20.55]. Low-frequency noise in the linear region also allows us to extract the average defect density over the entire channel region at the silicon–silicon dioxide interface [20.56, 57]. Noise experiments were performed on small-geometry polysilicon emitter bipolar transistors to investigate the number of interface states in the thin interfacial oxide layer between the monocrystalline and polycrystalline silicon [20.58–

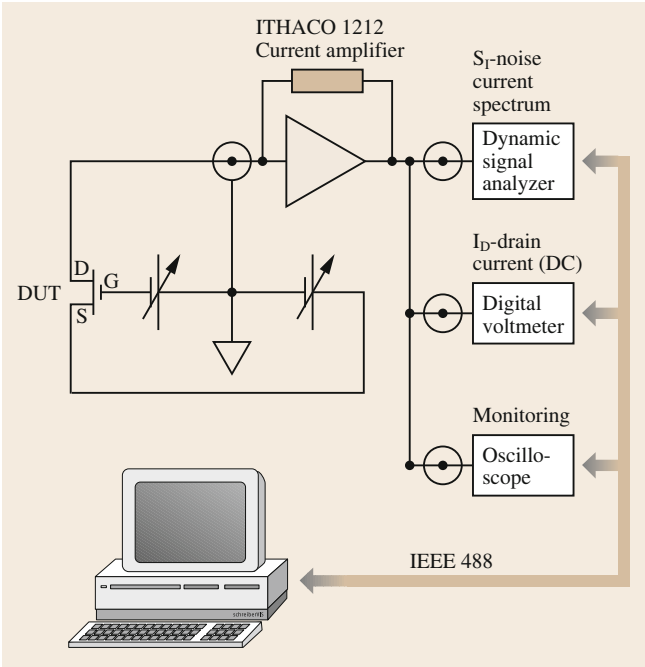


Fig. 20.30 System for measuring low-frequency noise. In this diagram, the device under test is any field-effect or thin-film transistor

68]. Recent experiments using body or substrate bias (V_B) in a MOS transistor allowed us to look at the contribution of bulk defects (defects away from the silicon–silicon dioxide interface) and their contribution to device noise [20.69–71]. This is important since substrate biasing has been proposed as a means to cleverly manage power dissipation and speed in emerging circuits and systems [20.72].

We will discuss how low-frequency noise (LFN) spectroscopy can be applied to the interfacial oxide layer between the mono-silicon and polysilicon emitter in bipolar junction transistors (BJTs) here. The experimental system shown in Fig. 20.30 is used for LFN measurements of field-effect transistors (FETs); the same system can also be used for BJTs.

As mentioned before, special attention must be paid to grounding and shielding in LFN measurements, as this is crucial to minimizing the effects of experimental and environmental noise sources on the device under test (DUT). Because electric power supplies are noisy, especially at 60 Hz (in North America) and its harmonics, and this noise can dominate the noise of the DUT, batteries are often used to supply the voltage. Metal film resistors are the preferred means of changing the biasing conditions, because of their better low-noise characteristics compared to carbon resistors, for example.

With these experimental precautions taken, the noise signal from the transistor might still be too low to be directly measured using a spectrum or signal analyzer. Therefore, a low-noise voltage or current amplifier, whose input noise sources are lower than that of the noise signal, is used to boost the noise signal. In addition, other instruments might be used to measure currents or voltages, or to display the waveforms (as shown in Fig. 20.30). An example of a low-frequency noise characterization system that we have used to study the noise in thin film polymer transistors is shown in Fig. 20.30. Note that LFN measurements are time intensive because a large number of averages are required for smooth spectra. Also, in noise measurements, the power spectrum densities S_V and S_I for the noise voltages and currents are measured, in units V^2/Hz and A^2/Hz, respectively.

20.6.2 Noise from the Interfacial Oxide Layer

Here we present some sample results and show how low-frequency noise spectra in ultrasmall devices can

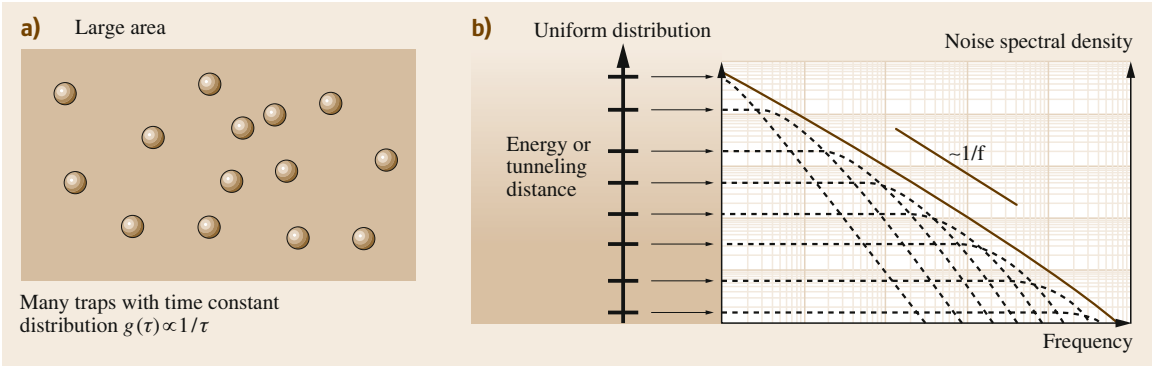


Fig. 20.31 Schematic representation of a *large*-area PE-BJT with many traps distributed uniformly across the band gap and the emitter area, and with a $g(\tau) = 1/\tau$ distribution for the time constant. The resulting spectrum is $1/f$ noise

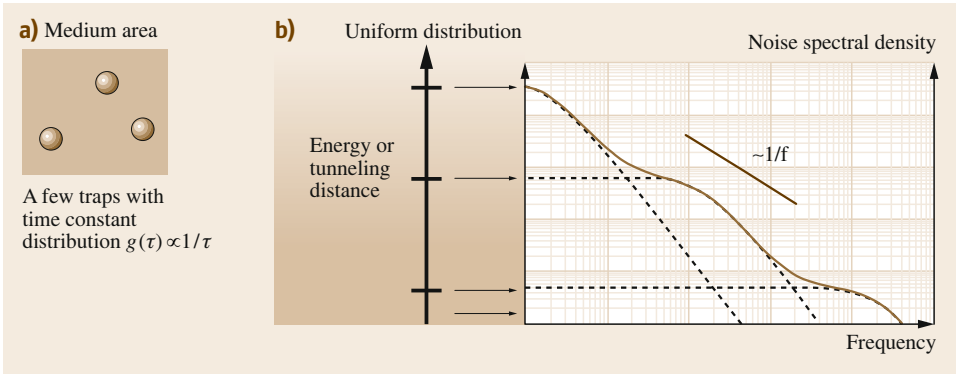


Fig. 20.32 Schematic representation of a *medium*-area PE-BJT ($\approx 0.5\,\mu m^2$) with a few traps. Note that g–r bumps appear for each trap since there are only a few traps

be used to estimate the oxide trap density. Generally, the low-frequency noise spectra of polysilicon emitter (PE) BJTs are made up of $1/f$ noise, generation–recombination (g–r) noise and shot noise sources. In the case of the base current, the noise spectra can be modeled as

$$S_{I_B} = \frac{K_F I_B^{A_F}}{f} + \sum_{i=1}^{n} \frac{B_i \tau_i}{1 + (2\pi f \tau_i)^2} + 2qI_B \,, \qquad (20.84)$$

where the symbols have their usual meanings, see [20.68] for example.

As described in [20.63–68], the LFN in PE BJTs originates from the thin layer of oxide between the monocrystalline and polycrystalline silicon emitter. The defects at this interface may be dangling oxygen bonds, oxygen vacancies, interface states or oxide traps [20.63]. Devices with large emitter areas have many traps, and these produce generation–recombination noise which produce $1/f$ noise when added. This is schematically shown in Fig. 20.31.

As the device area is reduced, and assuming a constant trap density (which is normally true for devices on the same wafer), then there are fewer traps in the interfacial oxide layer for smaller area devices. In this case, the spectral density of the noise changes and it gains characteristic *bumps* associated with resolvable g–r noise components. This is schematically shown in Fig. 20.32.

In very small devices with only a single trap, for example, the noise spectrum changes dramatically; only g–r noise is observed in the frequency domain along with a random telegraph signal (RTS) in the time domain. This is schematically shown in Fig. 20.33. Real experimental results are shown for three sizes of transistors (2.4, 0.64 and $0.16\,\mu m^2$) in Fig. 20.34. Here, one can see how $1/f$ noise is made up of g–r spectra as the emitter geometries are scaled to smaller and smaller values. For the PE BJT with an emitter area of $0.16\,\mu m^2$, a lower bound of $\approx 10^9\,cm^2$ can be approximated for the oxide trap density. Similar results have been obtained for MOSFETs [20.73].

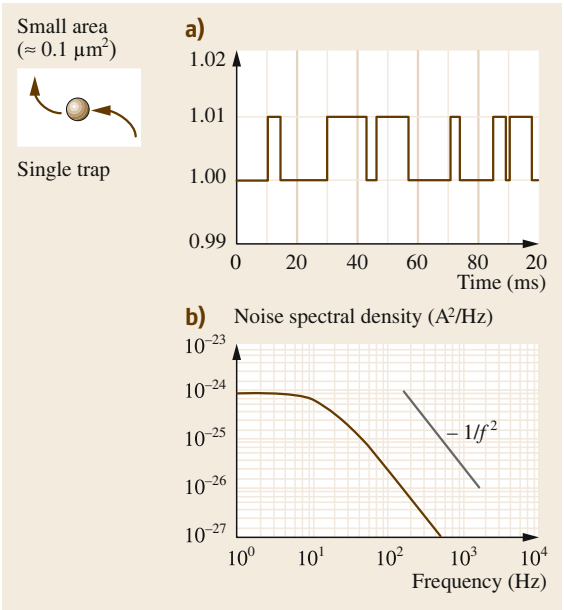


Fig. 20.33 Schematic representation of a *small*-area PE-BJT ($\approx 0.1\,\mu m^2$) with one trap. Note that **(a)** a single RTS and **(b)** g–r spectrum appear because there is only one trap

20.6.3 Impedance Considerations During Noise Measurement

Two basic circuits can be employed when measuring the low-frequency noise (LFN) in a device. These configurations are sketched in Fig. 20.35. In voltage noise measurement (Fig. 20.35a), a low-noise preamplifier senses the voltage across the device, and this signal is sent to a spectrum analyzer or a fast Fourier transform (FFT) analyzer. In current noise measurements (Fig. 20.35b), the low-noise preamplifier senses the current through the device, converts it into a voltage,

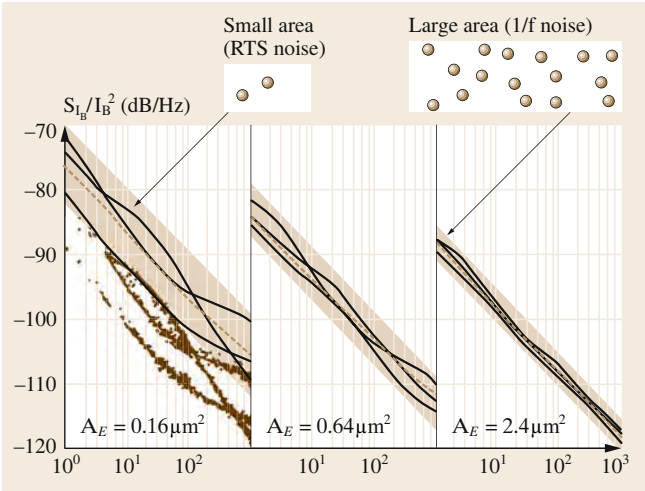


Fig. 20.34 Experimental results for low-frequency noise spectra from sets of large-, medium- and small-area PE-BJTs. In all cases, the average spectrum is $1/f$ noise and the relative magnitude of the $1/f$ noise is the same; that is, the area of $K_F\times$ is the same for the three sets of transistors. (After [20.59–62])

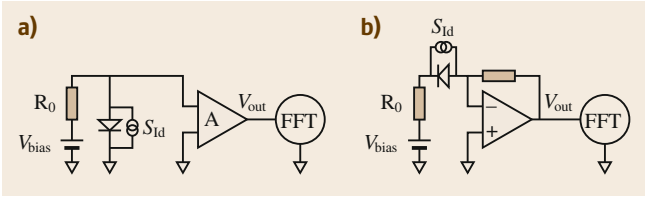


Fig. 20.35a,b Basic circuits used to measure the low-frequency noise (LFN) in a device (DUT). **(a)** Voltage noise measurement; **(b)** current noise measurement

and forwards the voltage to a FFT or spectrum analyzer.

In principle, both configurations can be used for LFN measurement, but the impact of the nonideality of the amplifier (such as the input impedance, noise voltage and current) changes when the device impedance changes. Also, the noise from the bias source varies with each measurement set-up.

The noise equivalent circuit used for voltage measurement is shown in Fig. 20.36. The noise from the amplifier is represented by the input-referred noise voltage (S_{Vn}) and noise current (S_{In}) sources. The noise voltage from the bias is represented by S_{V0}. The impedance of the bias source is R_0, whereas the input impedance of the amplifier is neglected, since it is usually very high compared to R_0. The impedance of DUT is r_d. The noise current S_{Id} of the device that can be measured, assuming that the noise voltage at the input of the amplifier $S_{Vm} = S_{Id}r_d^2$. However, the amplifier sees

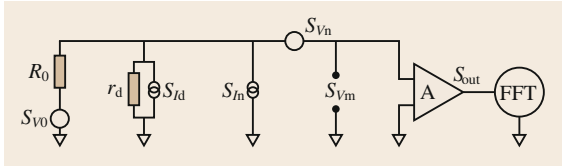


Fig. 20.36 Noise equivalent circuit for voltage noise measurements

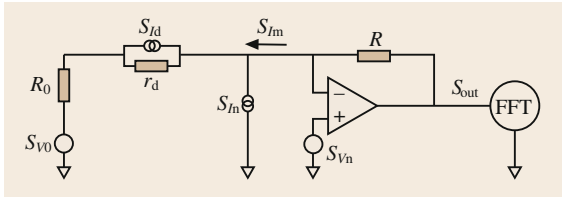


Fig. 20.37 Noise equivalent circuit of the current noise measurement

a different level of S_{Vm}, given by

$$S_{Vm} = \frac{S_{OUT}}{A^2} = S_{V0}\left(\frac{r_d}{r_d + R_0}\right)^2 + S_{Vn} + (S_{II} + S_{II})Z^2 \quad (20.85)$$

where

$$Z = (r_d//R_0) = \frac{r_d R_0}{r_d + R_0} \quad (20.86)$$

and A is the voltage gain of the amplifier. Therefore, the estimated value for S_{Id} is

$$S_{Id} = \frac{S_{Vm} - S_{Vn}}{Z^2} - \frac{S_{V0}}{R_0^2} - S_{In}\,. \quad (20.87)$$

The uncertainty in (20.87) is

$$\frac{\Delta S_{Id}}{S_{Id}} = \frac{\Delta S_{Vn}}{S_{Vm}} + \frac{\Delta S_{V0}}{S_{Vm}}\left(\frac{Z}{R_0}\right)^2 + \frac{\Delta S_{In}}{S_{Vm}}Z^2 \quad (20.88)$$

where $\Delta S \leq S$ denotes the uncertainty in each noise source. As seen from (20.88), the impact of the bias source noise ΔS_{V0} and the input current noise ΔS_{In} can be reduced if the impedance of the measurement circuit Z is low and the ratio r_d/R_0 is kept much less than 1; in other words, the *voltage noise measurement is more appropriate for low-impedance devices, such as diodes at forward biasing*, and the noise floor of the measurement is limited by the input-referred voltage noise S_{V_n} of the amplifier.

For the other (dual) case, current noise measurement, the noise equivalent circuit is shown in Fig. 20.37. The corresponding equations for the measured noise

current $S_{I\mathrm{m}}$, Z, the device noise $S_{I\mathrm{d}}$, and the uncertainty, respectively, are given by (20.89–20.92) below

$$S_{I\mathrm{m}} = \frac{S_{\mathrm{OUT}}}{R^2} = \frac{S_{V0}}{(r_\mathrm{d}+R_0)^2} S_{I\mathrm{n}} + \frac{S_{V\mathrm{n}}}{Z^2} + S_{I\mathrm{d}} \left(\frac{r_\mathrm{d}}{r_\mathrm{d}+R_0}\right)^2 \tag{20.89}$$

$$Z = (r_\mathrm{d}+R_0)\,//R = \left(\frac{1}{R} + \frac{1}{r_\mathrm{d}+R_0}\right)^{-1} \tag{20.90}$$

$$S_{I\mathrm{d}} = (S_{I\mathrm{m}} - S_{I\mathrm{n}}) \left(1+\frac{R_0}{r_\mathrm{d}}\right)^2 - \frac{S_{V0}}{r_\mathrm{d}^2} - \frac{S_{V\mathrm{n}}}{Z^2}\left(1+\frac{R_0}{r_\mathrm{d}}\right)^2 \tag{20.91}$$

$$\frac{\Delta S_{I\mathrm{d}}}{S_{I\mathrm{d}}} = \frac{\Delta S_{I\mathrm{n}}}{S_{I\mathrm{m}}} + \frac{\Delta S_{V0}}{S_{I\mathrm{m}}} \frac{1}{(r_\mathrm{d}+R_0)^2} + \frac{\Delta S_{V\mathrm{n}}}{Z^2 S_{I\mathrm{m}}} \tag{20.92}$$

As expected from duality, it is apparent from (20.92) that the impact of the bias source noise ΔS_{V0} and the input voltage noise $\Delta S_{I\mathrm{n}}$ can be reduced if the impedance of the measurement circuit Z and $(r_\mathrm{d}+R_0)$ are both high; in other words, the *current noise measurement is more appropriate for high-impedance devices, such as diodes at reverse biasing*, and the noise floor of the measurement is limited by the input-referred current noise S_{I_n} of the amplifier.

This analysis above demonstrates that the choice of the measurement configuration follows our expectation that voltage should be measured in low-impedance devices and current in high-impedance devices. Also, the noise floor limiting parameter of the preamplifier is of the same type as the type of measurement; that is, input-referred noise voltage for voltage noise measurement and input-referred noise current for current noise measurement. Note that there is a trade-off between the voltage and current noise in amplifiers, which implies that the measurement configuration – either voltage or current measurement – should also be carefully selected with respect to the impedance of the device under test. In addition, four-point connection can be used to measure the noise in very low impedance devices ($r_\mathrm{d} < 100$). These and other considerations for low-frequency noise instrumentation are discussed in many papers, for example [20.74–76].

20.7 Deep-Level Transient Spectroscopy

Deep-level transient spectroscopy (DLTS) is a fairly complicated electrical characterization technique where the temperature is varied in large range from cryogenic temperatures ($< 80\,\mathrm{K}$) to well above room temperature ($> 400\,\mathrm{K}$). However, it is a powerful and versatile technique for investigating deep-level defects and it also gives accurate values for the capture cross-sections of defects. There are several DLTS techniques and [20.77, 78] provide recent reviews of the subject. In DLTS, the semiconductor device or junction is pulsed with an appropriate signal, and the resulting transient (such as capacitance, voltage or current) is monitored at different temperatures. Using these recorded transients at different temperatures, it is possible to generate a spectrum with peaks, each of which is associated with a deep level. The heights of the peaks are proportional to the defect density.

Here, we will focus on a new version of DLTS: the constant resistance (CR) DLTS technique [20.79–81]. We were able to accurately investigate bulk defects in a variety of test structures with CR-DLTS. Using body bias in a MOS transistor, we were able to distinguish interfacial and bulk defects that are important for different applications. For example, interfacial defects are important for electronic applications, and bulk defects are important for imaging or radiation detection applications. Examples of results from DLTS studies with and without body bias will be discussed.

CR-DLTS is well-suited to investigations of electrically active point defects that are responsible for the creation of deep levels in the semiconductor band-gap. CR-DLTS can also be used to distinguish bulk traps and interface traps in MOSFETs.

A conventional DLTS system is shown schematically in Fig. 20.38. In DLTS, an excitation pulse is applied to the sample to fill all of the traps and then the pulsing is stopped. The next step is to detect the transient signal from the sample due to charge emission from the traps. The right side of Fig. 20.38 shows capacitance transients at eight different temperatures. By selecting a time window from t_1 to t_2, and then plotting $[C(t_1) - C(t_2)]$ as a function of temperature, a DLTS spectrum with a characteristic peak is obtained as shown in the bottom of Fig. 20.38.

This peak is a signature from a specific defect level. To determine the properties of the defect (its energy level and capture cross-section), the time window ($\tau = t_2 - t_1$) is changed. In this case, different DLTS spectra are obtained at different temperatures. Using the time difference τ and the temperatures at which the peaks occur,

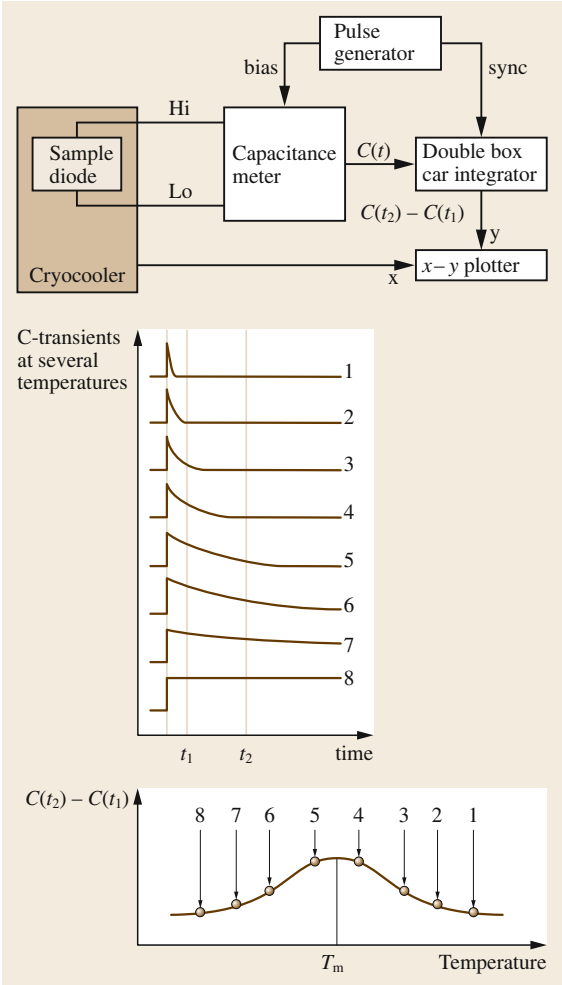


Fig. 20.38 Schematic representation of a conventional DLTS system. The time scans from which the DLTS temperature spectrum is obtained are shown on the *right*

Arrhenius plots are constructed in order to determine the defect energy level and its capture cross-section. Examples of DLTS spectra and Arrhenius plots associateed with CR-DLTS are presented later (in Fig. 20.40).

A block representation of the CR-DLTS system is shown in Fig. 20.39. More details can be found in [20.79–83]. A discussion of the signal processing and averaging techniques used with this DLTS technique can be found in [20.82]. Here, the gate bias voltage of the field-effect transistor is adjusted using a feedback circuit so that the resistance corresponding to the source–drain conductance matches that of a reference resistor R_{ref}, which is typically around 1 MΩ. The voltage transient due to the change in occupancy of the traps appears as a compensation voltage on the gate. This voltage change can be regarded as a threshold voltage change because the flat-band voltage of the device changes when the occupancy of the traps change. More details on how this change in the threshold is related to the traps can be found in [20.77, 79, 81].

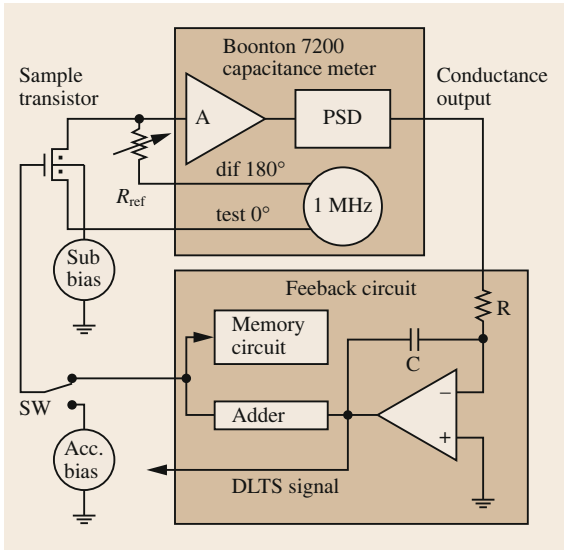


Fig. 20.39 Block diagram representation of the CR-DLTS system

Some important advantages of the CR DLTS technique are that the surface mobility of the MOS transistors does not need to be high, and that it is theoretically independent of the gate area of the transistor. This is expected, since the small amount of charge trapped beneath the gate must be balanced by a correction voltage applied across a relatively small gate–substrate capacitance.

Figure 20.40a shows six DLTS spectra for a junction field-effect transistor (JFET) damaged with 2.7×10^9 protons/cm^2 [20.79, 81] with six selected rate windows. Using the temperatures at which the peaks occur and the rate windows, Arrhenius plots can be constructed as shown in Fig. 20.40b, where the energies of five electron trap levels below the conduction band are also indicated. For the five traps, the extracted capture cross sections were 4.6×10^{-15} cm^2 (E1), 6.3×10^{-15} cm^2 (E2), 1.2×10^{-16} cm^2 (E3), 8.5×10^{-16} cm^2 (E4) and 3.4×10^{-15} cm^2 (E5).

Figure 20.41 shows CR-DLTS spectra as the source–body bias voltage is varied. The scans with a body bias of −1 V are lower in magnitude than those without substrate bias, except for the peak associated with the hole trap at 0.13 eV above the valence band [20.81]. When the reverse substrate bias is increased, the gate control of the space charge region near the channel decreases, meaning that fewer interface traps participate in the capture and emission of

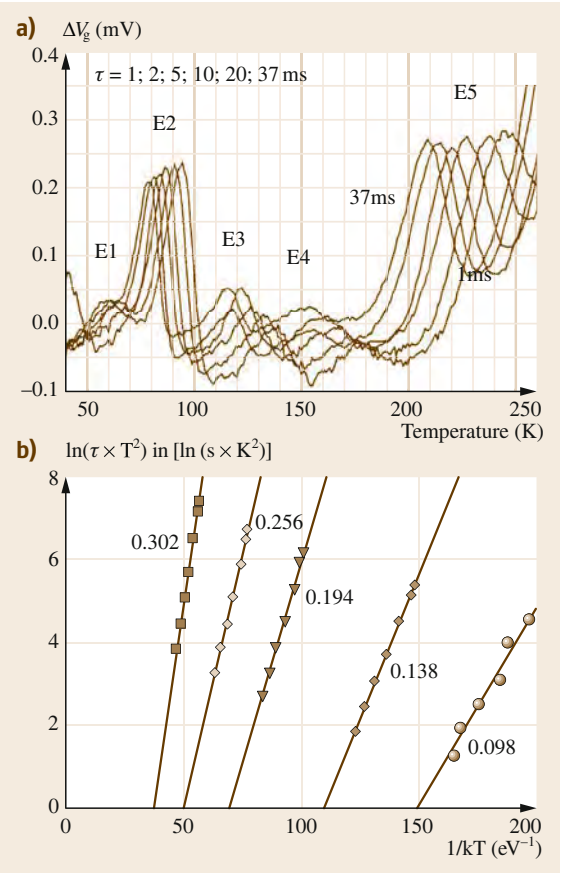


Fig. 20.40 **(a)** CR-DLTS spectra of a 50 µm × 20 µm MOSFET damaged with 2.7×10^9 proton/cm^2. **(b)** Arrhenius plot derived from the CR-DLTS spectra, showing the energies of the five traps E1–E5. (After [20.80, 83])

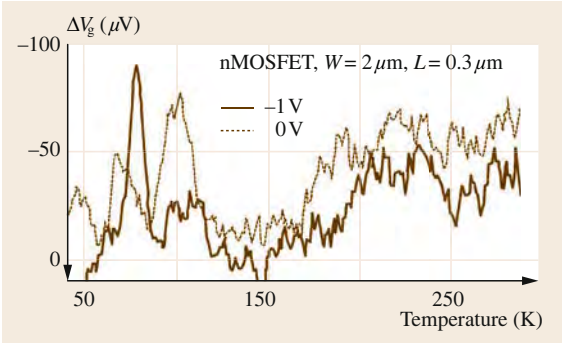


Fig. 20.41 Effect of body bias on CR-DLTS spectra. The body bias affects the surface and bulk traps in different ways

charges. However, the increased reverse substrate bias results in an increased space charge region in the silicon beneath the gate, so more bulk deep levels can participate in the capture and emission processes. This explains the increased deep-level peak (below 75 K) when −1 V is applied to the body. These differences between the CR-DLTS spectra demonstrate the ability to distinguish bulk traps from surface traps when the substrate bias of the MOSFET is varied.

Acknowledgments. The authors are very grateful to Drs. O. Marinov and D. Landheer for their careful review of the manuscript and their assistance. They are also grateful to several previous students and researchers whose collaborative research is discussed here. Finally, they are grateful to NSERC of Canada, the Canada Research Chair program and the CNRS of France for supporting this research.

References

20.1 M.J. Deen: Proc. 6th Symp. Silicon Nitride Silicon Dioxide Thin Insulating Films, ed. by R.E. Sah, M.J. Deen, D. Landheer, K.B. Sundaram, W.D. Brown, D. Misra (Electrochem. Soc. Paris, Paris 2003) p. 3

20.2 P. Rai-Choudhury, J. Benton, D. Schroder (Eds.): *Proceedings of the Symposium on Diagnostic Techniques for Semiconductor Materials and Devices*, Vol. 97-12 (The Electrochemical Society Press, New Jersey 1997)

20.3 D. Schroder: *Semiconductor Material and Device Characterization*, 2nd edn. (Wiley, New York 1998)

20.4 W. Runyan, T. Shaffner: *Semiconductor Measurements and Instrumentation*, 2nd edn. (McGraw Hill, New York 1997)

20.5 ITRS: International Technology Roadmap for Semiconductors 2001 version, http://public.itrs.net/Files/2001ITRS/Home.htm (2001)

20.6 R. Pierret, G.W. Neudeck: *Advanced Semiconductor Fundamentals*, Vol. 6 (Addison-Wesley, Reading 1987)

20.7 W. Sawyer: Proc. 1998 Int. Conf. Ion Implant. Technol., (IEEE Press, Piscataway 1998)

20.8 L.J. van der Pauw: Philos. Res. Rev. **13**, 1 (1958)

20.9 M. Newsam, A. Walton, M. Fallon: Proc. 1996 Int. Conf. Microelectron. Test Struct. (1996) p. 247

20.10 T. Noda, D. Lee, H. Shim, M. Sakuraba, T. Matsuura, J. Murota: Thin Solid Films **380**, 57 (2000)

20.11 P. De Wolf, R. Stephenson, S. Biesemans, P. Jansen, G. Badenes, K. De Meyer, W. Vandervorst: Proc. Int. Electron Devices Meet. (IEDM) Tech. Dig. (1998) p. 559

20.12 J.H. Orchard-Webb, R. Coultier: Proc. IEEE Int. Conf. Microelectron. Test Struct., (1989) p. 169

20.13 J.D. Plummer, P.B. Griffin: Proc. IEEE **89**(3), 240 (2001)

20.14 S. Sze: *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York 1981)
20.15 W.M. Loh, S.E. Swirhun, T.A. Schreyer, R.M. Swanson, K.C. Saraswat: IEEE Trans. Electron Devices **34**(3), 512 (1987)
20.16 S. Zhang, M. Östling, H. Norström, T. Arnborg: IEEE Trans. Electron Devices **41**(8), 1414 (1994)
20.17 A.S. Holland, G.K. Reeves: Microelectron. Reliab. **40**, 965 (2000)
20.18 K. von Klitzing: Prog. Math. Phys. **45**, 1 (2005)
20.19 E.H. Hall: Am. J. Math. **2**, 287 (1879)
20.20 P. Elias, S. Hasenohrl, J. Fedor, V. Cambel: Sens. Actuat. A **101**, 150 (2002)
20.21 A. Vandooren, S. Cristoloveanu, D. Flandre, J.P. Colinge: Solid-State Electron. **45**, 1793 (2001)
20.22 R.L. Petriz: Phys. Rev. **110**, 1254 (1958)
20.23 P. Terziyska, C. Blanc, J. Pernot, H. Peyre, S. Contreras, G. Bastide, J.L. Robert, J. Camassel, E. Morvan, C. Dua, C.C. Brylinski: Phys. Status Solidi (a) **195**(1), 243 (2003)
20.24 D.T. Lu, H. Ryssel: Curr. Appl. Phys. **1**(3–5), 389 (2001)
20.25 G. Rutsch, R.P. Devaty, D.W. Langer, L.B. Rowland, W.J. Choyke: Mat. Sci. Forum **264–268**, 517 (1998)
20.26 P. Blood, J.W. Orton: *The Electrical Characterization of Semiconductor: Majority Carriers and Electron States*, Techniques of Physics, Vol. 14 (Academic, New York 1992)
20.27 Q. Lu, M.R. Sardela Jr., T.R. Bramblett, J.E. Greene: J. Appl. Phys. **80**, 4458 (1996)
20.28 S. Wagner, C. Berglund: Rev. Sci. Instrum. **43**(12), 1775 (1972)
20.29 E.H. Nicollian, J.R. Brews: *MOS (Metal Oxide Semiconductor) Physics and Technology* (Wiley, New York 1982)
20.30 *Model 82-DOS Simultaneous C-V Instruction Manual* (Keithley Instruments, Cleveland 1988)
20.31 W. Beadle, J. Tsai, R. Plummer: *Quick Reference Manual for Silicon Integrated Circuit Technology* (Wiley, New York 1985)
20.32 J. Brews: J. Appl. Phys. **44**(7), 3228 (1973)
20.33 M. Kuhn: Solid-State Electron. **13**, 873 (1970)
20.34 C.N. Berglund: IEEE Trans. Electron Devices **13**(10), 701 (1966)
20.35 S. Witczak, J. Schuele, M. Gaitan: Solid-State Electron. **35**, 345 (1992)
20.36 M.J. Deen: Electron. Lett. **28**(3), 1195 (1992)
20.37 Z.P. Zuo, M.J. Deen, J. Wang: Proc. Can. Conf. Electr. Comput. Eng. (IEEE Press, Piscataway 1989) p. 1038
20.38 A. Raychaudhuri, M.J. Deen, M.I.H. King, W. Kwan: IEEE Trans. Electron Devices **43**(7), 1114 (1996)
20.39 A. Raychaudhuri, M.J. Deen, M.I.H. King, W. Kwan: IEEE Trans. Electron Devices **43**(1), 110 (1996)
20.40 W.S. Kwan, A. Raychaudhuri, M.J. Deen: Can. J. Phys. **74**, S167 (1996)
20.41 T. Matsuda, R. Takezawa, K. Arakawa, M. Yasuda, T. Ohzone, T. Kameda, E. Kameda: Proc. Int. Conf. Microelectron. Test Struct. (ICMTS 2001) (IEEE Press, Piscataway, 2001) p. 65
20.42 G. Groeseneneken, H. Maes, N. Beltram, R. DeKeersmaker: IEEE Trans. Electron Devices **31**, 42 (1984)
20.43 X. Li, M.J. Deen: Solid-State Electron. **35**(8), 1059 (1992)
20.44 X.M. Li, M.J. Deen: IEEE Int. Electron Devices Meet. (IEDM) (IEEE Press, Piscataway 1990) p. 85
20.45 D.S. Ang, C.H. Ling: IEEE Electron Device Lett. **19**(1), 23 (1998)
20.46 H. Uchida, K. Fukuda, H. Tanaka, N. Hirashita: Int. Electron Devices Meet. (1995) p. 41
20.47 N.S. Saks, M.G. Ancona: IEEE Trans. Electron Devices **37**(4), 1057 (1990)
20.48 S. Mahapatra, C.D. Parikh, V.R. Rao, C.R. Viswanathan, J. Vasi: IEEE Trans. Electron Devices **47**(4), 789 (2000)
20.49 Y.-L. Chu, D.-W. Lin, C.-Y. Wu: IEEE Trans. Electron Devices **47**(2), 348 (2000)
20.50 J.S. Bruglers, P.G. Jespers: IEEE Trans. Electron Devices **16**, 297 (1969)
20.51 D. Bauza: J. Appl. Phys. **94**(5), 3239 (2003)
20.52 L.M. Head, B. Le, T.M. Chen, L. Swiatkowski: Proc. 30th Annu. Int. Reliab. Phys. Symp. (1992) p. 228
20.53 S. An, M.J. Deen: IEEE Trans. Electron Devices **47**(3), 537 (2000)
20.54 S. An, M.J. Deen, A.S. Vetter, W.R. Clark, J.-P. Noel, F.R. Shepherd: IEEE J. Quantum Electron. **35**(8), 1196 (1999)
20.55 M.J. Deen, C. Quon: 7th Bienn. Eur. Conf. Insulating Films Semicond. (INFOS 91), ed. by W. Eccleston, M. Uren (IOP Publishing, Liverpool 1991) p. 295
20.56 Z. Celik-Butler: IEE Proc. Circuits Devices Syst. **149**(1), 23 (2002)
20.57 J. Chen, A. Lee, P. Fang, R. Solomon, T. Chan, P. Ko, C. Hu: Proc. IEEE Int. SOI Conf. (1991) p. 100
20.58 J. Sikula: Proc. 17th Int. Conf. Noise Phys. Syst. $1/f$ Fluct. (ICNF 2003), (CNRL, Prague 2003)
20.59 M.J. Deen, Z. Celik-Butler, M.E. Levinshtein (Eds.): SPIE Proc. Noise Devices Circuits, Vol. 5113 (2003)
20.60 C. R. Doering, L. B. Kish, M. Shlesinger: Proc. First Int. Conf. Unsolved Probl. Noise (World Scientific, Singapore 1997)
20.61 D. Abbot, L.B. Kish: *Proceedings of the Second International Conference on Unsolved Problems of Noise and Fluctuations*, American Institute of Physics Conference Proceedings, Vol. 511 (American Institute of Physics, Melville, New York 2000)
20.62 S.M. Bezrukov: *Proceedings of the Third International Conference on Unsolved Problems of Noise and Fluctuations, Washington DC*, American Institute of Physics Conference Proceedings, Vol. 665 (American Institute of Physics, Melville, New York 2000)
20.63 M.J. Deen, S.L. Rumyantsev, M. Schroter: J. Appl. Phys. **85**(2), 1192 (1999)
20.64 M. Sanden, O. Marinov, M. Jamal Deen, M. Ostling: IEEE Electron Device Lett. **22**(5), 242 (2001)
20.65 M. Sanden, O. Marinov, M. Jamal Deen, M. Ostling: IEEE Electron Device Lett. **49**(3), 514 (2002)
20.66 M.J. Deen, J.I. Ilowski, P. Yang: J. Appl. Phys. **77**(12), 6278 (1995)
20.67 M.J. Deen, E. Simoen: IEE Proc. Circuits Devices Syst. **49**(1), 40 (2002)
20.68 M.J. Deen: IEE Proc. Circuits Devices Syst. **151**(2) (2004) Special Issue on Noise in Devices and Circuits

20.69 M.J. Deen, O. Marinov: IEEE Trans. Electron Devices **49**(3), 409 (2002)

20.70 O. Marinov, M.J. Deen, J. Yu, G. Vamvounis, S. Holdcroft, W. Woods: Instability of the noise level in polymer field effect transistors with non-stationary electrical characteristics, 3rd Int. Conf. Unsolved Probl. Noise Fluct. (UPON 02), Washington DC (AIP, Melville 2002)

20.71 M. Marin, M.J. Deen, M. de Murcia, P. Llinares, J.C. Vildeuil: IEE Proc. Circuits Devices Syst. **151**(2), 95 (2004)

20.72 A. Chandrakasan: Proc. Eur. Solid-State Circuits Conf. (ESSCIRC 2002) (AIP, Melville 2002) p. 47

20.73 R. Brederlow, W. Weber, D. Schmitt-Landsiedel, R. Thewes: IEDM Tech. Dig. (1999) p. 159

20.74 L. Chaar, A. van Rheenen: IEEE Trans. Instrum. Meas. **43**, 658 (1994)

20.75 C.-Y. Chen, C.-H. Kuan: IEEE Trans. Instrum. Meas. **49**, 77 (2000)

20.76 C. Ciofi, F. Crupi, C. Pace, G. Scandurra: IEEE Trans. Instrum. Meas. **52**, 1533 (2003)

20.77 P. Kolev, M.J. Deen: Development and applications of a new DLTS method and new averaging techniques. In: *Advances in Imaging and Electron Physics*, Vol. 109, ed. by P. Hawkes (Academic, New York 1999)

20.78 P. McLarty: Deep level transient spectroscopy (DLTS). In: *Characterization Methods for Submicron MOSFETs*, ed. by H. Haddara (Kluwer, Boston 1996) p. 109

20.79 P.V. Kolev, M.J. Deen: J. Appl. Phys. **83**(2), 820 (1998)

20.80 P. Kolev, M.J. Deen, T. Hardy, R. Murowinski: J. Electrochem. Soc. **145**(9), 3258 (1998)

20.81 P. Kolev, M.J. Deen, J. Kierstead, M. Citterio: IEEE Trans. Electron Devices **46**(1), 204 (1999)

20.82 P.V. Kolev, M.J. Deen, N. Alberding: Rev. Sci. Instrum. **69**(6), 2464 (1998)

20.83 P. Kolev, M.J. Deen: Proc. 4th Symp. Low Temp. Electron. High Temp. Supercond., Vol. 97-2, ed. by C. Claeys, S.I. Raider, M.J. Deen, W.D. Brown, R.K. Kirschman (The Electrochemical Society Press, New Jersey 1997) p. 147

Part C Materials

Part C Materials for Electronics

21 **Single-Crystal Silicon: Electrical and Optical Properties**
Mark Auslender, Beer Sheba, Israel
Shlomo Hava, Beer Sheba, Israel

22 **Silicon-Germanium: Properties, Growth and Applications**
Yaser M. Haddara, Hamilton, Canada
Peter Ashburn, Southampton, UK
Darren M. Bagnall, Sydney, Australia

23 **Temperature-Insensitive Band-Gap III-V Semiconductors: TI-III-V and III-V-Bi**
Hajime Asahi, Osaka, Japan

24 **Amorphous Semiconductors: Structure, Optical, and Electrical Properties**
Kazuo Morigaki, Tokyo, Japan
Chisato Ogihara, Yamaguchi, Japan

25 **Amorphous and Microcrystalline Silicon**
Akihisa Matsuda, Osaka, Japan

26 **Ferroelectric Materials**
Roger Whatmore, London, UK

27 **Dielectric Materials for Microelectronics**
Robert M. Wallace, Richardson, USA

28 **Thin Films**
Robert D. Gould†
Safa Kasap, Saskatoon, Canada
Asim K. Ray, Middlesex, UK

29 **Thick Films**
Neil White, Southampton, UK

21. Single-Crystal Silicon: Electrical and Optical Properties

Mark Auslender, Shlomo Hava

Electrical and optical properties of crystalline semiconductors are important parts of pure physics and material science research. In addition, knowledge of parameters related to these properties, primarily for silicon and III–V semiconductors, has received a high priority in microelectronics and optoelectronics since the establishment of these industries. For control protocols, emphasis has recently been placed on novel optical measurement techniques, which have proved very promising as nondestructive and even non-contact methods. Earlier they required knowledge of the free-carrier-derived optical constants, related to the electrical conductivity at infrared frequencies, but interest in the optical constants of silicon in the visible, ultraviolet (UV) and soft-x-ray ranges has been revived since the critical dimensions in devices have become smaller.

This chapter surveys the electrical (Sect. 21.2) and optical (Sect. 21.3) properties of crystalline silicon. Section 21.2 overviews the basic concepts. Though this section is bulky and its material is documented in textbooks, it seems worth including since the consideration here focuses primarily on silicon and is not spread over other semiconductors – this makes the present review self-contained. To avoid repeated citations we, in advance, refer the reader to stable courses on solidstate physics [21.1, 2], semiconductor physics [21.3], semiconductor optics [21.4] and electronic devices [21.5]; seminal papers are cited throughout Sect. 21.2.

21.1 **Silicon Basics** 485
21.1.1 Structure and Energy Bands 485
21.1.2 Impurity Levels and Charge-Carrier Population 488
21.1.3 Carrier Concentration, Electrical and Optical Properties 490
21.1.4 Theory of Electrical and Optical Properties 491

21.2 **Electrical Properties** 495
21.2.1 Ohm's Law Regime 495
21.2.2 High-Electric-Field Effects 507
21.2.3 Review Material 514

21.3 **Optical Properties** 515
21.3.1 Diversity of Silicon as an Optical Material 515
21.3.2 Measurements of Optical Constants 515
21.3.3 Modeling of Optical Constants 516
21.3.4 Electric-Field and Temperature Effects on Optical Constants 520

References 520

We realize how formidable our task is – publications on electrical and optical properties of silicon amount to a huge number of titles, most dating back to the 1980s and 1990s – so any review of this subject will inevitably be incomplete. Nevertheless, we hope that our work will serve as a useful shortcut into the silicon world for a wide audience of applied physics, electrical and optical engineering students.

21.1 Silicon Basics

21.1.1 Structure and Energy Bands

Normally silicon (Si) crystallizes in a diamond structure on a face-centered cubic (fcc) lattice, with a lattice constant of $a_0 = 5.43$ Å. The basis of the diamond structure consists of two atoms with coordinates (0, 0, 0) and $a_0/4(1, 1, 1)$, as seen in Fig. 21.1. Other solids that can crystallize in the diamond structure are C, Ge and Sn. The important notion for the electronic band structure is the Brillouin zone (BZ). The BZ is a primitive cell in the reciprocal-space lattice, which proves to be a body-centered cubic (bcc) lattice for an fcc real-space lattice. For this case, the BZ with important reference points and directions within it is shown in Fig. 21.2.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_21

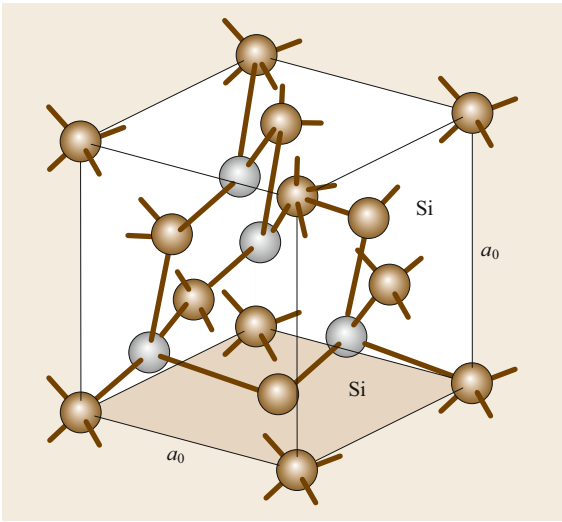


Fig. 21.1 Diamond crystal structure of Si

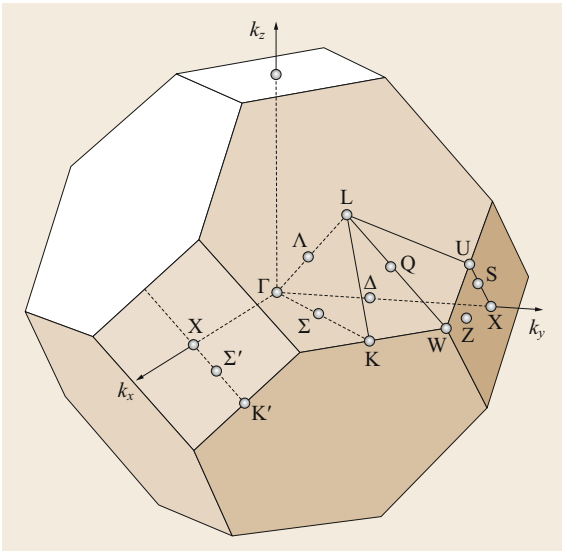


Fig. 21.2 Brillouin zone of the fcc lattice

The states of electrons in solids are described by wave functions of the Bloch type

$$\psi(\boldsymbol{r}) = \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} u_{s\boldsymbol{k}}(\boldsymbol{r}) \,, \tag{21.1}$$

where $\boldsymbol{k}$ is the wave vector that runs over reciprocal space, s is a band index and $u_{s\boldsymbol{k}}(\boldsymbol{r})$ is the periodic function of the direct lattice (Bloch amplitude). Both $u_{s\boldsymbol{k}}(\boldsymbol{r})$ and the corresponding energy-band spectrum $E_s(\boldsymbol{k})$ are periodic in $\boldsymbol{k}$, which allows one to restrict consideration to within the BZ. The bands are arranged so that there are energy regions for which no states given by (21.1) exist. Such forbidden regions are called energy gaps or band gaps and result from the interaction of valence electron with the ion cores of crystal. In semiconductor science the term *band gap* is accepted for the energy distance between the maximum of $E_s(\boldsymbol{k})$ for the highest filled (valence) band and the minimum of $E_s(\boldsymbol{k})$ for the lowest empty (conduction) band (denoted by E_g). The band gap is called *direct* if the aforementioned maximum and minimum occur at the same point of the BZ, e.g., Γ (Fig. 21.2), and *indirect* if they occur at different points of the BZ, e.g., Γ and X (Fig. 21.2).

Si is an indirect-band-gap semiconductor with $E_\mathrm{g} = 1.1700\,\mathrm{eV}$ at 4.2 K. The calculated energy-band structure, that is the curves of $E_\nu(\boldsymbol{k})$ for selected directions in the BZ, is shown in Fig. 21.3a. The conduction-band minimum lies at six equivalent points Δ on the Γ–X lines (Fig. 21.2). In some vicinity (called the *valley*) of every such point the band spectrum is quadratic in $\boldsymbol{k}$, e.g., for the valley ⟨100⟩

$$E_\mathrm{c}(\boldsymbol{k}) = E_{\mathrm{c}0} + \frac{\hbar^2 (k_x - k_0)^2}{2m_\mathrm{l}} + \frac{\hbar^2 \left(k_y^2 + k_z^2\right)}{2m_\mathrm{t}} \,, \tag{21.2}$$

where $k_0 \approx 1.72\pi/a_0$, m_l and m_t are the longitudinal and transverse effective masses. The spectra for other five valleys are obtained from (21.2) by 90° rotations and inversions $k_0 \to -k_0$. Though the constant-energy surface for (21.2) is an ellipsoid (Fig. 21.3b), the density of states (DOS) proves to be the same as for an isotropic parabolic spectrum with an effective mass

$$m_\mathrm{de} = 6^{2/3} m_\mathrm{l}^{1/3} m_\mathrm{t}^{2/3} \,, \tag{21.3}$$

which is called the *DOS effective mass*. Another mass, m_ce, which appears in the direct-current (DC) and optical conductivity formulas, is defined via the harmonic mean

$$\frac{1}{m_\mathrm{ce}} = \frac{1}{3}\left(\frac{1}{m_\mathrm{l}} + \frac{2}{m_\mathrm{t}}\right) . \tag{21.4}$$

For this reason, m_c is called the *conductivity/optical effective mass*. Equation (21.2) holds at $E_\mathrm{c} - E_{\mathrm{c}0} < 0.15\,\mathrm{eV}$, but at larger energies the ellipsoids warp strongly, especially near the X point; the change of the spectrum with energy is mostly due to the increasing m_t, while m_l increases weakly [21.6].

The valence-band maximum is at the Γ point ($\boldsymbol{k} = 0$), where the Bloch-wave state $u_{n0}(\boldsymbol{r})$ has the full symmetry of an atomic p-orbital, being six-fold degenerate in the nonrelativistic limit. The spin–orbit interaction splits off from the bare band top a four-fold degenerate $\mathrm{p}_{3/2}$ level up by $1/3\,\Delta_\mathrm{so}$, and a two-fold degenerate $\mathrm{p}_{1/2}$

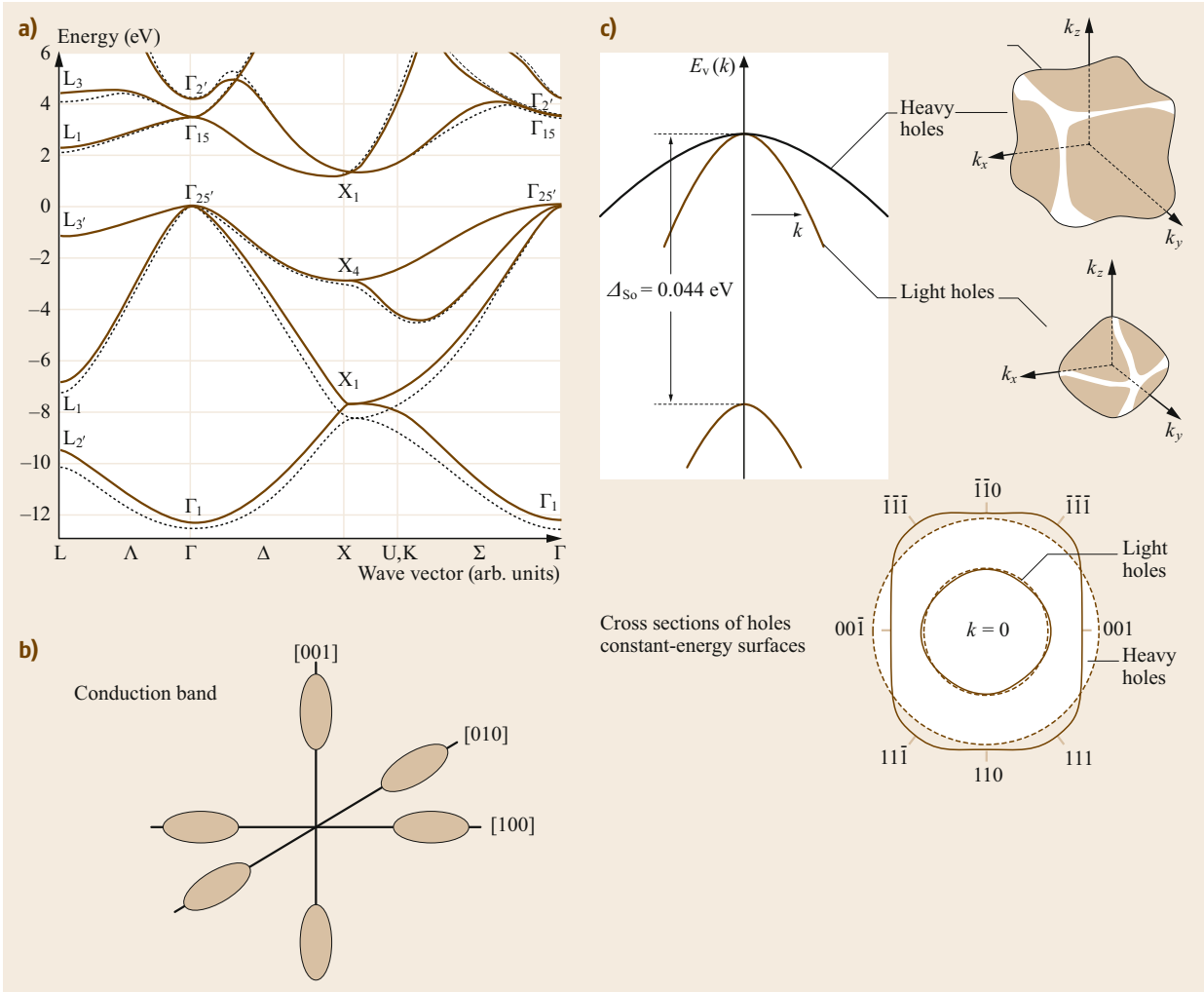


Fig. 21.3a–c Electronic band structure of Si: **(a)** Energy dispersion curves near the fundamental gap. (After [21.7] with permission); **(b)** The constant-energy ellipsoids of the conduction band; **(c)** The constant-energy warped spheres and schematic of the valence band

level down by $-2/3\Delta_{so}$. The spectrum at $\boldsymbol{k} \neq 0$, even at energies near the top E_{v0}, is very complex. It consists of three branches, which are in general nonparabolic and nonisotropic [21.8]. At $E_{v0} - E_v \ll \Delta_{so}$ there are two nonparabolic anisotropic $p_{3/2}$-derived sub-bands with the energy spectra

$$E_{v1,2}(\boldsymbol{k}) = E_{v0} + \frac{\hbar^2}{2m_0} \times \left[Ak^2 \pm \sqrt{B^2k^4 + C^2\left(k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2\right)}\right], \quad A < 0 \tag{21.5}$$

and at $E_{v0} - E_v \ll 2\Delta_{so}$ there is an isotropic parabolic $p_{1/2}$-derived split-off band with the energy spectrum

$$E_{v3}(\boldsymbol{k}) = E_{v0} - \Delta_{so} + \frac{\hbar^2k^2}{2m_0}A', \quad A' < 0. \tag{21.6}$$

Here m_0 is the free electron mass, A, B, C and A' are inverse hole-mass parameters (for small Δ_{so}, $A' \approx A$). In (21.5) the $-$ sign corresponds to heavy holes and the $+$ sign to light holes. The constant-energy surfaces for (21.5) are warped spheres (Fig. 21.3c), the DOS is nevertheless parabolic and is described by the effective

Table 21.1 The conduction-band effective masses and valence-band parameters for Si

Conduction band					
Mass	m_l/m_0	m_t/m_0	m_{de}/m_0	m_{ce}/m_0	
Experiment	0.97 ± 0.04[a]	0.19 ± 0.01[a]	1.08 ± 0.05	0.26 ± 0.01	
	0.9163 ± 0.04[b]	0.1905 ± 0.0001[b]	1.0618 ± 0.0005	0.2588 ± 0.0001	
Theory	0.971[c]	0.205[c]	1.137	0.2652	
	0.9716[d]	0.1945[d]	1.0978	0.2780	
Valence band					
Parameter	Δ_{so}	A	$\lvert B\rvert$ or B	$\lvert C\rvert$	m_3/m_0
Experiment	0.0441 ± 0.004[g]	-4.1 ± 0.2[a]	1.6 ± 0.2[a]	3.3 ± 0.2[a]	0.24 ± 0.01
		-4.28 ± 0.02[e]	-0.75 ± 0.04[e]	5.25 ± 0.05[e]	0.234 ± 0.001
		-4.27 ± 0.02[h]	-0.63 ± 0.08[h]	4.93 ± 0.15[h]	0.234 ± 0.001
Theory		−4.38[c]	0.84[c]	4.11[c]	0.23
	0.04[d]	−4.38[d]	−1.00[d]	4.80[d]	0.23
	0.044[f]	−4.22[f]	−0.78[f]	4.80[f]	0.24

[a] After [21.9], [b] after [21.10], [c] after [21.11], [d] after [21.12], [e] after [21.13], [f] after [21.14], [g] after [21.15], [h] after [21.16]

Table 21.2 The valence-band effective masses calculated using experimental data

Mass	m_{d1}/m_0	m_{d2}/m_0	m_d/m_0	m_{c1}/m_0	m_{c2}/m_0
Exp.[a]	0.55 ± 0.12	0.16 ± 0.01	0.61 ± 0.12	0.51 ± 0.10	0.16 ± 0.01
Exp.[b]	0.58 ± 0.02	0.151 ± 0.001	0.63 ± 0.02	0.43 ± 0.01	0.145 ± 0.001
Exp.[c]	0.49–0.56	0.153–0.158	0.54–0.62	0.40–0.43	0.147–0.152

[a] After [21.9], [b] after [21.13], [c] after [21.16]

mass

$$m_{dh} = \left(m_{d1}^{3/2} + m_{d2}^{3/2}\right)^{2/3}, \tag{21.7}$$

where $m_{d1,2}$ are the partial DOS effective masses. Quite different masses enter various physical quantities for heavy (light) holes, and are complicated functions of A, B and C [21.17], however, the split-off band is characterized by only one effective mass: $m_3 = m_0/|A'|$.

These band-structure parameters, obtained from cyclotron-resonance data and calculations, are presented in Tables 21.1 and 21.2. The unreferenced values of m_{de}, m_{ce} and m_3 were calculated from the referenced data using (21.3) and (21.4) and the assumption $A' = A$. Optical data for m_{ce} are discussed in Sect. 21.3, while the specific-heat data for m_{de} are not considered here.

Considerable uncertainty and errors in the values of B and C have a small effect on the light-hole effective-mass values (m_{d2} and optical m_{c2}) but lead to an ambiguity in the heavy-hole effective-mass (m_{d1} and optical m_{c1}) values (Table 21.2). In each data set $m_{d2} \approx m_{c2}$, which means that an isotropic approximation is reasonable for the light-hole spectrum. The experimental data [21.13, 16] and band-structure calculations [21.14] are in good agreement; the former is used in theoretical papers on hole transport in Si [21.18, 19].

Nonparabolic parts in the electron [21.6] and hole spectra [21.8] lead to apparent dependence of effective-mass measurements on the temperature and carrier concentration [21.20]. Recently the DOS mass issue was revisited [21.21] in connection with the intrinsic carrier concentration.

21.1.2 Impurity Levels and Charge-Carrier Population

In highly purified Si electrons and holes are thermally generated with equal concentrations ($n = p$), equal the intrinsic concentration n_i given by

$$n_i = 2\left(\frac{2\pi\sqrt{m_{de}m_{dh}}k_B T}{h^2}\right)^{3/2} \exp\left(-\frac{E_g}{2k_B T}\right), \tag{21.8}$$

where k_B and h are the Boltzmann and Planck constants, respectively, and T is the absolute temperature. To assess $n_i(T)$ theoretically one should take into account possible changes of $m_{de,h}$ and E_g with T. An early experimental plot of $n_i(T)$ for Si is shown in Fig. 21.4, giving $n_i(300\,\mathrm{K}) = 1.38 \times 10^{10}\,\mathrm{cm}^{-3}$. This and the early textbook value ($1.45 \times 10^{10}\,\mathrm{cm}^{-3}$) do not fit (21.8) with the established low-temperature values of $m_{de,h}$ and $E_g(300\,\mathrm{K}) = 1.1242\,\mathrm{eV}$, giving $(0.662–0.694) \times 10^{10}\,\mathrm{cm}^{-3}$ for the best available data. This discrepancy was eliminated [21.21] by critically evaluating measurements of $n_i(T)$ and incorporating the latest data for the hole DOS effective mass. The reassessed value of

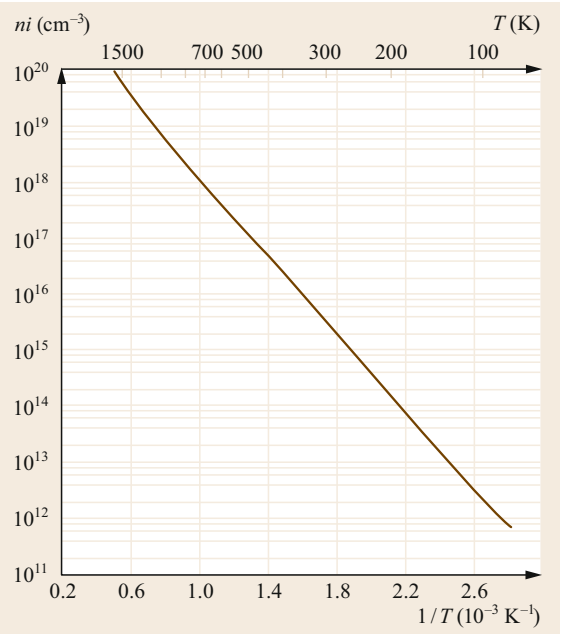


Fig. 21.4 Intrinsic concentration in Si versus temperature. (After [21.24] with permission)

n_i (300 K), verified by measurements on solar-cell devices [21.22], is $1.08(8)\times10^{10}\mathrm{cm}^{-3}$ (in a very recent textbook [21.23] the unreferenced value $n_i(300\,\mathrm{K}) = 1.00\times10^{10}\,\mathrm{cm}^{-3}$ is adopted). This is obtained from (21.8) using $m_{\mathrm{de}}(300\,\mathrm{K}) = 1.08m_0$, which is quite close to the tabular values presented, and $m_{\mathrm{dh}}(300\,\mathrm{K}) = 1.15m_0$, which is 1.8–2.1 times larger than the tabular values presented. Such a strong increase of m_{dh} when increasing T from 4.2 to 300 K is anticipated from the nonparabolic part of the hole spectra [21.8].

In impure samples $n \neq p$, the charge-neutrality condition is then

$$n - N_{\mathrm{d}}^{+} = p - N_{\mathrm{a}}^{-} \tag{21.9}$$

together with the known expressions for n, p (with $np = n_i^2$ holding until the carriers become degenerate) and the concentrations of charged donors N_{d}^{+} and acceptors N_{a}^{-} via the Fermi energy E_{F} and impurity ionization energies E_{d} and E_{a} allows one to calculate E_{F} (and hence all the concentrations) as a function of T and doping.

An impurity is called *shallow* if $|E_{c(v)0} - E_{\mathrm{d(a)}}| \ll E_{\mathrm{g}}$, and *deep* if $|E_{c(v)0} - E_{\mathrm{d(a)}}| \approx 0.5E_{\mathrm{g}}$. Shallow group V donors (Sb, P, As) and group III acceptors (B, Al, Ga, In) are well soluble in Si [21.25]. The ionization energy was calculated using the effective-mass approximation, and the Bohr model for donors (*Kohn* and *Luttinger*, 1955 [21.26, 27]) and acceptors (*Kohn* and *Schechter*, 1955 [21.28]). The value of $E_{\mathrm{d(a)}}$ calculated in this way is insensitive to the specific shallow donor (acceptor). Actual thermal (i. e., retrieved from electrical measurements) ionization energies are given in Table 21.3. It is seen that shallow impurities have different ionization energies. This difference is small for donors and larger for acceptors. The same trend was observed for optical ionization energies, although they are different from the thermal values [21.29, 30]. For shallow impurities, $E_{\mathrm{d(a)}}$ decreases as doping becomes heavier, and becomes zero as $N_{\mathrm{d(a)}}$ approaches the corresponding insulator–metal transition concentration [21.31]. Deep impurities (except for Mn, Fe and Zn) are amphoteric, i. e., they act simultaneously as donors and acceptors. For such a donor (acceptor) state, the level may lie closer to the valence (conduction) than the conduction (valence) band. The deep impurities are mostly unionized at room temperature due to their large $E_{\mathrm{d(a)}}$, so their direct contribution to n or p is negligible. The unionized deep impurities may, however, trap the carriers available from the shallow impurities or injection, thus decreasing the conductivity or the minority-carrier lifetime. Atoms that behave in Si in this manner, for example Au, Ag and Cu, are added for lifetime control. The properties of these impurities in Si have been studied in detail [21.32, 33].

Table 21.3 Energy levels of impurities in Si

Impurity	Donor/ acceptor	Below conduction band (eV)	Above valence band (eV)
Ag	D	0.310	
	A		0.210
Al	A		0.057
As	D	0.049	
Au	D		0.330
	A	0.540	
B	A		0.045
Cu	A		0.490
	D		0.240
Fe	D	0.550	
	D		0.400
Ga	A		0.065
In	A		0.160
Li	D	0.033	
Mn	D	0.530	
P	D	0.044	
Sb	D	0.039	
S	D	0.180	
	D	0.370	
Zn	A	0.550	
	A		0.300

21.1.3 Carrier Concentration, Electrical and Optical Properties

Concentration and Electrical Measurements

Measurements of carrier concentrations, as well as electrical and optical characteristics are most tractable if either $n \gg p$ (strongly n-type conduction) or $p \gg n$ (strongly p-type conduction). Since the np product is constant versus doping, the contribution of minority carriers to the conductivity becomes unimportant when $N_{d(a)}$ increases significantly over n_i. A standard route for determining $n(p)$ is Hall-effect measurements. The Hall coefficient R^H, measured directly on a long thin slab in a standard crossed electric and magnetic field configuration, is retrieved by

$$R^H = 10^{-8}\frac{V_H d}{IB}\,, \tag{21.10}$$

where V_H is the Hall voltage (volt), I is the current (ampere), d is the sample thickness (cm) in the z-direction, and B is the magnetic field strength (Gauss) applied in this direction. There are two limiting cases. One, the high-field regime is defined by $qB\tau/mc \gg 1$, where m and τ are the appropriate mass and relaxation-time parameters, respectively. In this case

$$\begin{aligned} R^H_e(\infty) &= -\frac{1}{qn}\,, \\ R^H_h(\infty) &= \frac{1}{qp}\,. \end{aligned} \tag{21.11}$$

The other, low-field, regime holds with the opposite inequality; in this regime

$$\begin{aligned} R^H_e(0) &= -\frac{r_e}{qn}\,, \\ R^H_h(0) &= \frac{r_h}{qp}\,, \end{aligned} \tag{21.12}$$

where the constant of proportionality $r_{e(h)}$, called the electron (hole) Hall factor, depends on the details of the scattering process and band structure. Thus the majority-carrier concentration is determined directly from $R^H_{e(h)}$ using a high-field Hall measurement. For typical laboratory magnetic fields, this regime is attainable only with extremely high mobility and low effective mass, which excludes moderately and heavily doped Si, for which very high magnetic fields are required. In some cases the Hall factor is quite close to unity, e.g., $r_e = 3\pi/8$ for the phonon scattering in the isotropic and parabolic (standard) band.

The electrical properties are fully described by the drift-diffusion relation for the electron (hole) current density $\boldsymbol{j}_{e(h)}$

$$\begin{aligned} \boldsymbol{j}_e &= -qn\boldsymbol{v}_{de} + qD_e\nabla n\,, \\ \boldsymbol{j}_h &= qp\boldsymbol{v}_{dh} - qD_h\nabla p\,, \end{aligned} \tag{21.13}$$

where $\boldsymbol{v}_{de} = -\mu_e E$ ($\boldsymbol{v}_{dh} = \mu_h E$) is the drift velocity, $\mu_{e(h)}$ is the drift mobility, E is the electric field strength, and $D_{e(h)}$ is the diffusion coefficient; in general, $\mu_{e(h)}$ and $D_{e(h)}$ depend on E. In the homogeneous case (21.13) converts into the material equation $\boldsymbol{j}_{e(h)} = \sigma_{e(h)}E$, where $\sigma_{e(h)} = qn\mu_{e(h)}$ is the electron (hole) conductivity; the total conductivity equals $\sigma = \sigma_e + \sigma_h$. In the weak-field DC (ohm) and alternating current (AC: microwave or light, except for intense laser, irradiation) regimes, $D_{e(h)}$ is proportional to $\mu_{e(h)}$ being both constant versus E, depending on the radiation frequency ω.

Combining the high-induction Hall and Ohm resistivity ($\rho = \sigma^{-1}$) measurements one obtains the drift mobility

$$\mu_{e(h)} = R^H_{e(h)}(\infty)\sigma\,. \tag{21.14}$$

Replacing $R^H_e(\infty)$ by $R^H_e(0)$ in the right-hand side of (21.14), one arrives at the so-called Hall mobility

$$\mu^H_{e(h)} = R^H_{e(h)}(0)\sigma = r_{e(h)}\mu_{e(h)}\,, \tag{21.15}$$

which never equals the drift mobility, although it may be fairly close to it in the cases mentioned above. In general, when $n(p)$ is extracted from $R^H_{e(h)}(0)$, the calculation of the $r_{e(h)}$ factor is completed. Magnetoresistance (MR), i. e., ρ versus B measurement, is an important experimental tool as well. Another established method is the Haynes–Shockley experiment, which allows one to measure the minority-carrier drift mobility. In high-electric-field conditions, a noise-measurement technique is used. A relatively novel, time-of-flight technique was used in the latest (to our knowledge) mobility and diffusion-coefficient measurements on lightly doped crystalline Si samples, both in the low- and high-field regimes [21.19].

Basic Optical Parameters

The electromagnetic response of homogeneous nonmagnetic material is governed by the dielectric constant tensor ε, which connects the electric displacement vector $\boldsymbol{D}$ inside the material to $\boldsymbol{E}$ through the material equation $\boldsymbol{D} = \varepsilon E$. For cubic crystals, such as Si, ε is a scalar. An effective-medium homogeneous dielectric constant may be attributed to inhomogeneous and composite materials if the nonhomogeneity feature size is smaller than the radiation wavelength $\lambda = 2\pi c/\omega$. Actually, ε characterizes the material's bulk and therefore loses its sense in nanoscale structures (superlattices,

quantum wells etc.). The dependence $\varepsilon(\omega)$ expresses the optical dispersion in the material. The dielectric constant is usually represented via its real and imaginary parts: $\varepsilon = \varepsilon_1 + \mathrm{i}\varepsilon_2$ $(\varepsilon_2 \geq 0)$, connected to each other by the Kramers–Kronig relation (KKR)

$$\varepsilon_1(\omega) = 1 + \frac{2}{\pi} \int_0^\infty \frac{\Omega}{\Omega^2 - \omega^2} \varepsilon_2(\Omega) \mathrm{d}\Omega \,. \tag{21.16}$$

At low frequency (radio, microwave), in the absence of magnetic fields, $\varepsilon_1 \approx \varepsilon(0)$ and ε_2, which is responsible for dielectric loss, is small. At optical wavelengths, from far-infrared (IR) to soft x-rays, the basic quantity is the complex refractive index $\sqrt{\varepsilon} = n + \mathrm{i}k$. The real refractive index n, which is responsible for wave propagation properties, and the extinction index k, responsible for the field attenuation, are referred to as *optical constants*. They are related to the dielectric constant via

$$\begin{aligned} \varepsilon_1 &= n^2 - k^2 \,, \\ \varepsilon_2 &= 2nk \,, \\ n &= \sqrt{\frac{\left(\varepsilon_1^2 + \varepsilon_2^2\right)^{1/2} + \varepsilon_1}{2}} \,, \\ k &= \sqrt{\frac{\left(\varepsilon_1^2 + \varepsilon_2^2\right)^{1/2} - \varepsilon_1}{2}} \,. \end{aligned} \tag{21.17}$$

21.1.4 Theory of Electrical and Optical Properties

Boltzmann-Equation Approach

The response of carriers in a band to perturbations away from the thermal-equilibrium state, such as applied electric and magnetic fields or impinging electromagnetic radiation, is described by the deviation of the carrier distribution function $f_\mathrm{s}(\boldsymbol{k}, \boldsymbol{r}, t)$ from the equilibrium Fermi–Dirac distribution $f_0[E_s(\boldsymbol{k})]$. The current density equals $\boldsymbol{j}_\mathrm{s} = q \int \boldsymbol{v}_\mathrm{s}(\boldsymbol{k}) f_\mathrm{s}(\boldsymbol{k}, r, t) \mathrm{d}\boldsymbol{k}$, where $\boldsymbol{v}_\mathrm{s}(\boldsymbol{k}) = \partial E_s(\boldsymbol{k}) / (\hbar \partial \boldsymbol{k})$ is the microscopic carrier velocity and the integration is performed over the BZ. The process that balances the external perturbations is scattering of carriers by lattice vibrations (phonons), impurities and other carriers. Impurity scattering dominates transport at low temperatures and remains important at room temperature for moderate and high doping levels, although carrier–carrier scattering also becomes appreciable. Under appropriate conditions, one being that $\hbar\omega \ll \bar{E}$ (where $\bar{E}$ is the average carrier kinetic energy), $f_\mathrm{s}(\boldsymbol{k}, \boldsymbol{r}, t)$ satisfies the quasi-classical Boltzmann kinetic equation. In the opposite, quantum, range, radiation influences the scattering process. Generalized kinetic equations, which interpolate between the quasi-classical and quantum regimes, have also been derived [21.34].

There exist various methods of solving the quasi-classical Boltzmann equation. The relaxation-time method, variational method [21.35–37] for low electric fields, and displaced-Maxwell-distribution approximation [21.38] for high electric fields, were used in early studies. In the last four decades the Monte Carlo technique [21.39], which overcomes limitations inherent to these theories and allows one to calculate subtle details of the carrier distribution, has been applied to various semiconductors, including crystalline Si [21.19]. If $\hbar\omega \gg \bar{E}$, the kinetic and optical characteristics are calculated well using transition probabilities between carrier states, with the radiation quantum absorbed or emitted [21.40]. The most problematic is the intermediate range – generalized kinetic equations have been shown to recover the extreme ranges, but no working methods have been developed for solutions at $\hbar\omega \approx \bar{E}$, to the best of our knowledge. The relaxation-time approximation has proved to work well in many cases of scattering in Si. In this framework, the basic quantity is the relaxation-time tensor $\tau_\mathrm{s}(E)$, where s is the index indicating the conduction or valence band, and E is the carrier kinetic energy. For electron valleys, $\tau_\mathrm{c}(E)$ has the same symmetry as the respective effective-mass tensor – it is diagonal with principal values of, e.g., τ_l, τ_t, τ_t for $\langle 100 \rangle$ etc. For scalar holes, $\tau_{1,2}(E)$ can only be introduced in the isotropic-bands approximation. The mobility and the Hall factor for weak electric fields in the relaxation-time approximation are given by

$$\begin{aligned} \mu_\mathrm{e}(\omega) &= \frac{q}{3} \left\langle \frac{\tau_\mathrm{l}/m_\mathrm{l}}{1 - \mathrm{i}\omega\tau_\mathrm{l}} + \frac{2\tau_\mathrm{t}/m_\mathrm{t}}{1 - \mathrm{i}\omega\tau_\mathrm{t}} \right\rangle , \\ r_\mathrm{e} &= \frac{3 \left\langle 2\tau_\mathrm{l}\tau_\mathrm{t}/m_\mathrm{l}m_\mathrm{t} + (\tau_\mathrm{t}/m_\mathrm{t})^2 \right\rangle}{\langle \tau_\mathrm{l}/m_\mathrm{l} + 2\tau_\mathrm{t}/m_\mathrm{t} \rangle^2} ; \end{aligned} \tag{21.18}$$

$$\begin{aligned} \mu_\mathrm{h}(\omega) &= \frac{q}{1+\beta} \left\langle \frac{\tau_1/m_1}{1 - \mathrm{i}\omega\tau_1} + \frac{\beta\tau_2/m_2}{1 - \mathrm{i}\omega\tau_2} \right\rangle , \\ r_\mathrm{h} &= \frac{(1+\beta) \left\langle (\tau_1/m_1)^2 + \beta\,(\tau_2/m_2)^2 \right\rangle}{\langle \tau_1/m_1 + \beta\tau_2/m_2 \rangle^2} , \end{aligned} \tag{21.19}$$

where the angular brackets indicate averaging with the weight $-E^{3/2} f_0'(E)$, $\beta = (m_{\mathrm{d}2}/m_{\mathrm{d}1})^{3/2}$ is the density ratio of light holes to heavy holes, and the option for a nonparabolic band is retained.

Lattice Scattering

Deformational phonons – longitudinal, transverse acoustical (LA, TA) and optical (LO, TO) – mediate carrier–lattice scattering in Si. The phonon modes are

Table 21.4 Parameters of the phonon modes in crystalline Si

Mode	Energy (K)				Sound velocity (cm/s)
LO	760	700–735[a]	560	580	
TO	760	–	630–690[a]	680	
LA	0	240–260	500–510[a]	580	8.99×10^5
TA	0	140–160	210–260	220	5.39×10^5
q	Γ	Δ	S	X	

[a] After [21.9]

Table 21.5 Deformation-potential parameters [21.41]

T (K)	Ξ_u (eV)	$\Xi_d + 1/3\Xi_u - a$ (eV)	$\lvert b\rvert$ (eV)	$\lvert d\rvert$ (eV)
80	8.6 ± 0.2	3.8 ± 0.5	2.4 ± 0.2	5.3 ± 0.4
295	9.2 ± 0.3	3.1 ± 0.5	2.2 ± 0.3	–

presented in Table 21.4, where **q** is a point in the phonon BZ. The points Δ and S correspond to the scattering process, where the electron transits between the bottoms of two perpendicularly (f) and parallel (g) oriented valleys, respectively. The phonon energies are precise at the points Γ and X, as determined by neutron-scattering techniques [21.42], but uncertain at Δ and S, since in this case only estimation and fitting methods were available.

The rigid- and deformable-ion lattice models have been used to obtain the carrier–phonon interaction for electrons [21.43] and holes [21.44]. A deformation-potential theory of the interaction with long-wavelength phonons, which takes the crystal symmetry and band structure fully into account in a phenomenological manner, has been developed. This theory deduces two constants, Ξ_u, Ξ_d, for the conduction band [21.45] and four, a, b, d, d_{opt}, for the valence band [21.46, 47], which are presented in Table 21.5. The deformation-potential theory has been used to calculate the acoustic scattering-limited mobility in n-Si [21.48] and p-Si [21.49]. Later, optical-phonon scattering, along with an approximate valence-band spectrum instead of (21.5), were taken into account [21.18, 19].

The matrix elements of electron–phonon interaction between wave functions of different valleys are not taken into account by the deformation-potential theory. For inter-valley transitions, other than the three marked in Table 21.4 by the superscript a, the matrix elements calculated at the valley-bottom wave vectors are zero [21.43, 50]. The actual scattering probabilities are never zero; for those forbidden by selection rules [21.43, 50] one should take into account the wave-vector offset at the final scattering state, which gives nominally small, but unknown values. Several inter-valley scattering models have been tried to fit the theoretical formulas to the mobility data in lightly doped n-Si: with one allowed TO and one forbidden TA phonon [21.51], one allowed TO phonon [21.52] and more involved combinations of the transitions [21.53, 54]. The scattering of electrons by long-wavelength optical phonons, regarded as a cause of drift-velocity saturation at high electric fields [21.38], is forbidden in Si [21.43]. In n-Si, by all accounts, the cause may be the allowed g-phonon (Table 21.4) scattering [21.54].

Impurity Scattering

There are two types of impurity scattering – by ionized and neutral impurities. The latter is the dominant impurity scattering for uncompensated, light or moderate shallow-impurity doping, at low T. In samples doped with deep impurities, neutral-impurity scattering may also show up. At elevated T, when shallow impurities are increasingly ionized, ionized-impurity scattering is the dominant impurity scattering and may compete with phonon scattering, depending on $N_{\text{d(a)}}$. Lastly, in heavily doped samples, where the impurities are ionized for all T, ionized-impurity scattering is dominant up to 300 K.

Early theories of impurity scattering were developed for carriers in parabolic bands, scattered by hydrogen-like centers. For neutral impurities the s-scattering cross section [21.55] and a cross section that takes allowance of the scattered carrier's bound state [21.56] were adopted. For ionized centers, use was made of the Coulomb scattering cross section, cut off at a small angle depending on $N_{\text{d(a)}}$ [21.57], and the screened Coulomb potential cross section, calculated in the Born approximation [21.58–60]. These theories consider scattering by the donors and acceptors on an equal footing. The use of the Conwell–Weisskopf formula for $\tau_s(\varepsilon)$ declined towards the end of the 1950s, while the corresponding Brooks–Herring (BH) formula became widespread, mostly due the consistency of its derivation, even though none of the assumptions for its validity are completely satisfied. This formula was corrected [21.61], on account of the

band carrier's degeneracy, compensation and screening by carriers on impurity centers. The discrepancy between experiment and the BH formula, revealed during three decades of studies, have been thoroughly analyzed [21.62].

Modifications and developments made to overcome the drawbacks of the BH formula are worth mentioning. Taking the multi-valley band structure into account did not invalidate the relaxation-time method as such, but resulted in essentially different $\tau_l(E)$ and $\tau_t(E)$ [21.63, 64]. These formulas have also been discussed [21.65] in light of the scattering anisotropies measured in n-Si [21.66]. To overcome limitations of the Born approximation exact, although limited only to the standard band, phase-shift analysis was employed [21.67]. Including a non-Coulomb part of the impurity potential [21.68] made it possible to explain in part the difference in mobility of n-Si samples doped with different donors [21.25]. Lastly, Monte Carlo simulations of the impurity scattering, improving the agreement between theory and experiment for n-Si, have recently been reported [21.69]. The ionized-impurity scattering in p-Si was considered in the approximation of isotropic hole bands [21.70]. We are not aware of any theoretical papers on the subject, which used anisotropic energy spectra given by (21.26) in the case of p-Si.

Carrier–Carrier Scattering

Carrier–carrier scattering becomes important as n or p increases, along with increasing N_d^+ or N_a^-. The relaxation-time concept does not apply for this mechanism. Carrier–carrier collisions redistribute the carrier's energy in a chaotic manner that was presumed to cause a decrease in the net mobility due to other mechanisms [21.71]. For the standard band, the effects of electron–electron scattering were modeled using the variational method, which predicted about 30% reduction in the ionized-impurity scattering-limited mobility [21.72]; close results were obtained using another, quite different, method [21.73]. Hole–hole scattering and electron–hole scattering were also considered [21.74] in the standard band. Due to ignorance of the specific band-structure features, the results of these papers had limited relevance to Si. The effect of electron–electron scattering has been recast [21.75] for the multi-valley band structure using the generalized Drude approximation (GDA). At DC, the GDA corresponds to the zeroth-order approximation of the variational method, which highly overestimates [21.72] the effect considered.

Dielectric Constant

In Si the current carriers are well decoupled from the host electrons, so the Maxwell equations result in a unique decomposition of the dielectric constant

$$\varepsilon(\omega) = \varepsilon_L(\omega) + \varepsilon_C(\omega)\,,$$
$$\varepsilon_C(\omega) = \mathrm{i}\frac{4\pi\sigma(\omega)}{\omega}\,. \tag{21.20}$$

Here $\varepsilon_L(\omega)$ is the host contribution, which is indirectly influenced by the carriers. The direct effect of the carriers on the dielectric constant is the conductivity contribution $\varepsilon_C(\omega)$. As seen from (21.20), doped Si behaves at DC as a metal. The asymptote at high frequencies (IR for Si) is $\varepsilon_{C,s}(\omega) \approx -(\Omega_{pl,s}/\omega)^2$, where $s = $ e or h, and $\Omega_{pl,s}$ are the bare plasma frequencies given by

$$\Omega_{pl,e}^2 = \frac{4\pi q^2 n}{m_{ce}}\,, \quad \frac{1}{m_{ce}} = \frac{1}{3}\left\langle\frac{1}{m_l} + \frac{2}{m_t}\right\rangle,$$
$$\Omega_{pl,h}^2 = \frac{4\pi q^2 p}{m_{ch}}\,, \quad \frac{1}{m_{ch}} = \frac{1}{1+\beta}\left\langle\frac{1}{m_1} + \frac{\beta}{m_2}\right\rangle, \tag{21.21}$$

resulting from (21.18) and (21.20), irrespective of the scattering model. Due to this asymptote and (21.20) $\varepsilon_1(\omega)$ should become zero at some frequency $\omega_{pl,s}$ and doped Si should behave optically as a dielectric at $\omega < \omega_{pl,s}$ and as a metal at $\omega > \omega_{pl,s}$. True plasma frequencies are estimated roughly as $\omega_{pl,s} \approx \Omega_{pl,s}/n_L$, where $n_L = 3.42$–3.44 is the Si host refractive index in the IR. For the parabolic bands, m_{ce} has already been presented in Table 21.1, and $m_{ch} = (0.33$–$0.39)m_0$ using the data of Table 21.1, with the assumption that $m_{1,2} = m_{c1,2}$.

General formulas for $\varepsilon_C(\omega)$ are rather involved because of the averaging over E they contain, and so are rarely used. The Drude formula

$$\varepsilon_{C,s}(\omega)|_{\mathrm{Drude}} = \frac{\mathrm{i}}{\omega}\,\frac{\Omega_{pl,s}^2}{\gamma_s - \mathrm{i}\omega}\,, \tag{21.22}$$

where γ_s are adjustable phenomenological damping parameters, is often employed instead [21.76]. To match the behavior of $\varepsilon_C(\omega)$ at $\omega \to 0$ one should put $\gamma_s = 1/\tau_{0,s}$ in (21.22), where

$$\tau_{0,e} = \frac{m_{ce}}{3}\left\langle\frac{\tau_l}{m_l} + \frac{2\tau_t}{m_t}\right\rangle;$$
$$\tau_{0,h} = \frac{m_{ch}}{1+\beta}\left\langle\frac{\tau_1}{m_1} + \frac{\beta\tau_2}{m_2}\right\rangle \tag{21.23}$$

are the DC mobility relaxation times. Such an adjustment was shown to work poorly in n-Si [21.77]. On the

other hand, putting $\gamma_s = \gamma_{\infty,s}$ in (21.22), where

$$\gamma_{\infty,e} = \frac{m_{ce}}{3}\left\langle \frac{1}{m_l\tau_l} + \frac{2}{m_t\tau_t}\right\rangle;$$
$$\gamma_{\infty,h} = \frac{m_{ch}}{1+\beta}\left\langle \frac{1}{m_1\tau_1} + \frac{\beta}{m_2\tau_2}\right\rangle, \qquad (21.24)$$

allows one using the Drude formula to match *two* leading at $\omega \to \infty$ terms in the power series expansion of $\varepsilon_C(\omega)$ with respect to ω^{-1}. Thus (21.22) with the above adjustments may serve as an overall interpolation if the *high-frequency* relaxation time, $\tau_{\infty,s} = 1/\gamma_{\infty,s}$ turns out to be close to $\tau_{0,s}$. A Drude formula, empirically adjusted in IR has been devised for n-Si [21.78].

Using $\tau_{\infty,s}$ instead of $\tau_{0,s}$ in the mobility is a prerequisite for GDA at DC. As discussed above, the Boltzmann-equation-based formulas are valid in the range $\lambda \gg \lambda_q = hc/\bar{E}$. For nondegenerate carriers $\lambda_q(\text{cm}) \approx 1.4388/T$, while in Si for n or p up to $\approx 10^{20}\,\text{cm}^{-3}$ the carrier degeneracy (if present) stops much below 300 K, so the validity of the $\varepsilon_C(\omega)$ formulas considered is restricted to the far-IR and longer wavelengths. To properly describe $\varepsilon_C(\omega)$ in the near- and mid-IR range, ω-dependent GDA has been suggested [21.79]. In this approximation one replaces γ_s in (21.22) by an ω-dependent damping $\gamma_s(\omega)$, which is then determined by comparison of the first imaginary term in the expansion of the thus-generalized Drude formula, i. e., $(\Omega_{pl,s}/\omega)^2\gamma_s(\omega)/\omega$, with $\varepsilon_2(\omega)$ calculated using the methods of transition probabilities or perturbations for correlation functions.

In Si, unlike semiconductors with ionic bonds (e.g., $A^{III}B^{V}$), the elementary cell has no dipole moment and hence no quasi-classical optical-phonon contribution in $\varepsilon_L(\omega)$ is present. However, several weak IR absorption bands, attributed to two-phonon interaction of light with the Si lattice, are observed. Of these bands the most prominent is that peaked at 16.39 μm, which undergoes about a twofold increase in absorption upon increasing the temperature from 77 to 290 K. Comparable to that, the 9.03 μm absorption band, observed in pulled Si crystals, was attributed to Si−O bond stretching vibrations [21.80]. At $\lambda < 1.2\,\mu\text{m}$, where $\varepsilon_L(\omega)$ dominates the dielectric constant irrespective of the doping, $\varepsilon_{2L}(\omega)$ is accounted for by inter-band electronic transitions. The spectral bands of $\varepsilon_{2L}(\omega)$ correspond to the absorption of photons with energies close to the band gaps. Bare indirect-band-gap transitions, that necessitates lowest energy, is forbidden, but perturbation correction in the electron–phonon interaction to $\varepsilon_{2L}(\omega)$ suffices to describe the observed indirect absorption band. Due to the phonons the lowest fundamental absorption, although smaller than in direct bands, increases with increasing T. In principle, $\varepsilon_{2L}(\omega)$ may be calculated by the band-structure simulation route (e.g., [21.6, 7] and *Kleinmann* and *Phillips* [21.11]), especially with the present state-of-the-art theory – $\varepsilon_{1L}(\omega)$ is then calculated using the KKR (21.16). However, applications need fast modeling, and such formulas have been developed for Si [21.81–83]. There are a few distinct doping effects on $\varepsilon_L(\omega)$:

1. Effects on the lowest band edges, both direct and indirect: the Burstein–Moss shift with increasing $n(p)$ in heavily doped samples at the degeneracy due to filling of the conduction (valence) band below (above) E_F; band-gap shrinkage due to carrier–carrier and carrier–impurity interactions [21.84, 85] that work against the Burstein–Moss shift; and the formation of band tails because of the random potential of impurities [21.31].
2. Effects on higher edges, such as: E_1 (3.4 eV) – due to transitions between the highest valence band and the lowest conduction band along the Λ line in a region from $\pi/4a_0(1, 1, 1)$ to the L point on the BZ edge; and E_2 (4.25 eV) – due to transitions between the valence band at the X point and the conduction band at $2\pi/a_0(0.9, 0.1, 0.1)$ [21.7]. In this case [21.86], the electron–electron interaction plays a small role because carriers are located in a small region of the BZ, different from that where the transitions take place, and the effect of the electron–impurity interaction is calculated using standard perturbation theory.
3. Absorption due to direct inter-conduction-band (inter-valence-band) transitions specific to the type of doping. In n-Si this is a transition from the lowest conduction band to the band that lies higher at the Δ point but crosses the former at the X point, and which gives rise to a broad absorption band peaked around 0.54 eV [21.87] and tailing of the indirect gap at heavy doping [21.88]; the theory of this contribution to $\varepsilon_{2L}(\omega)$ has been developed [21.85]. In p-Si these transitions are those between the three highest valence bands [21.8]. Absorption due to the 1→2 transition has no energy threshold, and so resembles the usual free-carrier absorption; that due to the 1→3 and 2→3 transitions appears at $\hbar\omega = \Delta_{so}$. A high-energy threshold, above which the inter-valence-band absorption becomes negligible, exists due to the near congruency of all three valence sub-bands [21.8] at large well $\boldsymbol{k}$ (this is not accounted for by (21.5) and (21.6), which are valid at small $\boldsymbol{k}$). In contrast, with p-Ge, the manifestation of the inter-valence-band transitions in the

reflection, was proposed for p-Si [21.89], but fully reconciled later [21.85, 90]. Kane's theory of the inter-valence-band absorption [21.8] has also been revisited [21.90].

21.2 Electrical Properties

An extensive investigation of basic electrical properties was started 60 years ago, when polycrystalline Si containing B and P, was reported [21.91]. This seminal work was necessarily limited because neither single crystals nor the means for measuring below 77 K were then available. The first papers on single crystals were published five years afterwards. Since then, as techniques for fabricating quality single-crystalline silicon, such as the pulling, e.g., Czochralski (CZ), Teal–Little (TL), and floating-zone (FZ) techniques, became highly developed, many experiments on electrical properties have been published. A number of papers are considered below in historical retrospect. In the accompanying graphs additional, less cited, papers are referenced.

Though the physical mechanisms behind the electrical properties of crystalline Si have been studied and partially understood for a long time, the resulting formulas and procedures are too complicated and time-consuming to be used in electronics device modeling. In this connection, several useful, simplified but accurate, procedures for modeling mobility versus temperature, doping, injection level and electric field strength have been developed. For this issue we refer to items 2–4 in Sect. 21.2.3.

21.2.1 Ohm's Law Regime

Drift Measurements

1. The minority-carrier mobility as a function of N_d, N_a, T and ρ in n- and p-type samples is in the range 0.3–30 Ω cm [21.92]. The cited paper revealed for the first time the inapplicability, at least for holes, of the simple $T^{-1.5}$ lattice mobility law, and presented curves of ρ versus exhaustion concentration $N = |N_d - N_a|$ in the range $10^{14}\,\mathrm{cm}^{-3} \le N \le 10^{17}\,\mathrm{cm}^{-3}$.
2. Measurements of μ_e in p-type and μ_h in n-type Si on 11 single crystals ranging in ρ from 19 to 180 Ω cm [21.93]. In the purest crystals, in the range 160–400 K, μ_e and μ_h obeyed the dependencies $T^{-2.5\pm0.1}$ and $T^{-2.7\pm0.1}$, respectively. The conductivity of some of these crystals, measured from 78 to 400 K, provided independent evidence for the temperature dependencies of the mobility quoted above.
3. Room-temperature drift and conductivity plus Hall-effect measurements [21.94] of μ_e and μ_h versus resistivity on an unprecedentedly large number of samples cut from CZ crystals. The largest ρ was above 200 Ω cm. The values of μ_e and μ_h obtained from both experiments in the purest crystals were reported and compared with those obtained by other authors.

Resistivity and Galvanomagnetic Measurements

1. Room-temperature $\mu^{H}_{e(h)}$ as a function of $\rho_{e(h)}$ [21.95]. The crystals used were grown from Dupont hyper-pure material, with ρ of 0.01–94 Ω cm for n-type and 0.025–110 Ω cm for p-type samples. Curves of $\mu^{H}_{e(h)}$ versus $\rho_{e(h)}$ were calculated using the BH and combined-mobility [21.71] formulas for $m_{e(h)} = m_0$ and compared with experimental curves.
2. The first extensive experimental study of electrical conductivity and the Hall effect in TL silicon [21.96]. The properties were measured at temperatures of 10–1100 K on six arsenic-doped n-type samples, and one undoped, plus five boron-doped, p-type samples, covering the range from light ($N = 1.75\times10^{14}$ and $3.1\times10^{14}\,\mathrm{cm}^{-3}$) to heavy ($N = 2.7\times10^{19}$ and $1.5\times10^{19}\,\mathrm{cm}^{-3}$) doping. Compensation by unknown acceptors (donors) occurred in four lightly and moderately doped n(p)-type samples. A deviation of the lattice mobility from the $T^{-1.5}$ dependence was reported for both electrons and holes. Curves of $\mu^{H}_{e(h)}$ against $\rho_{e(h)}$ at 300 K were computed in the same way as by Debye and Kohane, but incomplete ionization of impurity centers was additionally taken into account.
3. First systematic study of μ_e and μ_h versus N_d and N_a, respectively, at $T = 300$ K [21.97]. Measurements were taken with several group V and group III impurities up to 6×10^{19} and $6\times10^{18}\,\mathrm{cm}^{-3}$ for n- and p-Si, respectively. Impurity concentrations were obtained by radioactive tracers or from thermal neutron activation analysis; μ_e and μ_h were calculated from these data by considering the N_d^+ and N_a^- percentages. The combination with measured $\mu^{H}_{e(h)}$ resulted in $r_{e(h)}$ values in agreement with theory. A comparison with a BH-formula-based theory yielded semiquantitative agreement for μ^{H}_{e}, while measured values of μ^{H}_{h} proved to be much smaller than the theoretical values.

4. Galvanomagnetic effects in p-Si: ρ and R^{H} versus T and B [21.98] and MR [21.99]. Boron-doped samples cut from CZ crystals were used. In the first paper four samples, two with $\rho(300\,\mathrm{K}) = 35\,\Omega\,\mathrm{cm}$ and two with $\rho(300\,\mathrm{K}) = 85\,\Omega\,\mathrm{cm}$, were measured in the range 77–320 K. The dependence $\mu_{\mathrm{h}} \propto T^{-2.7\pm0.1}$ at $B = 0$, as observed by Ludwig and Watters, was typical of the results obtained on all the samples; r_{h} was observed to exhibit a weak linear decrease with T in the range 200–320 K, and to be almost entirely independent of B up to $B = 1.3\,\mathrm{T}$ in the temperature interval studied. The dependence of MR on the relative directions of current, fields and crystallographic axes was studied at 77 and 300 K as a function of B. Large values of longitudinal MR, as large as the transverse effects in some cases, were observed, contradicting the only calculations available at that time [21.17]. To obtain data sufficient for constructing a more satisfactory model, the above study was continued in the second paper on 10 more samples, with ρ of 0.15–115 Ω cm. Measurements of three MR coefficients were carried out at a number of temperatures in the range 77–350 K. The results showed a marked dependence of the band structure and scattering anisotropies on the temperature, yet no definite model of these effects was arrived at.
5. A comparative study of mobility in pulled and FZ crystals [21.100]. The question of the dependence of the intrinsic mobility on temperature was recast. The authors found that in FZ, contrary to pulled crystals, $\mu_{\mathrm{e}}^{\mathrm{H}}$ followed the $T^{-1.5}$ law in the range 20–100 K, although $\mu_{\mathrm{h}}^{\mathrm{H}}$ still displayed a different, viz. T^{-2}, variation with temperature. It was argued that such a disagreement with the work of Morin and Maita was due to the large content, up to $10^{18}\,\mathrm{cm}^{-3}$, of oxygen impurities in the pulled crystals they used, which resulted in scattering that obscured the phonon scattering.
6. Solid analysis of electrical properties of n-Si with respect to: the ionized-impurity scattering in isotropic approximation [21.101], scattering anisotropies [21.66] and lattice scattering [21.51]. The measurements were made from 30 to 100–350 K using a set of P-doped, B-compensated, n-type samples of rather wide impurity content, yet in the range from light to moderate doping ($N_{\mathrm{d}} = 4.5\times10^{15}\,\mathrm{cm}^{-3}$ at most). The purest samples were cut from FZ crystals while others were from CZ crystals. These authors developed a sophisticated, but robust, method of determining N_{d} and N_{a} by analysis of the R^{H} versus T data. With this method, in the first paper they obtained curves of μ_{e} versus T in the range 30–100 K, which were used to test the BH formula. In comparing the formula with the data, correction to the observed μ_{e} because of the phonon-scattering contribution was necessary. The BH formula was shown to provide a good quantitative description of the data when $m_{\mathrm{e}} = 0.3m_0$ was used, provided that ion scattering was not too strong. When ion scattering was dominant, viz. in moderately doped samples at low T, they observed a discrepancy between the theory and data, which was attributed to electron–electron interaction. For the purpose of detecting scattering anisotropy, the MR coefficients were measured in the second paper on several relatively pure ($N_{\mathrm{d}} = 8.0\times10^{14}\,\mathrm{cm}^{-3}$ at most) samples. The results indicated that $\tau_{\mathrm{l}}/\tau_{\mathrm{t}} \approx 0.67$ and $\tau_{\mathrm{l}}/\tau_{\mathrm{t}} > 1$ for acoustic-phonon and ionized-impurity scattering, respectively. The inter-valley phonon scattering, important at higher T, proved to be isotropic. In the third paper lattice scattering was treated. A model assuming inter-valley scattering by two-phonon modes, in addition to the intra-valley acoustic-phonon scattering, was applied to the results of the electrical and MR measurements in the purest samples. By using the coupling constants with the phonon modes as fitting parameters, fairly good agreement was achieved between the model and experimental data on curves of μ_{e} versus T in the range 30–350 K, though one of the inter-valley phonon energies corresponded to a forbidden transition (Table 21.4).
7. Experimental verification of the anisotropic ion-scattering theory [21.102]. Measurements of saturated longitudinal MR and μ_{e} were made in fields up to 9 T at 78 K on a series of [111]-oriented phosphorus-doped samples with N_{d} in the range 2×10^{13}–$6\times10^{16}\,\mathrm{cm}^{-3}$. A quantitative analysis was made that involved combined relaxation times from impurity, intra-valley, and inter-valley lattice scattering. The data agreed very well with the theory [21.103]. In samples with $N_{\mathrm{d}} > 10^{16}\,\mathrm{cm}^{-3}$ the neutral-impurity scattering effect was observed.
8. Refined analysis of the Hall effect and mobility in n-Si [21.53]. μ_{e} and n, as determined from the high-field Hall effect, were numerically analyzed for a series of n-type samples doped with Sb, P, and As. The calculations were based on the general treatment for an anisotropic parabolic band [21.48]. Lattice-scattering parameters for the inter-valley and acoustic modes were determined from a comparison of the results between theory and experiment, using as many as four inter-valley phonons. The conclusions supported the two-phonon Long model. Ionized-impurity scattering was calculated from the BH formula and a newer theory [21.103] and compared with ex-

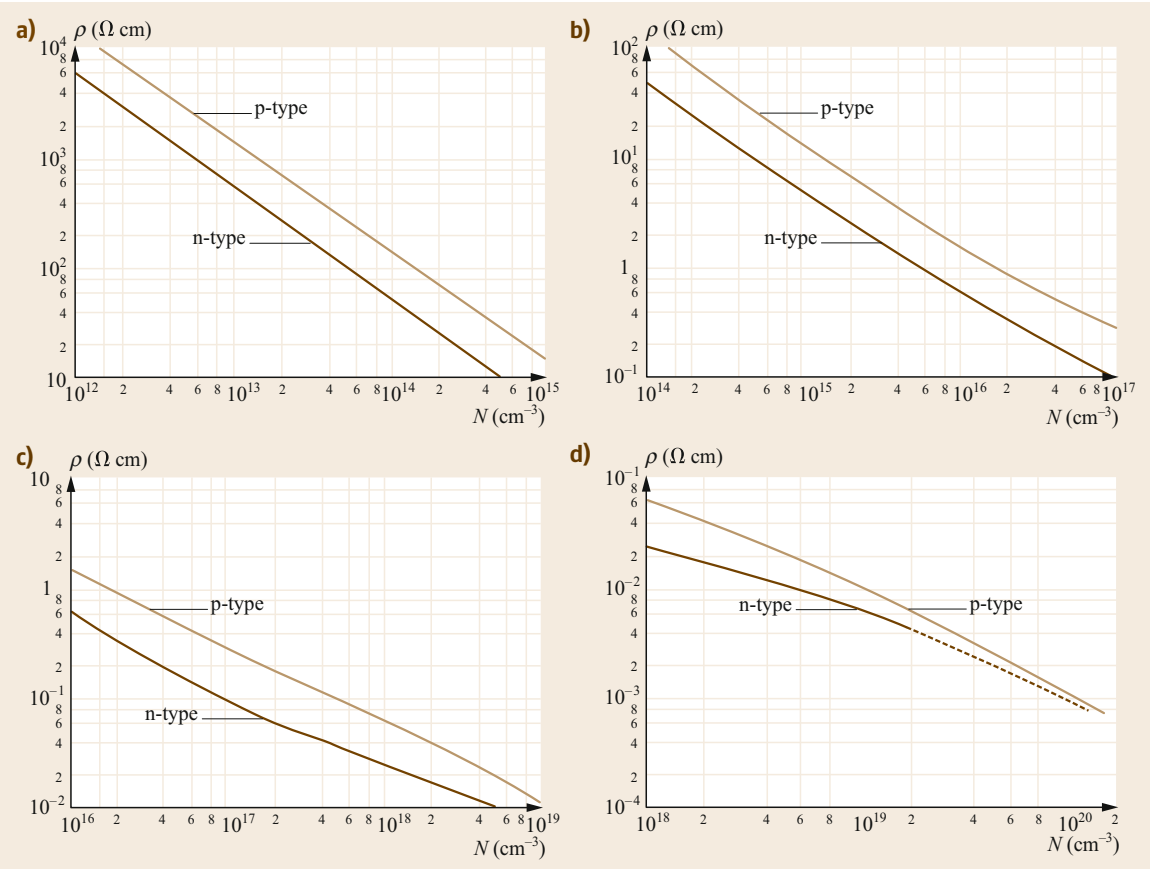


Fig. 21.5a–d Room-temperature resistivity of Si versus N (defined in the text) for overlapping ranges of doping: **(a)** light to moderate; **(b)** moderate; **(c)** moderate to heavy; **(d)** heavy. (After [21.24] with permission)

periment in favor of the latter. Neutral-impurity scattering proved to be temperature-dependent, unlike the available model [21.55].

9. The Hall factor in n-Si versus $N_I = N_d^+ + N_a^-$ [21.104]. The experimental value of r_e was obtained from the ratio of $R_e^H(0)$ to $R_e^H(\infty)$, determined by independent methods at 77 and 350 K over a wide range of N_I. Samples were cut from non-oxygenated ingots grown in vacuum, in which $R_e^H(0)$ was measured by the standard method (Sect. 21.2). Discs were then cut from the same samples, in proximity to Hall probes, and their total resistances (given by $BR_e^H(\infty)/d$ at $\mu_e B/c \gg 1$ (Sect. 21.2), where d is the disc width) were measured in fields up to $B = 3.5$ T. To check these results, standard measurements of R_e^H in pulsed fields (Sect. 21.2) up to $B = 15$ T were made. The values of r_e were calculated from theory [21.103]. Good agreement of experiment with theoretical calculations was obtained.

Minority Carriers – Miscellaneous

Minority carriers play a crucial role in silicon-based electronic devices such as bipolar junction transistors (BJTs) and solar cells, so their charge-transport parameters – concentration, recombination time, mobility and diffusion coefficient – are importany for device modeling. We do not present a literature overview nor graphical and tabular material here, as we do for the case of the majority carriers, and refer the reader directly to the review papers outlined in items 5 and 6 of Sect. 21.2.3.

Irvin Curves

The curves of $\rho(300\,\mathrm{K})$ versus N, over the entire range up to the solubility limit, compiled using published data [21.105, 106] are shown in Fig. 21.5a–d, with the high-doping range shown Fig. 21.6a–d. The resistivity decreases with increasing N since the conductivity is proportional to $n(p)$ while $n(p) = N$ at $T = 300$ K. However, at large N the decrease is slower than $\propto N^{-1}$

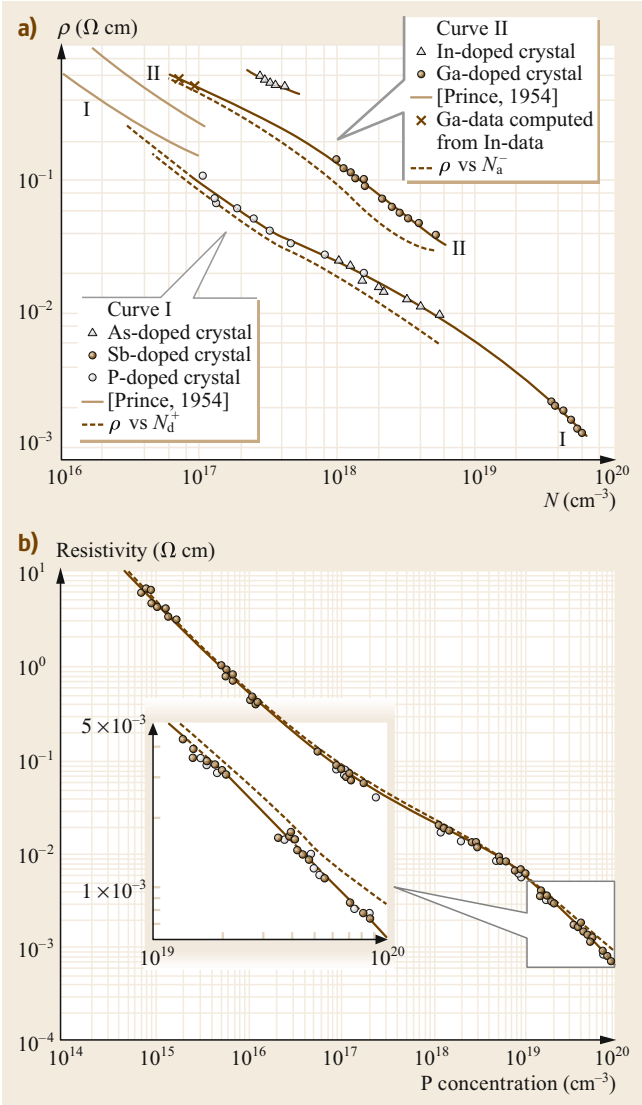


Fig. 21.6a,b Room-temperature resistivity of highly doped Si versus: **(a)** N in compensated n- and p-type crystals; the curves corrected for ionized impurities content are also presented. (After [21.97] with permission); **(b)** P content in uncompensated P-doped Si. *Dashed curve* is due to *Irvin* [21.105]. P concentration by: *solid circle* – neutron activation analysis, *open circle* – Hall-effect measurement. *Inset* – heavily doped range enlarged. (After [21.107] with permission)

due to the notable dependence of $\mu_{e(h)}$ on doping (see below).

Mobility and Hall Effect Versus Resistivity and Doping

Many studies have reported μ versus ρ graphically. Figure 21.7a–d summarize this information obtained

Table 21.6 Best fit parameters for the Caughey–Thomas formula (21.25)

Parameters	Electrons	Holes	Units
μ_{min}	92.0	47.7	$cm^2/(V\,s)$
μ_{max}	1360	495	$cm^2/(V\,s)$
N_{ref}	1.3×10^{17}	6.3×10^{16}	cm^{-3}
α	0.91	0.76	–

Table 21.7 Best fit parameters for (21.26), (21.31) and (21.32)

Parameters	Electrons	Holes	Units
μ_0	1400	480	$cm^2/(V\,s)$
N_{eff}	3×10^{16}	4×10^{16}	cm^{-3}
S	350	81	–
E_1	3.5×10^{3}	6.1×10^{3}	V/cm
E_2	7.4×10^{3}	2.5×10^{3}	V/cm
F	8.8	1.6	–
v_m	$1.53\times10^{9}T^{-0.87}$	$1.62\times10^{8}T^{-0.87}$	cm/s
E_c	$1.01T^{1.55}$	$1.01T^{1.68}$	V/cm
β	$2.57\times10^{-2}T^{0.66}$	$0.46T^{0.17}$	–

before 1965. For the Hall measurements the drift mobility was recalculated from the Hall value using the Hall factor for the standard band (Sect. 21.1.3). The dependencies of $\mu_{e(h)}$ on N, $N_{d(a)}$ and on N_I at low T are more physical and allow comparison with available theories. Examples of such dependencies obtained by different authors are presented in Fig. 21.8a–e. Several phenomenological expressions for $\mu_{e(h)}(N, 300\,K)$ have been given in the literature (Sect. 21.2.3) in agreement with experimental data on charge transport in Si. We consider here two of them, for obvious reasons of space. The first [21.108] reads

$$\mu(N_i) = \mu_{min} + \frac{\mu_{max} - \mu_{min}}{1 + (N_i/N_{ref})^{\alpha}}\,, \qquad (21.25)$$

where $i = a, d$, and the other fitting parameters are given in Table 21.6, the electron column of which corresponds to a later best fit for Si:P [21.109], while the hole column is due to *Caughey* and *Thomas* [21.108]. The particular case of (21.25) with $\mu_{min} = 0$, $\alpha = 0.5$, $N_{ref} = 10^{17}\,cm^{-3}$ [21.111] has also proved useful. The second expression for Ohm's mobility versus impurity concentration,

$$\mu(N_i) = \frac{\mu_0}{\sqrt{1 + \frac{(N_i/N_{eff})S}{N_i/N_{eff}+S}}}\,, \qquad (21.26)$$

where N_{eff} and S are fitting parameters, follows from the weak electric field limit of the more general phenomenological expression (21.31) for v_d versus E and N_i. The best-fit parameters for (21.26) can be found

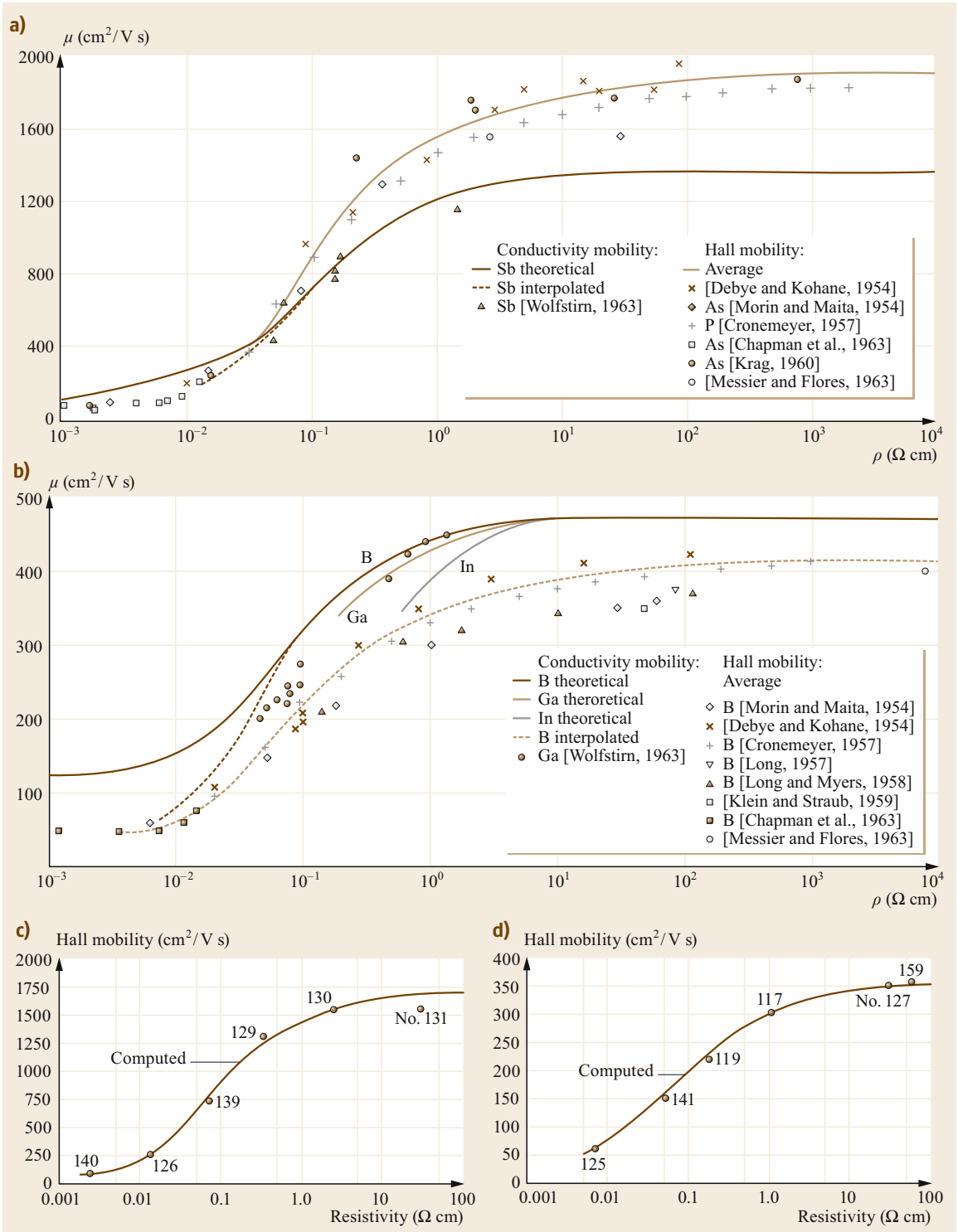


Fig. 21.7 Mobility versus resistivity in Si at room temperature compiled from different authors: **(a)** Electron mobility. (After [21.24] with permission); **(b)** Hole mobility. (After [21.24] with permission); **(c)** Hall mobility of electrons in Si:As. (After [21.96] with permission); **(d)** Hall mobility of holes in Si:B. (After [21.96] with permission). Numbers on graphs **(c)** and **(d)** denote samples, details of which are given in Table 21.10

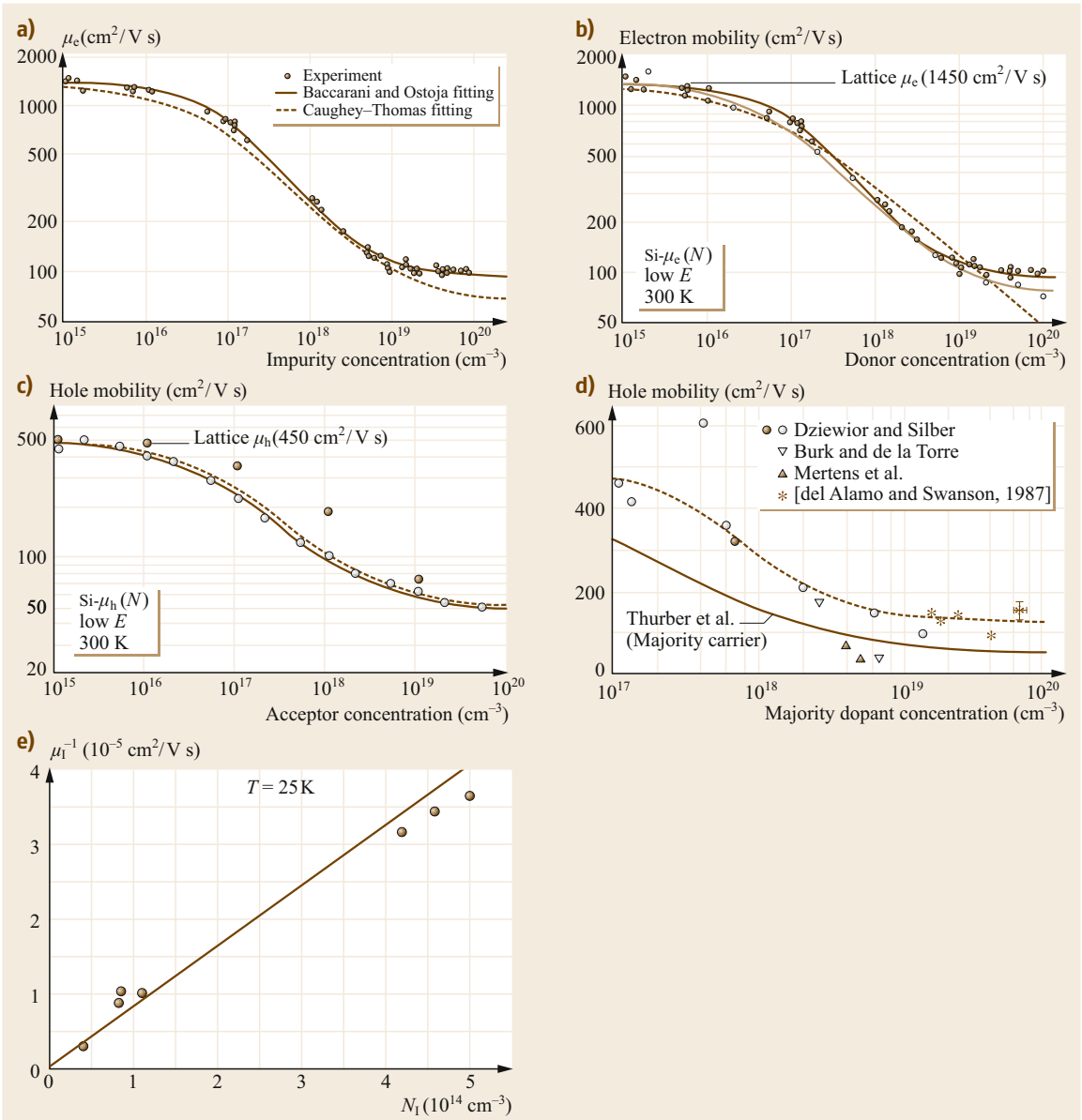


Fig. 21.8a–e Mobility versus doping in Si: **(a)** Room-temperature electron mobility in n-type Si: P versus P concentration, *solid square* – experimental data [21.107], *full line* – the best fit due to Baccarani and Ostoja ((21.25), Table 21.6) and *dashed line* – fitting of Irvin's data by (21.25) due to Caughey and Thomas [21.109], with permission; **(b)** Summary of data on and fits to room-temperature electron mobility versus donor concentration in n-Si: *open circle* – compiled experimental data [21.105], *solid circle* – experimental data [21.107], *full line* – the best fit due to Baccarani and Ostoja ((21.25), Table 21.6), *broken line* – the best fit by (21.25) due to Hilsum and *dash-dotted line* – the best fit due to Scharfetter and Gummel ((21.26), Table 21.9) [21.19].; **(c)** Summary of data on and fits to room-temperature hole mobility versus acceptor concentration in p-Si, *open circle* – compiled experimental data [21.105], *solid circle* – lattice mobility, *full line* – fitting of Irvin's data by (21.25) due to Caughey and Thomas, and *dash-dotted line* – the best fit due to Scharfetter and Gummel ((21.26), Table 21.9) [21.19] with permission; **(d)** Mobility of holes as being minority carrier (*symbols* and *dashed line*), measured on devices by different authors indicated in captions, and majority carrier (*full line*) due to Thurber et al. (After [21.110] with permission, the references are found in this review); **(e)** Inverse ionized-scattering mobility in n-Si at $T = 25$ K versus N_I. (After [21.100] with permission)

Table 21.8 Intrinsic mobility in crystalline Si at room temperature

Carriers	Hall mobility ($cm^2/(V\,s)$)			Drift mobility ($cm^2/(V\,s)$)			
Electrons	1610[a]	1450[b]	1560[c]	1500[d]	1610[a]	1350[e]	1360[c]
Holes	365	298	345	500	360	480	510

[a] After [21.95], [b] after [21.96], [c] after [21.94], [d] after [21.92], [e] after [21.93]

Table 21.9 Room-temperature mobility of Si at $n(p) = 2\times 10^{18}\,cm^{-3}$. (After [21.112])

	Donor		Acceptor	
Impurity	Sb	As	B	Ga
Ionization energy (eV)	0.039	0.049	0.045	0.065
Mobility ($cm^2/(V\,s)$)	235	220	110	100

in Table 21.7. Fits to (21.25) and (21.26) are used in Fig. 21.8a–c. Physically, μ_{max} in (21.25) and μ_0 in (21.26) are $\mu_{e,hL}(300\,K)$, the intrinsic mobility (see below) at room temperature. Experimental data on this parameter due to different authors are presented in Table 21.8. Figure 21.8d shows a comparison between the mobility of holes, as minority and majority carriers, plotted against the concentration of majority impurities. At the same concentration the minority-hole mobility is strikingly larger than the majority-hole mobility. This effect may be due to the difference between scattering by attractive and repulsive centers as well as to hole–hole interaction. The variation of $\mu(N, 300\,K)$ with shallow donor and acceptor species, noted in Sect. 21.1.4, is shown in Table 21.9. As can be seen, the effect – the larger $E_{d(a)}$ the lower $\mu_{e(h)}$ – is not drastic but quite noticeable. At low temperatures the impurity-scattering-limited mobility depends on the density of ionized centers, as predicted by the BH formula (or its modifications), e.g., for electrons

$$\mu_{eI} = \frac{\eta T^{3/2}}{N_I(\ln b - 1)}\,,$$
$$\eta = \frac{2^{7/2}k_B^{3/2}\varepsilon'^2}{\pi^{3/2}q^3 m^{*1/2}}\,, \qquad (21.27)$$

where m^* is an adjusted effective mass, $\varepsilon' = \varepsilon(0)$ is the static dielectric constant and

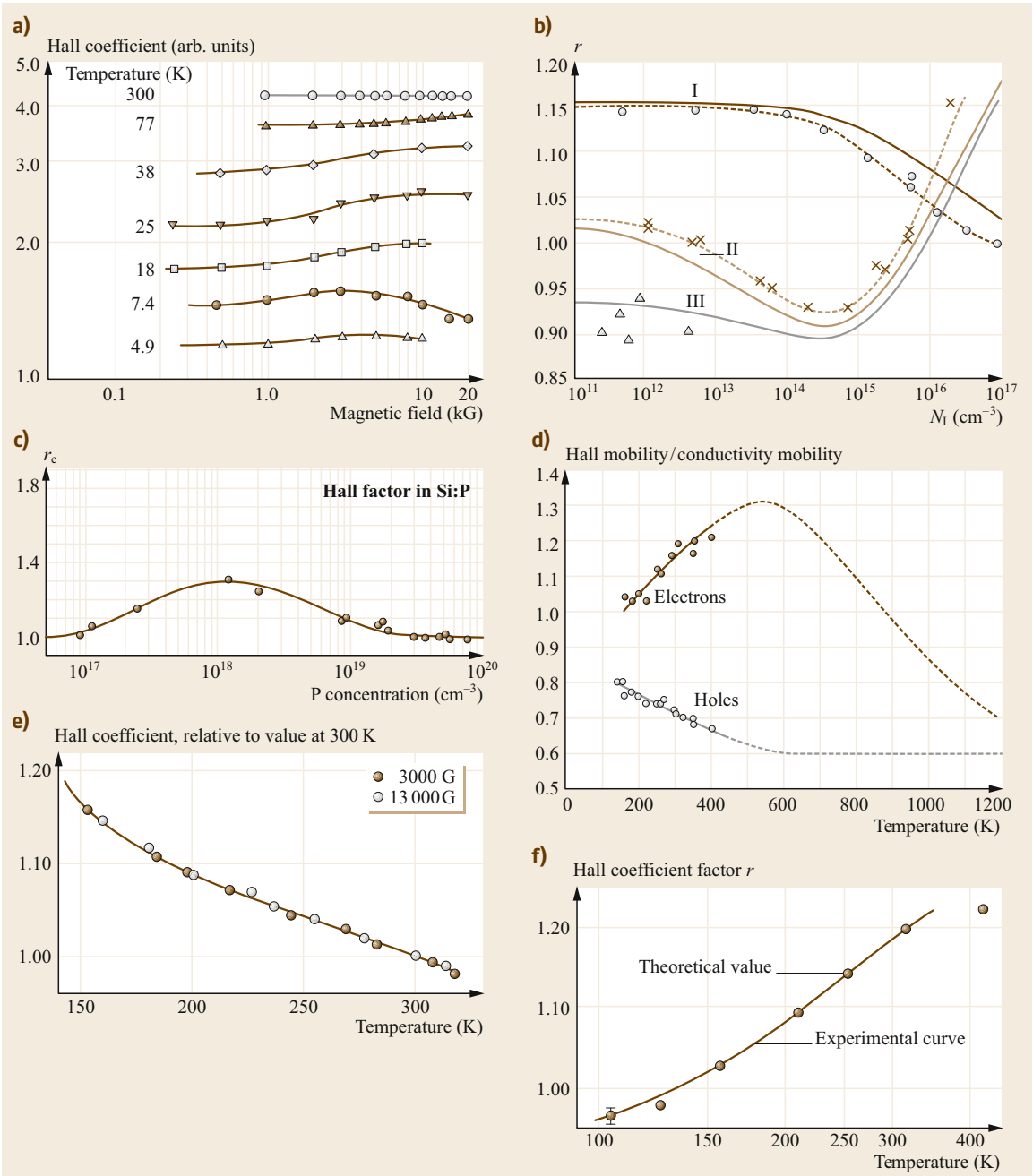
$$b = \frac{24\pi m\varepsilon'(k_B T)^2}{n' q^2 h^2}\,,$$
$$n' = n + (n + N_a)\left(1 - \frac{n + N_a}{N_d}\right). \qquad (21.28)$$

For compensated n-Si $n \ll N_a$ at sufficiently low T, so $N_I = 2N_a$ and $n' = N_a(N/N_d) = N_I(1-K)/2$, where K is the compensation ratio. Thus, disregarding the weak logarithmic dependence, μ_{eI}^{-1} proves to be linear with N_I. Such dependence is seen in the experimental data shown in Fig. 21.8e.

The Hall coefficient for lightly compensated n-Si as a function of B is shown in Fig. 21.9a. The curve for $T = 300\,K$ shows no visible variation when sweeping from low to high fields, which means that either the high-field regime is not attained or $r_e = 1$. The transition between the two regimes is clearly seen with $r_e < 1$, at the lower indicated values of T. The data below 20 K were obtained by sweeps of the photo-Hall coefficient, so the decrease of R_e^H at high B is fictitious, due to the field dependence of the photo-carriers lifetime [21.53]. Figure 21.9b,c display the dependence of r_e on the impurity concentration for light to moderate and high doping, respectively. Figure 21.9a,b are in qualitative agreement with the inequality $r_e < 1$ at

Table 21.10 Si samples. (After [21.96])

Sample number	Impurity	E_d or E_a (eV)	N (cm^{-3})	N_a or N_d (cm^{-3})	m^*/m_0
n-type					
131	As	0.056	1.75×10^{14}	1.00×10^{14}	0.5
130	As	0.049	2.10×10^{15}	5.25×10^{14}	1.0
129	As	0.048	1.75×10^{16}	1.48×10^{15}	1.2
139	As	0.046	1.30×10^{17}	2.20×10^{15}	1.0
126	As	?	2.00×10^{18}	–	–
140	As	Degenerate	2.70×10^{19}	–	–
p-type					
159	B	0.045	3.10×10^{14}	4.10×10^{14}	0.4
127	B	0.045	7.00×10^{14}	2.20×10^{14}	0.4
117	B	0.043	2.40×10^{16}	2.30×10^{15}	0.6
119	B	0.043	2.00×10^{17}	4.90×10^{15}	0.7
141	B	?	1.00×10^{18}	–	–
125	B	Degenerate	1.50×10^{19}	–	–



$T = 77\,\mathrm{K}$ for comparable concentrations. The overall observed behavior of $r_e(N)$ at $T = 300\,\mathrm{K}$, as shown in Fig. 21.9c, viz. $r_e \approx 1$ on both the moderate and heavy doping sides and a maximum in between at which $r_e > 1$, is deduced from (21.23) using knowledge of the anisotropy of the mass and relaxation-time tensors. Compared to these, results for the Hall effect for holes are not available, to our knowledge.

Hall Effect and Mobility Versus Temperature

The dependencies of $r_e(T)$ and $r_h(T)$ for samples 131, 130 and 159, 127 from Table 21.10 are shown in Fig. 21.9d. The temperature dependence of R_h^H relative to its value at $T = 300\,\mathrm{K}$ for another p-type sample is shown in Fig. 21.9e. Since, in the considered temperature range, p is constant with T, the dependence is fully congruent to $r_h(T)$ – in this regard the lower curve in

Fig. 21.9a–f Hall-effect parameters in Si: **(a)** Hall coefficient in n-Si versus magnetic field strength as measured for Si:As1 (Table 21.13) at several indicated temperatures. (After [21.53] with permission); **(b)** Hall factor in n-Si versus N_I, *cross* and *open circle* – experimental data at $T = 77$ and 300 K, respectively, *open triangle* – experimental data at $T = 77$ K for samples with lowest μ_e (supposedly compensated), *dashed lines* – guides for the eye, *full lines* are the dependencies calculated using theory [21.103] at $T = 77$ K (I) and 300 K (II) without account of neutral-impurity scattering, and at $T = 77$ K on account of neutral impurities with the density 6×10^{16} cm^{-3} (III). (After [21.104] with permission); **(c)** Hall factor in highly doped n-Si:P at $T = 300$ K versus phosphorous concentration [21.107]; **(d)** Hall factors for electrons and holes versus T, *solid circle* – measured, *dashed line* – computed dependencies. (After [21.96] with permission); **(e)** Hall coefficient, relative to its value at 300 K, versus T at two indicated strengths of magnetic field in a sample of p-Si with room-temperature $\rho = 35\,\Omega$ cm. (After [21.98] with permission); **(f)** Electron Hall factor versus T in n-type (SP6A, Table 21.12) sample, *full line* – experimental data, *open circle* – theory. (After [21.51] with permission) ◀

Fig. 21.9d and the curve in Fig. 21.9e are in agreement. There is good quantitative agreement between experiment and theory for $r_e(T)$ using an exact effective-mass tensor, as shown in Fig. 21.9f. On the contrary, the Hall effect and MR in p-Si are poorly understood [21.98, 99] as the energy spectra of holes is much more complicated than (21.5) and (21.6), even for energies well above 200 K [21.8].

The $\mu_{e,h}(T)$ curves have been the subject of extensive, but to our opinion not exhaustive, studies. As discussed in Sect. 21.2, these dependencies reflect the scattering mechanisms. The basic question is that of temperature dependence due to lattice scattering, which can be measured on technologically pure, very lightly doped and negligibly compensated samples. Since such samples were not available in early studies, researchers tried to apportion the intrinsic $\mu_{e,h}(T)$ from the drift measurements, as shown in Fig. 21.10a, or from conductivity and Hall measurements combined with calculations, as shown in Fig. 21.10b. In practice the intrinsic $\mu_{e,h}(T)$ is described by the phenomenological best-fit equation

$$\mu_{e,hL}(T) = AT^{-\gamma} \,, \tag{21.29}$$

where A and γ may vary with the carrier and measurement type, fitting interval and sample quality. Different drift measurements, in the ranges 150–300 K [21.92], and 160–400 K [21.93], resulted in very different values of A and γ for both electrons and holes. The first measurements of the Hall effect and conductivity against temperature [21.96] gave the same γ for holes as that obtained by Prince, but a drastically larger γ for electrons, as the n-type samples with the lowest N_d used were inappropriate for that purpose – one sample (131, Table 21.10), though lightly doped, was compensated ($K \approx 0.57$) and the other (130, Table 21.10) was doped in excess of 10^{15} cm^{-3}. Therefore, Morin and Maita numerically extracted the ionized-impurity scattering mobility portion (21.27) and (21.28) and then determined γ. This procedure seems unreliable in view of the inapplicability of the BH formula with strong compensation, neglect of scattering by neutral impurities and of the spectrum and scattering anisotropies. Analogously, using (21.29) for a high-resistivity p-type Si sample [21.99], fitting to $\mu_h(T)$ resulted in the same γ as that obtained by Ludwig and Waters while fitting to $\mu_h^H(T)$ (Fig. 21.10e) resulted in an even larger γ. Using high-purity samples, for which the $\mu_e(T)$ and $\mu_h(T)$ values are presented in Fig. 21.10c, made it possible to determine a γ for $\mu_e(T)$ in agreement with that obtained by Prince, but failed to fit $\mu_h(T)$ to (21.29) [21.100]. *Long* did not fit using (21.29) and calculated $\mu_{eL}(T)$ using a realistic electron spectrum and a specific inter-valley electron–phonon scattering model (Sect. 21.1.4), and compared the calculation to experiment [21.51], with excellent agreement, as can be seen in Fig. 21.10d. The validity of the model [21.51] was confirmed by an independent study [21.53]. The calculation of $\mu_{eL}(T)$ was revisited on the basis of the Monte Carlo method and extended inter-valley scattering models [21.113]; the results are displayed in Fig. 21.11a along with the experimental data and fit line. We are unaware of analysis for $\mu_{hL}(T)$ comparable to that done for $\mu_{eL}(T)$, taking both the realistic hole spectrum [21.8] and full hole–phonon scattering into consideration. Another analysis [21.70] greatly simplified the problem by using relaxation-time equalization ($\tau_1(E) = \tau_2(E)$), resulting from the isotropic approximation adopted for the hole spectra (21.5), but proved insufficient to account for both Ohm's mobility and hot-holes phenomena [21.18]. The results of calculations based on the relaxation-time method, using only one isotropic but nonparabolic heavy-hole band [21.18], are displayed in Fig. 21.11b along with the experimental data and fit line. The best-fit parameters of (21.29) obtained by different authors are given in Table 21.14.

The effect of doping on mobility is shown in Fig. 21.12. In doped samples $\mu_{e,h}(T)$ increases with decreasing T up a maximum at some T_{max} (the lower the sample's purity the larger T_{max}) and then starts to decrease. This behavior is due to competition be-

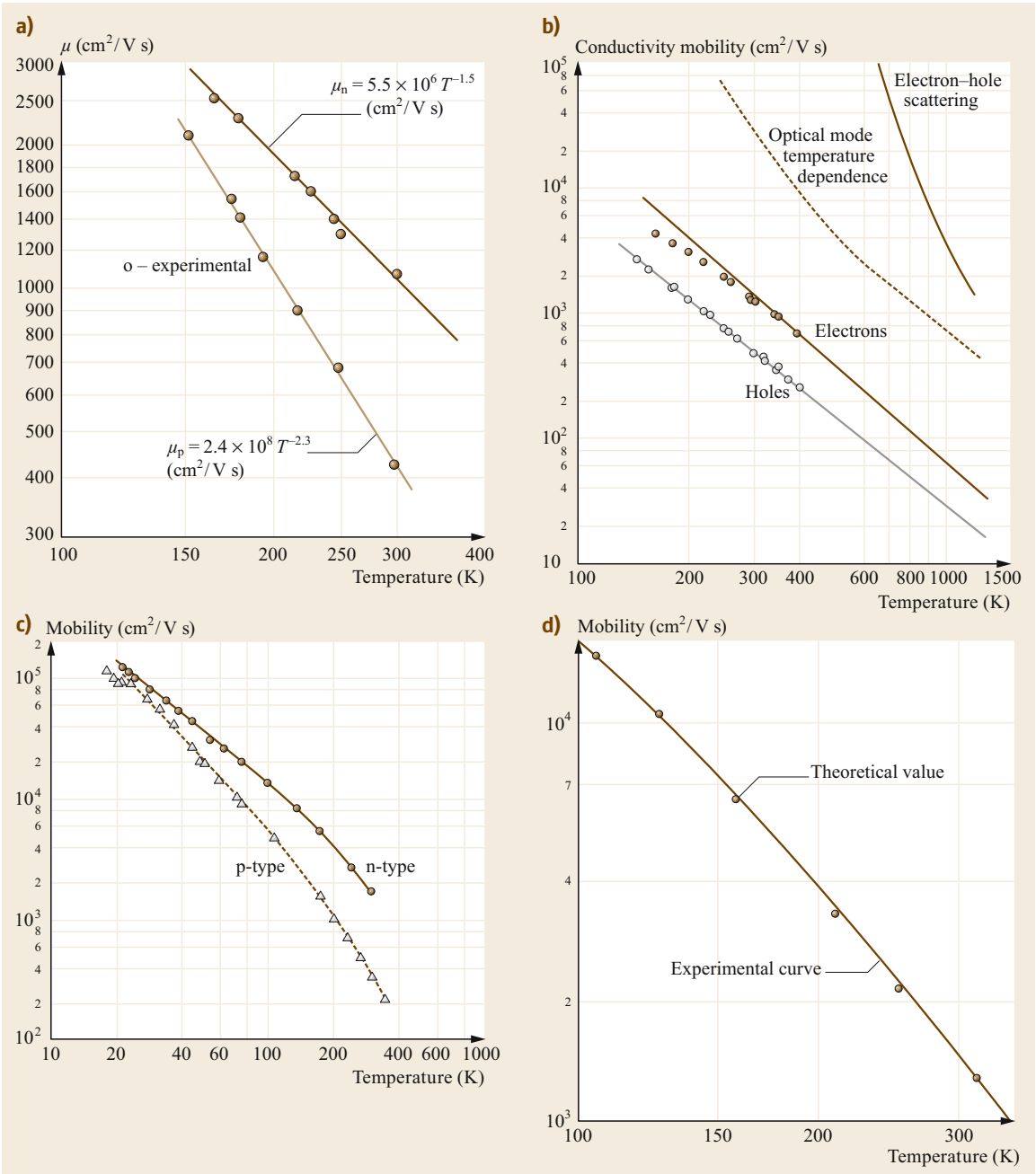


Fig. 21.10a–e Apportionment of dependence of lattice mobility on T in Si: **(a)** Minority electron and hole drift mobility in high-resistivity crystals. (After [21.92] with permission); **(b)** Conductivity mobility in TL crystals of n-type (130, 131, Table 21.10) and p-type (127, 159, Table 21.10). (After [21.96] with permission); **(c)** Conductivity mobility in FZ crystals of n-type (*A*, Table 21.11) and of p-type (*F*, Table 21.11). (After [21.100] with permission); **(d)** Conductivity mobility in FZ crystal of n-type (SP6A, Table 21.12), *full line* – experimental data, *open circle* – theory. (After [21.51] with permission); **(e)** Hall mobility in a CZ crystal of p-type with the indicated resistivity. (After [21.99] with permission)

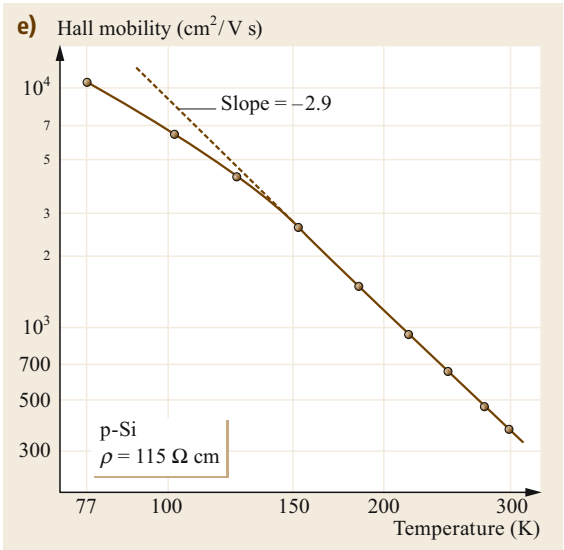


Fig. 21.10 (continued)

tween lattice scattering, which weakens with decreasing T (21.29), and impurity scattering, which strengthens with decreasing T, (21.27) and (21.28). In heavily doped samples scattering by impurities, while dominating the mobility, ceases to depend notably on T ((21.27) and (21.28) do not hold in this case), which results in an overall weak dependence of the observed $\mu_{e,h}(T)$ on T. These types of behavior are clearly seen in Fig. 21.12a,b. The evolution of T_{max} with the impurity content for lightly doped, differently compensated, n-Si is readily traced in Fig. 21.12c. The compensation is crucial for the rate of downturn of the mobility, as follows from (21.27) and (21.28) and is confirmed by the curves for samples C and D (Table 21.11) in Fig. 21.12c. Figure 21.12d summarizes the information on $\mu_{e,h}(T)$ for doped Si. For estimations of the combined-scattering mobility, e.g., for electrons, the Mattisen rule $\mu_e^{-1}(T) = \mu_{eL}^{-1} + \mu_{eI}^{-1}$, where μ_{eL} and μ_{eI} are given by (21.29) and (21.27) and (21.28), respectively, is used [21.100, 101]. The formula, consistently derived for nondegenerate carriers in the standard band, is

$$\mu_e(T) = \mu_{eL} + \mu_{eL}x^2 \times \left[\mathrm{Ci}(x)\cos x + \left(\mathrm{Si}(x) - \frac{\pi}{2}\right)\sin x\right], \quad x = \sqrt{6\frac{\mu_{eL}}{\mu_{eI}}}, \tag{21.30}$$

where $\mathrm{Ci}(x)$ and $\mathrm{Si}(x)$ are integral cosine and sine, respectively, and μ_{eL} is due to acoustic-phonon scat-

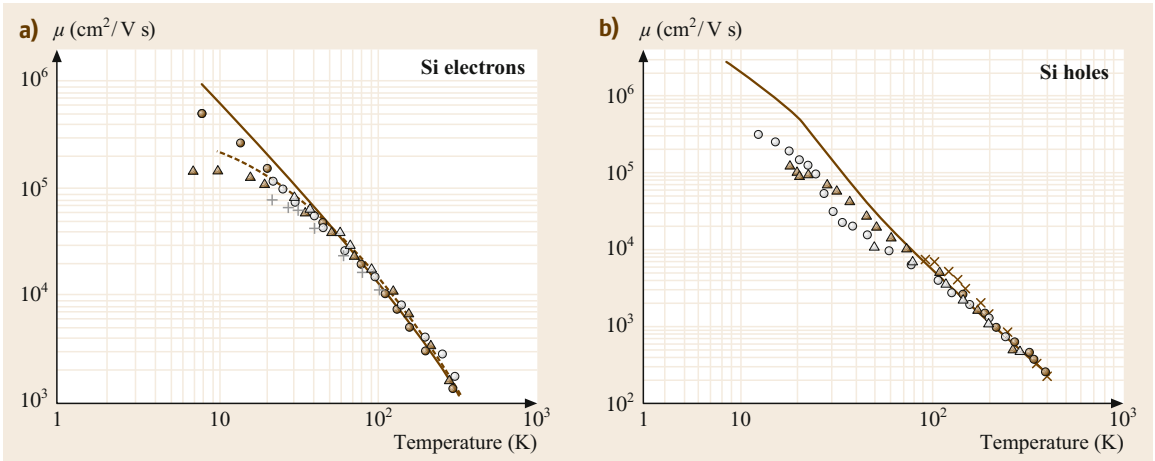


Fig. 21.11a,b Summary of data on μ versus T in nearly intrinsic to lightly doped Si; *symbols* and *lines* indicate experimental data and simulations, respectively: **(a)** n-type. *Solid circle* – time-of-flight technique on high-purity n-Si (Ohm's regime achieved only at $T \geq 45$ K) by Canali et al., *plus* – [21.114], *open triangle* – [21.101], *open circle* – [21.100], *solid triangle* – [21.53], *dashed* and *full lines* – Monte Carlo method by Canali et al. using three allowed and six lowest inter-valley phonon modes (Table 21.4), respectively, and no impurities, *dash-dotted line* – the same as full, but with inclusion of 10^{13} cm^{-3} ionized donors. (After [21.115] with permission); **(b)** p-type. *Open circle* – time-of-flight technique on high-purity p-Si (Ohm's regime achieved only at $T \geq 100$ K) by Ottaviani et al., *solid triangle* – [21.100], *cross* – minority hole drift mobility [21.93], *open triangle* – [21.96], *solid circle* – [21.96], *full line* – relaxation time method by Ottaviani et al. with the use of simplified models of the hole energy spectrum and hole–phonon interactions. (After [21.18] with permission)

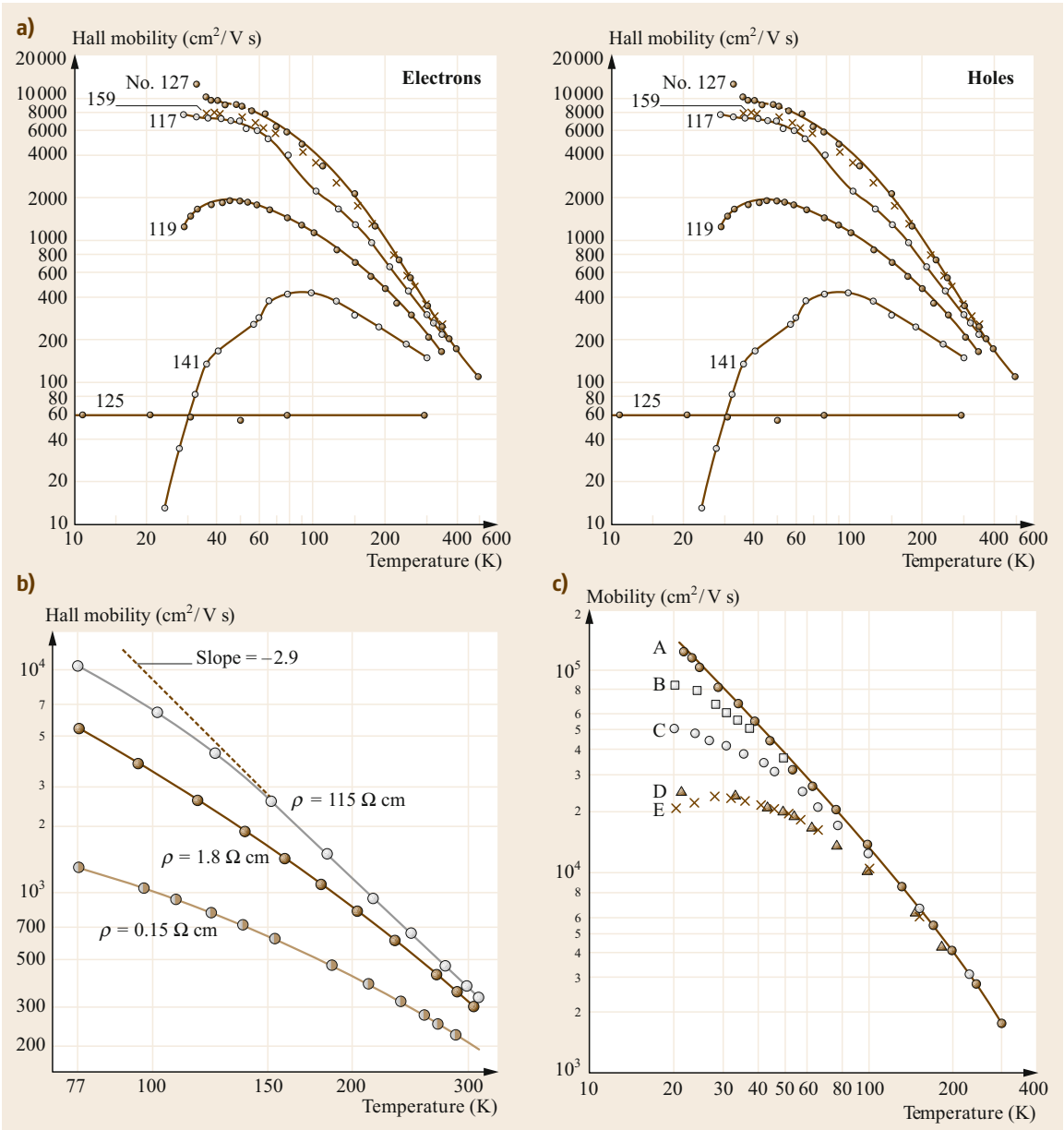


Fig. 21.12a–d Effects of doping on mobility versus T curves in Si: **(a)** Conductivity mobility of differently doped n- and p-type TL crystals. Numbers indicate samples from Table 21.10. (After [21.96] with permission); **(b)** Hall mobility of p-type CZ crystals with different ρ_{300}. (After [21.99] with permission); **(c)** Conductivity mobility versus T of lightly doped, differently compensated, n-Si samples from Table 21.11. (After [21.100] with permission); **(d)** Summary of data on μ versus T; *symbols* and *lines* indicate experimental data and simulations, respectively. Electrons: *solid circle* – high purity Si ($N \leq 10^{12}$ cm^{-3}, time-of-flight) [21.115], *solid triangle* – lightly doped compensated (Si: P 6, Table 21.13, photo–Hall effect) [21.53], *open square* – moderately doped with $K \approx 0.01$ (139, Table 21.10, Hall effect) [21.96], *full line* indicates the theoretical results for the lattice mobility [21.115], *dot-dashed line* gives the best fit of that mobility by an inverse power of T around room temperature ((21.29), Table 21.9); Holes: *solid circle* – high purity Si ($N \leq 10^{12}$ cm^{-3}, time-of-flight) [21.18], *open triangle* – lightly doped with $K \approx 0.01$ (F, Table 21.11, Hall effect) [21.100], *open square* – moderately doped with $K \approx 0.1$ (119, Table 21.10, Hall effect) [21.96], *full line* indicates the theoretical results for the lattice mobility, *dot-dashed line* gives the best fit of that mobility by an inverse power of T around room temperature ((21.29), Table 21.9) [21.18]. (After [21.19] with permission)

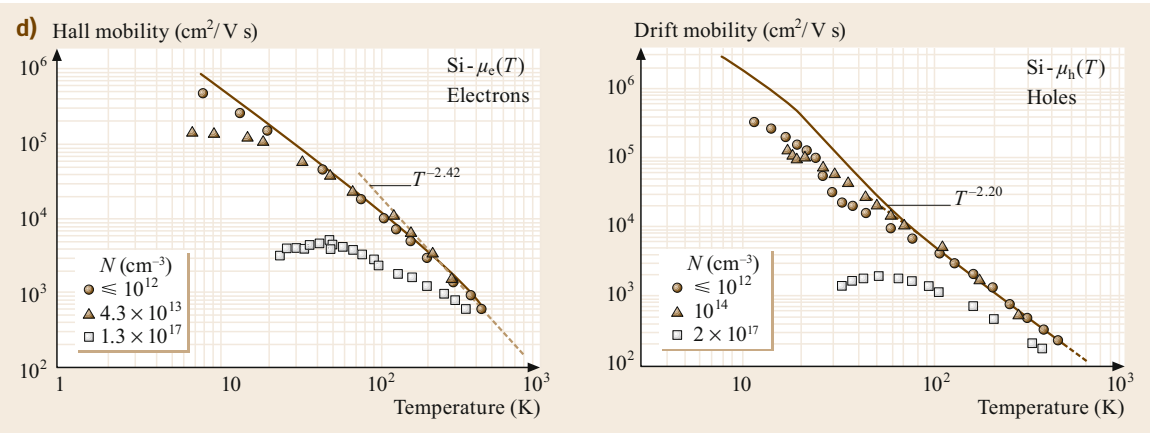


Fig. 21.12 (continued)

Table 21.11 Si samples [21.100]

Sample	Impurity	N_d (cm^{-3})	N_a (cm^{-3})	N_{Oxygen} (cm^{-3})
A (FZ)	P	1.14×10^{14}	4.00×10^{12}	$\approx 10^{16}$
B (FZ)	P	9.00×10^{13}	2.00×10^{13}	$\approx 10^{16}$
C (CZ)	P	2.50×10^{14}	5.50×10^{13}	$\approx 10^{18}$
D (CZ)	P	4.90×10^{14}	2.10×10^{14}	5.0×10^{17}
E (CZ)	P	3.30×10^{14}	2.30×10^{14}	7.7×10^{17}
F (FZ)	B	2.00×10^{12}	3.40×10^{14}	$\approx 10^{16}$

Table 21.12 Compensated Si:P samples [21.51, 66]

Sample	N_d (cm^{-3})	N_a (cm^{-3})
SP6X	1.7×10^{13}	8.0×10^{12}
SP6S	2.1×10^{13}	8.0×10^{12}
SP6A	2.8×10^{13}	7.5×10^{12}
SP8A	5.4×10^{13}	1.0×10^{13}
SP4M	1.3×10^{14}	1.0×10^{13}
SP4A	2.4×10^{14}	1.0×10^{13}
SP1A	6.7×10^{14}	3.8×10^{14}
SP1D	8.0×10^{14}	4.0×10^{14}
SP2A	2.3×10^{15}	8.3×10^{14}
SM 2	3.9×10^{15}	3.3×10^{15}
SM 3	4.5×10^{15}	3.3×10^{15}

tering (with $\gamma = 1.5$) [21.71], which proved more appropriate for quantitative analysis. The comparison of (21.30) and the expression for $\mu_e^H(T)$, derived under the same conditions, with the corresponding experimental curves, using $\eta = 8.6\times10^{17}\,\mathrm{K}^{-1.5}\,\mathrm{cm}^2\,\mathrm{V}^{-2}\,\mathrm{s}^{-1}$ (21.27) for an effective isotropic band mass of $m^* = 0.3m_0$, is shown in Fig. 21.13. Good quantitative agreement between the theory and experiment is observed except for the samples SM3 and SM2, in which the ionized-impurity scattering is strongest (to avoid confusion the data points for SM2 are not shown but they are as discrepant with respect to the observed mobility curve as those shown for SM3). The generalization of (21.30) to the case with spectrum and scattering anisotropies [21.103] was employed to calculate $r_e(T)$ [21.65] and to extract N_a from the mobility data [21.53]. Dakhoskii numerically explained why the above isotropic approximation [21.101] worked well, but presented no graphs of mobility versus T. Norton et al. presented graphically calculated $\mu_e(T)$ only for Si:P6, the purest of their samples. No analysis for $\mu_h(N,T)$ comparable to that done for $\mu_e(N,T)$, i. e., taking both the realistic hole spectrum and hole–phonon and hole–impurity scatterings into account, exists to our knowledge in the literature. Success of modeling based on empirically adjusted phenomenological expressions (items 2–4 in Sect. 21.2.3), though valuable for the device community, cannot be considered as such from the viewpoint of solid-state and semiconductors theories.

21.2.2 High-Electric-Field Effects

The group of effects associated with the carrier transport in high electric fields is traditionally called *hot-carrier phenomena*. This name was borrowed from the displaced-Maxwell-distribution method, where an effective parameter is the carrier temperature, which becomes much higher than that of the lattice, as if the field heats the carriers [21.38]. Not too much work has been published on hot-carrier phenomena in Si, but most

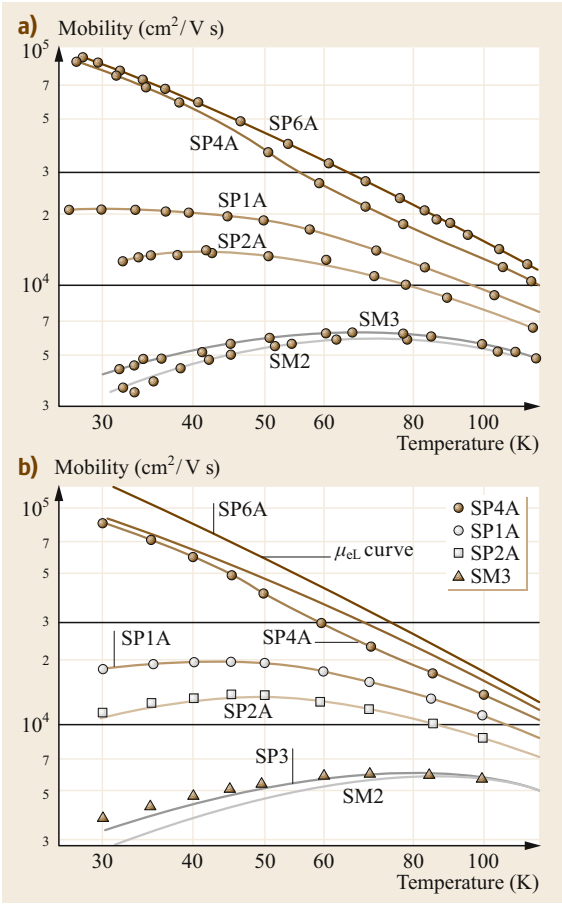


Fig. 21.13a,b Comparison of the theoretical results on $\mu_e(T)$ and $\mu_e^H(T)$ (*symbols*) with the experimental data (*full lines*). The indicated samples are those from Table 21.12. (After [21.101] with permission)

of the existing work was overviewed in the two referencess cited below, and in item 3 in Sect. 21.2.3:

1. The electron drift velocity [21.113]. Experimental results for electrons in Si obtained using a time-of-flight technique were presented for $8\,\mathrm{K} \le T \le 300\,\mathrm{K}$ and $\boldsymbol{E}$ in the magnitude range 1.5×10^4–$5\times10^4\,\mathrm{V/cm}$, oriented along the ⟨111⟩, ⟨110⟩ and ⟨100⟩ crystallographic directions. At 8 K the dependence of the transit time upon sample thickness allowed a measurement of the valley repopulation time when the electric field is ⟨100⟩ oriented. These experimental results were interpreted with Monte Carlo calculations in the same ranges of T and $\boldsymbol{E}$. The model included the many-valley structure of the conduction band of Si, acoustic intra-valley scattering with the correct momentum and energy relaxation and correct equilibrium phonon population, several inter-valley scatterings, and ionized-impurity scattering.
2. Hole drift velocity [21.18]. Drift velocities for holes in high-purity Si, were measured by a time-of-flight technique with $\boldsymbol{E}$ in the amplitude range from 3×10^4–$5\times10^4\,\mathrm{V/cm}$ along the ⟨100⟩, ⟨110⟩, and ⟨111⟩ directions, and at $8\,\mathrm{K} \le T \le 300\,\mathrm{K}$. The Ohm's mobility is theoretically interpreted on the basis of a two-band model consisting of a spherical parabolic and a spherical nonparabolic band, and relaxation-time approximation. The low-T Ohm's mobility proved to be strongly influenced by the nonparabolic structure of the heavy-hole band. The high-field region, $E = |\boldsymbol{E}| \ge 10^3\,\mathrm{V/cm}$, was analyzed using a single, warped heavy-hole band model and a Monte Carlo technique. Anisotropy of the hot-hole drift velocity due to warping of the valence band was measured. Optical- and acoustic-phonon scattering were found to have comparable rates.

Impact ionization phenomenon are different in appearance and are not considered here in detail; see item 7 in Sect. 21.2.3.

Drift Velocity – Electric-Field Relationship

As the field strength $\boldsymbol{E}$ overcomes some limit (which depends on the lattice temperature, doping and type of carrier) the drift velocity (Sect. 21.2.3) becomes a sub-linear function of $\boldsymbol{E}$, i. e., the mobility starts to decrease with increasing $\boldsymbol{E}$. As revealed by the early study for n-Si [21.116], in the range 10^3–$10^5\,\mathrm{V/cm}$, n does remain constant, but the mobility decreases appreciably, as seen in Fig. 21.14. Further studies obtained, in addition, the following patterns [21.18, 19, 113]:

1. Drift velocity $\boldsymbol{v}_d(\boldsymbol{E})$ is anisotropic in the orientation of $\boldsymbol{E}$ relative to the crystallographic axes: different values of $\boldsymbol{v}_d$ are obtained for the same E for $\boldsymbol{E}$ along different high-symmetry directions ⟨111⟩, ⟨110⟩ and ⟨100⟩. For less-symmetric directions $\boldsymbol{v}_d$ is not even parallel to $\boldsymbol{E}$, which is known as the Sasaki–Shibuya effect [21.38]. The $\boldsymbol{v}_d$–$\boldsymbol{E}$ relation depends strongly on T, as seen from Fig. 21.15a,b for the ⟨111⟩ and ⟨100⟩ directions. The anisotropy becomes stronger as T decreases, as seen from Fig. 21.16a,b. For electrons, the curves of $\boldsymbol{v}_d$ versus $\boldsymbol{E}$ for different T (Fig. 21.15a) tend to join together at some value of $\boldsymbol{E}$, which increases with T, even though such a rejoining is not reached at all considered T. For holes (Fig. 21.15b), this tendency is much weaker and the rejoining field is much higher than for electrons.
2. For electrons at the highest $\boldsymbol{E} \| \langle 111 \rangle$, a region where $\boldsymbol{v}_d$ is independent of E (saturation) is obtained

Table 21.13 Weakly compensated n-Si samples [21.53]

Sample	N_d (cm^{-3})	N_a (cm^{-3})	E_d (meV)	m^*/m_0	ρ_{300} (Ω cm)
Si:P 1	9.5×10^{15}	4.2×10^{12}	45.64	0.3218	0.66
Si:P 3	2.5×10^{14}	2.3×10^{13}	45.14	0.3218	17.1
Si:P 4	9.6×10^{14}	2.0×10^{14}	43.39	0.3218	4.76
Si:P 5	3.3×10^{14}	8.6×10^{12}	45.38	0.3218	13.6
Si:P 6	4.3×10^{13}	7.7×10^{12}	45.42	0.3218	123
Si:As 1	1.4×10^{16}	6.0×10^{12}	53.64	0.3218	0.43
Si:As 2	7.9×10^{15}	4.3×10^{13}	52.54	0.3218	0.66
Si:As 3	7.5×10^{16}	1.8×10^{13}	52.52	0.3218	0.13
Si:Sb 1	7.4×10^{15}	5.3×10^{12}	42.59	0.3218	0.68

Table 21.14 Parameters of the best fit of the intrinsic mobility to (21.29)

Carriers	γ						$A(10^8\,K^\gamma cm^2V^{-1}s^{-1})$			
Electrons	1.5[a]	2.6[b]	2.5[c]	–	1.5[e]	2.42[f]	5.5[a]	40[b]	21[c]	14.3[f]
Holes	2.3[a]	2.3[b]	2.7[c]	2.7–2.9[d]	2.0[e]	2.20[f]	2.4[a]	2.5[b]	23[c]	1.35[f]

[a] After [21.92], [b] after [21.96], [c] after [21.93], [d] after [21.99], [e] after [21.100] at $T \le 100$ K, [f] after [21.19]

(Fig. 21.15a). For holes, this type of saturation is not approached for all available $\boldsymbol{E}$ (Fig. 21.15b).

3. Ohm's law is reached at $T > 45$ and 100 K for electrons and holes, respectively, and the results for mobility are in agreement with those obtained by standard techniques for the linear regime. At $T < 45$ K for electrons, and $T < 100$ K for holes, even at the lowest $\boldsymbol{E}$, $\boldsymbol{v}_d$ is not linear with $\boldsymbol{E}$. At low temperatures the mobility is as high as $5\times10^5\,cm^2/(V\,s)$ at $E = 1.5$ V/cm for electrons ($T = 8$ K), and $3.5\times10^5\,cm^2/(V\,s)$ at $E = 3$ V/cm for holes ($T = 6$ K). A negative differential mobility (NDM) was found with $\boldsymbol{E}\|\langle100\rangle$ at $T < 40$ K for electrons (Fig. 21.16a). The field at which $\boldsymbol{v}_d$ reaches the maximum before the NDM region (NDM threshold), decreases with decreasing T. For holes, a net low-field saturation region of $\boldsymbol{v}_d$ (50 V/cm $< \boldsymbol{E} <$ 150 V/cm at $T = 6$ K) shows up for $\boldsymbol{E}\|\langle111\rangle$, $\langle110\rangle$ and $\langle100\rangle$ (see top of Fig. 21.16b). This effect gradually disappears as T increases, and at $T > 38$ K it is no longer visible.

For electrons, the anisotropy is due to a repopulation of the valleys: when $\boldsymbol{E}\|\langle111\rangle$ the six valleys (Fig. 21.3b) are equally oriented with respect to $\boldsymbol{E}$ and all of them give the same contribution to $\boldsymbol{v}_d$. When, for example, $\boldsymbol{E}\|\langle100\rangle$, two valleys exhibit the effective mass m_l in the direction of the field, while the remaining four exhibit the transverse mass $m_t < m_l$. Electrons in the transverse valleys respond with a higher mobility, are heated to a greater extent by the field and transfer electrons to the two longitudinal, *colder* and slower valleys, which results in a lower $\boldsymbol{v}_d$ than for $\boldsymbol{E}\|\langle111\rangle$, as seen, e.g., in Fig. 21.16a. For holes, the anisotropy is due to the two warped and degenerate valence sub-

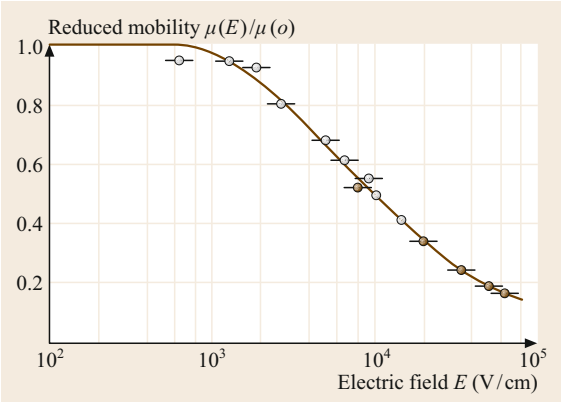


Fig. 21.14 Variation of reduced mobility with electric field applied to a 70 Ω cm n-Si sample along the direction $\langle111\rangle$. The electron concentration is assumed to be constant. (After [21.24] with permission)

bands (Fig. 21.3c) resulting in different effective masses for holes with different $\boldsymbol{k}$, and in the lowest and highest $\boldsymbol{v}_d$ for $\boldsymbol{E}\|\langle110\rangle$ and $\boldsymbol{E}\|\langle100\rangle$, as seen in Fig. 21.16b. A summary of the curves of $\boldsymbol{v}_d$ versus $\boldsymbol{E}$ is presented on Fig. 21.16c.

Anisotropy becomes stronger as T decreases, since a lower T leads to less-effective relaxation effects. In particular, at $T \approx 45$ K the repopulation of electron valleys may be so rapid with increasing $\boldsymbol{E}\|\langle100\rangle$ that NDM occurs with $\boldsymbol{E}$ in the range 20–60 V/cm. NDM was observed via oscillations of the current [21.132], and in the I–V [21.128, 132, 133] and $\boldsymbol{v}_d$–$\boldsymbol{E}$ [21.113, 134, 135] characteristics. Whether the curves of electron $\boldsymbol{v}_d$ versus $\boldsymbol{E}$ for $\boldsymbol{E}\|\langle100\rangle$ and $\boldsymbol{E}\|\langle111\rangle$ join together at the high-field limit is still an open point. Theoretical considerations [21.136] seem to indicate that a small

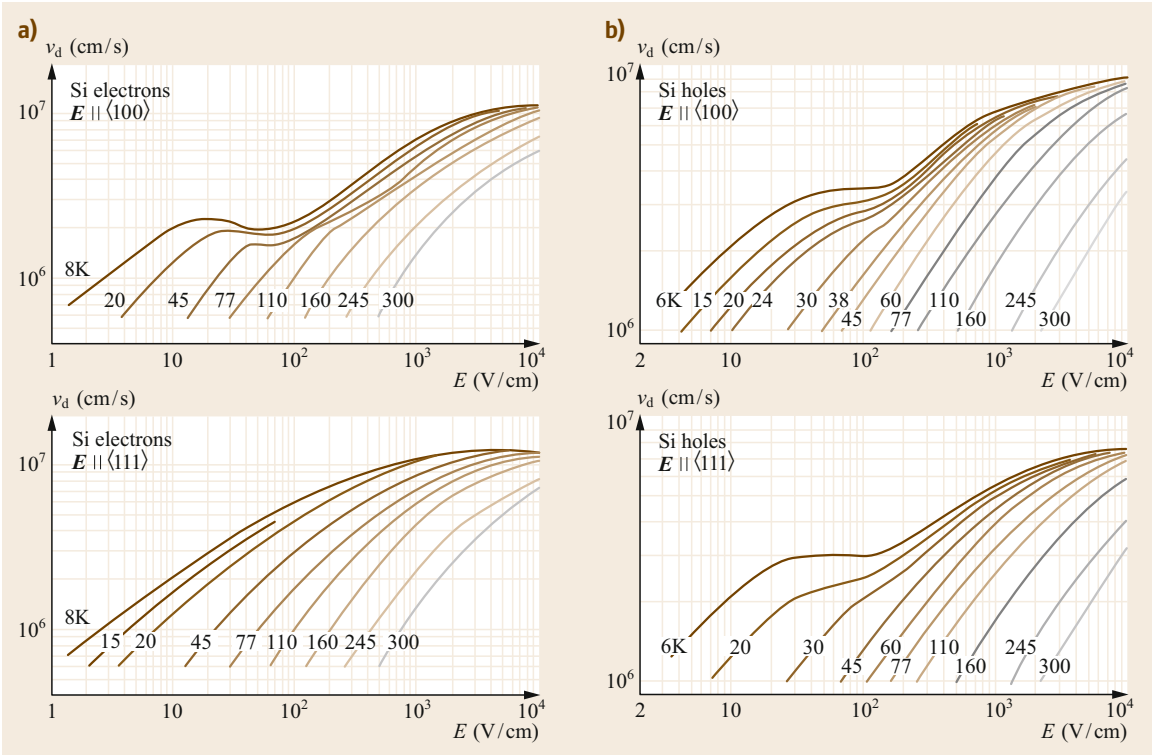


Fig. 21.15 Carriers' drift velocity for E, applied along two indicated directions, measured at different indicated T by the time-of-flight method: **(a)** Electrons. (After [21.117] with permission); **(b)** Holes. (After [21.18] with permission)

difference ($\approx 5\%$) between the two curves should remain. However, this difference is comparable to experimental error and merging of the two curves has been claimed at several temperatures in the experimental results [21.113, 118]. In the case of holes, both theoretical and experimental results [21.18] indicate that, up to the highest fields attained (5×10^4 V/cm), no merging occurs. The aforementioned anomalous behavior of the curves of $\boldsymbol{v}_d$ versus $\boldsymbol{E}$ for holes at low temperatures ($T < 38$ K) – that these curves tend to *saturate* at intermediate field strengths and then rise again – has been explained [21.18, 117] by a nonparabolic distortion of the heavy-hole spectrum.

A general tendency of $\boldsymbol{v}_d(\boldsymbol{E})$ to saturate at the highest fields is an important phenomenon which was considered from the very beginning of the hot-electron investigations [21.137–139]. The original explanation [21.138] for this behavior was based on a rough physical model, in which it was assumed that the carriers emit an optical phonon as soon as they reach the energy $\hbar\omega_0$ of the phonon that limits the drift velocity by the saturation value $v_s \approx \sqrt{\hbar\omega_0/m^*}$ [21.138]. For electrons, inter-valley instead of optical phonons should be considered (Sect. 21.2.3). Subsequent calculations [21.140, 141] showed, that the Shockley formula should be considered only as a rough evaluation, since at very high electric field the transport process depends on many factors including several phonon dispersion curves, phonon emission and absorption, and nonparabolic band corrections. Good agreement between the Monte Carlo calculations and the experimental data for high-purity n-Si is seen in Fig. 21.17a. In the case of holes the saturation was neither predicted theoretically, nor found experimentally. A nearly saturated drift velocity was observed [21.122] only at room temperature, and the saturation should occur, by extrapolation, at $\boldsymbol{E} \geq 2 \times 10^5$ V/cm with v_{sh} around 10^7 cm/s (Fig. 21.16c).

Empirically adjusted formulas for $v_d(E)$ have been proposed for varying N_i and $T = 300$ K [21.142]

$$v_d = \frac{\mu_0 \boldsymbol{E}}{\sqrt{1 + \frac{S(N_i/N_{ref})}{N_i/N_{ref}+S} + \frac{(E/E_1)^2}{E/E_1+F} + \left(\frac{E}{E_2}\right)^2}}, \quad (21.31)$$

and for $N_i = 0$ and varying T [21.108]

$$v_d = v_m \frac{E/E_c}{[1 + (E/E_c)^\beta]^{1/\beta}} ; \quad (21.32)$$

the best-fit parameters are presented in Table 21.7. Also a formula was suggested [21.19] for the electron satu-

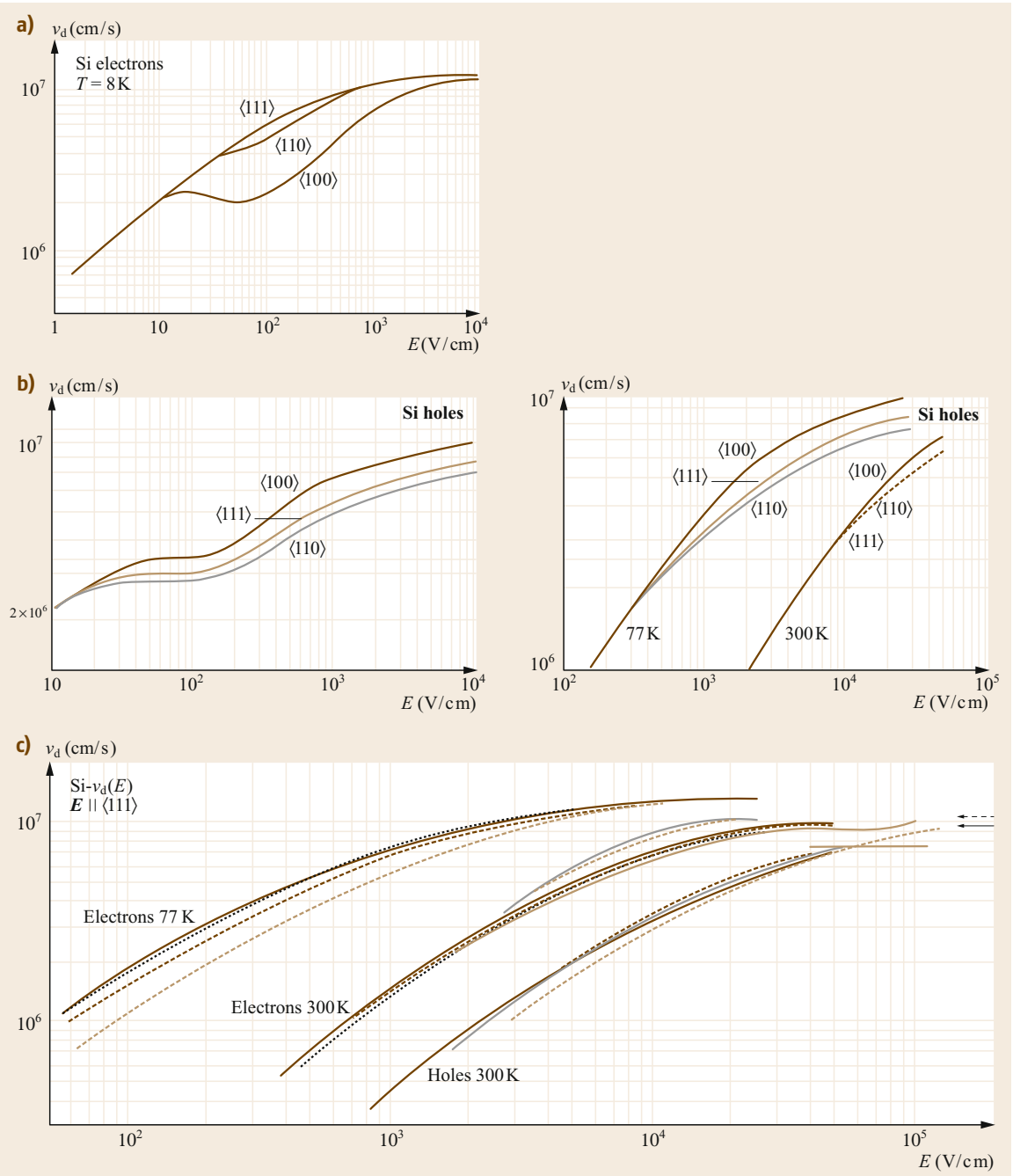


Fig. 21.16a–c Carriers' drift velocity anisotropy in details at different indicated T: **(a)** Electrons. (After [21.117] with permission); **(b)** Holes. (After [21.18] with permission); **(c)** Summary of the data on the drift velocity obtained by different techniques. Holes, $T = 300$ K: *full* – time-of-flight [21.118], *dashed* – time-of-flight [21.119], *double-dot-dashed* – time-of-flight [21.120], *cross-dashed* – $I(V)$ in space-charge-limited current (SCLC) regime [21.121], *dot-dashed* – $I(V)$ [21.122], *arrow* – extrapolated value of v_{sh} [21.122]. Electrons, $T = 300$ K: *full* – time-of-flight [21.118], *long-dash-dot* – time-of-flight [21.119], *dashed* – time-of-flight [21.120], *double-dot-dash* – $I(V)$ in the SCLC regime [21.123], *dots* – $I(V)$ [21.124], *dash-crossed* – $I(V)$ [21.125], *dashed arrow* – v_{se} from $I(V)$ in avalanche diodes [21.126, 127]. Electrons, $T = 77$ K: *full* – time-of-flight [21.118], *dashed* – $I(V)$ [21.128], *dash-dot* – $I(V)$ [21.129], *dotted* – $I(V)$ [21.130, 131]. (After [21.19] with permission)

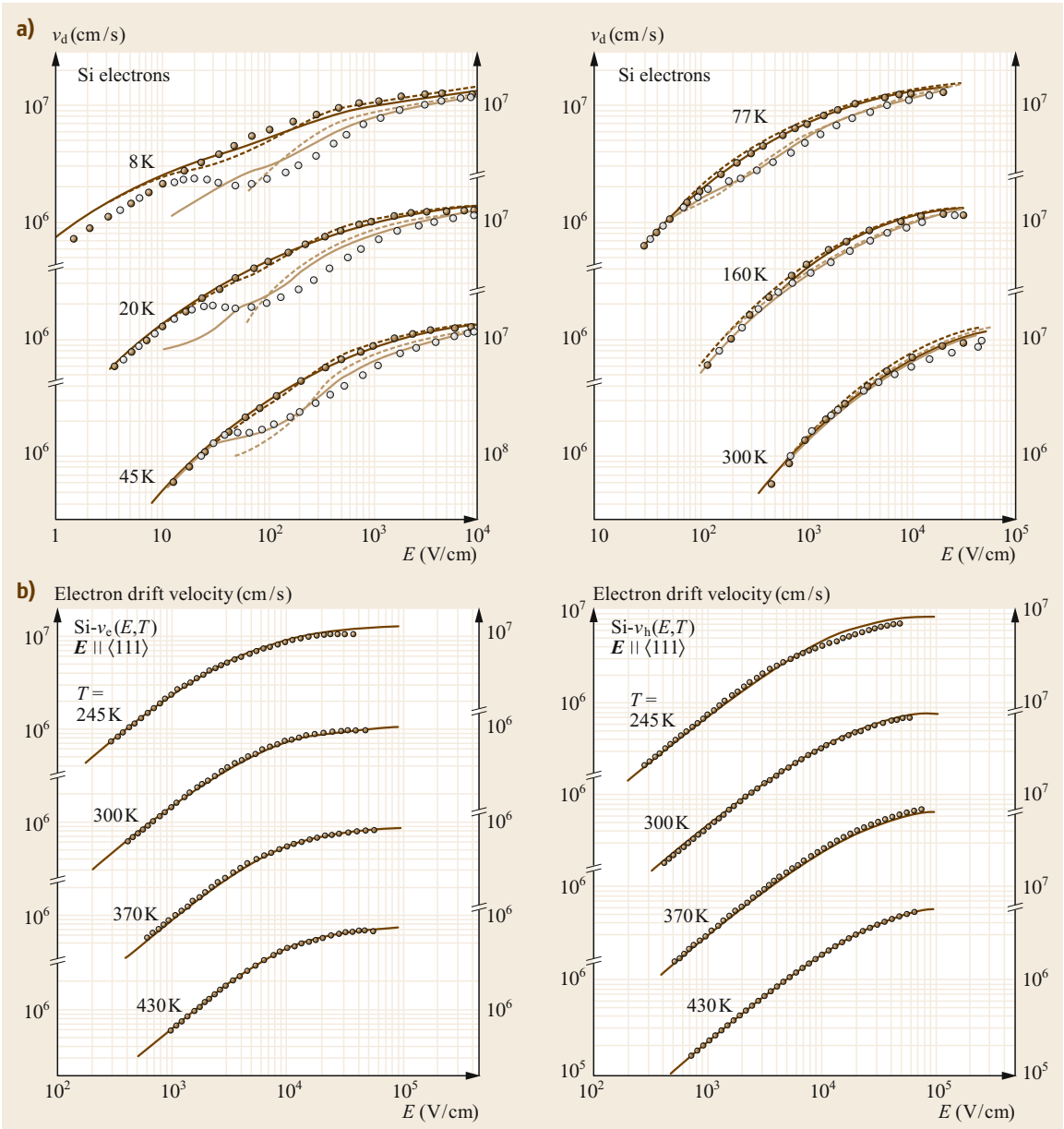


Fig. 21.17a–c Measured versus simulation results on drift velocity in Si: **(a)** $v_{de}(E)$ at different indicated temperatures. *Circles* indicate experimental data by time-of-flight method, *solid circle* – $E \| \langle 111 \rangle$ and $E \|$ *open circle* – $\langle 100 \rangle$, and lines Monte Carlo simulations results neglecting impurity scattering, *full* and *broken* – using six lowest and three allowed inter-valley phonons (Table 21.4), respectively. For $T \leq 45$ K, the theoretical curves which refer to the $\langle 100 \rangle$ direction are interrupted, as it was not possible to reach a sufficient precision in the simulated v_d. (After [21.113, 115] with permission); **(b)** The $v_{de}(E)$ and $v_{de}(E)$, for $E \| \langle 111 \rangle$, in high-purity Si at four different T. *Solid circle* – the experimental data [21.118] and *lines* are the best-fit curves obtained with (21.32) using the parameters listed in Table 21.7. (After [21.19] with permission); **(c)** $v_{se}(T)$ obtained experimentally with different techniques and numerically fitted. *Solid circle* – time-of-flight [21.113, 115, 118, 143], *open circle* – $I(V)$ in avalanche diodes [21.126, 127], *open triangle* – $I(V)$ in the SCLC regime [21.123]; the latter data have been normalized to 9.6×10^6 cms^{-1} at 300 K; *line* represents the best fitting curve of (21.33). (After [21.19] with permission)

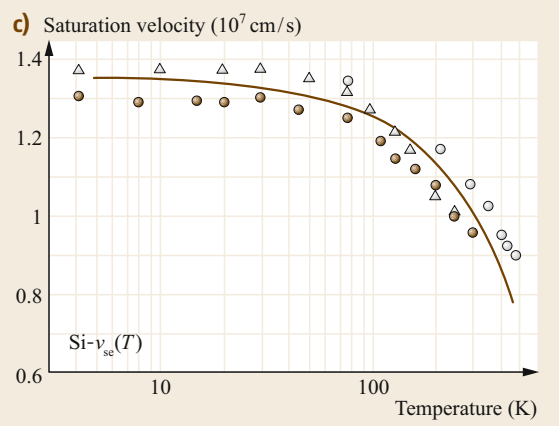


Fig. 21.17 (continued)

rated drift velocity as a function of temperature

$$v_{se}(T) = \frac{2.4\times10^7\,\mathrm{cm/s}}{1+0.8\times\exp\left(\frac{600\,\mathrm{K}}{T}\right)}. \qquad (21.33)$$

The applicability of (21.32) and (21.33) is clearly demonstrated by Fig. 21.17b,c, respectively.

Hot-Carrier Diffusion

Knowledge of diffusion processes is useful for a better understanding of charge-transport phenomena and correct simulation of high-frequency devices. At low fields the diffusion coefficient D is related to the mobility by the Einstein relation (Sect. 21.1.3). At high fields it becomes a field-dependent tensor, which describes the diffusion with respect to $\boldsymbol{E}$ depending on its direction in the crystallographic frame. The difference between diffusion parallel to and perpendicular to $\boldsymbol{E}$ is greater than the variations caused by changes in the crystallographic direction. Therefore, investigations set out to determine the field dependence of both the longitudinal ($D_\parallel$) and transverse ($D_\perp$) components of the diffusivity tensor with respect to $\boldsymbol{E}$ irrespective of its orientation in the crystal.

The longitudinal diffusion coefficient can be measured by the Haynes–Shockley or time-of-flight [21.120] technique by observing the difference between the fall and rise times of the current pulse. This difference is caused by the spread of the carriers traveling across the sample and is simply related to $D_\parallel$ [21.144]. Analogously, $D_\perp$ can be obtained by observing the spread of the current perpendicular to the direction of the field. The current is originated by a point excitation on one surface of a Si wafer and is collected on the opposite surface by several electrodes of appropriate geometry [21.145]; this technique is sometimes called *geometrical* [21.19]. Finally, both

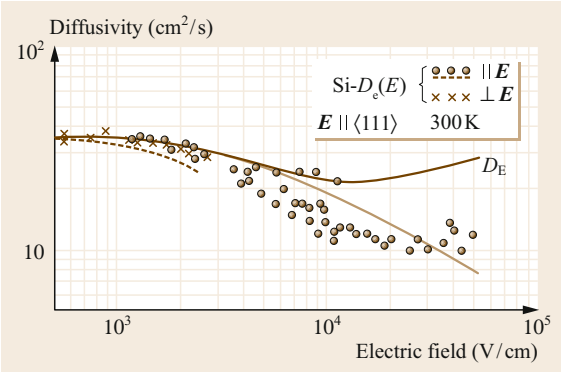


Fig. 21.18 Diffusion coefficients of electrons versus E in Si at $T = 300$ K: *solid circle* – $D_\parallel$, time-of-flight data [21.113, 115]; *dashed* – $D_\parallel$, noise data [21.146]; *crosses* – $D_\perp$, the geometrical technique data [21.145]; *full lines* – the Monte Carlo computed $D_\parallel$ and D_E [21.113, 115]. (After [21.19] with permission)

$D_\parallel$ and $D_\perp$ have been related to noise measurements, parallel and perpendicular respectively to the current direction [21.146].

Figure 21.18 shows some experimental results on the field dependence of $D_\parallel$ and $D_\perp$ for electrons in Si at room temperature with $\boldsymbol{E}\|\langle111\rangle$. The data obtained by the noise measurements are in a reasonable agreement with the time-of-flight results, although the former cover a narrower range of $\boldsymbol{E}$, just outside Ohm's region. As $\boldsymbol{E}$ increases $D_\parallel$ decreases to about one third of its low-field value ($\approx 36\,\mathrm{cm^2/s}$), which is in substantial agreement with theoretical Monte Carlo computations for the nonparabolic band [21.113]. The results for transverse diffusion showed that, as $\boldsymbol{E}$ increases, $D_\perp$ also decreases, but to a lesser extent than $D_\parallel$. There exists a hypothesis of validity to extrapolate the Einstein relation outside the linear region by $D_E = \frac{2\bar{E}}{3q}\mu(E)$. As seen from Fig. 21.18, this yields a qualitative interpretation of $D_\perp$ for not too high fields. As far as $D_\parallel$ is concerned, the diffusion process seems much more complex than pictured by the Einstein relation. For holes, the dependence of $D_\parallel$ on $\boldsymbol{E}$ was found to be similar to that for electrons [21.19].

Impact Ionization

Impact ionization is an important charge-generation mechanism. It occurs in many silicon-based devices, either determining the useful characteristic of the device or causing an unwanted parasitic effect. The breakdown of a silicon p-n diode is caused by impact ionization if its breakdown voltage is larger than about 8 V. The operation of such devices as thyristors, impact avalanche transit time (IMPATT) diodes and trapped plasma avalanche-triggered transit (TRAPATT) diodes

is based on avalanche generation, the phenomenon that results from impact ionization. The avalanche generation also plays an increasing role in degradation due to hot-carrier effects and bipolar parasitic breakdown of metal–oxide–semiconductor (MOS) devices, the geometrical dimensions of which have been scaled down recently.

The ionization rate is defined as the number of electron–hole pairs generated by a carrier per unit distance traveled in a high electric field, and is different for electrons and for holes. Impact ionization can only occur when the particle gains at least the threshold energy for ionization from the electrical field. This can be derived from the application of the energy and momentum conservation laws to the amount $E_i \approx 1.5E_g$ (assuming that the effective masses of electron and hole are equal). A large spread of experimental values for E_i exists, with a breakdown field of order of 3×10^5 V/cm. For more detailed consideration we refer the reader to the review article [21.147] noted in Sect. 21.2.3.

21.2.3 Review Material

The following materials may be recommended for further reading:

1. Electrical properties of Si [21.24]. Summary of papers on the subject that were published over a decade until 1965 are overviewed. Miscellaneous properties, such as piezoresistance and high-electric-field mobility, were also presented.
2. Electron mobility and resistivity in n-Si versus dopant density and temperature [21.148]. An improved model for computing μ_e as a function of N_d and T in uncompensated n-Si was formulated. The effects of electron–electron interaction on conventional scattering processes, as well as their anisotropies were incorporated empirically. The model was verified to ±5% of the mobility measured on wafers doped by phosphorous in the range $10^{13}-10^{19}\,\mathrm{cm}^{-3}$.
3. Bulk charge-transport properties of Si [21.19]. Review of knowledge on the subject with special emphasis on application to solid-state devices. Most attention was devoted to experimental findings at room temperatures and to high-field properties. The techniques for drift-velocity measurements and the principles of Monte Carlo simulation were overviewed. Empirical expressions were given, when possible, for the most important transport quantities as functions of T, $N_{d(a)}$ and E.
4. Semi-empirical relations for the carrier mobilities [21.149]. From a review of different publications on $\mu_{e,h}$ in Si, the authors proposed an approximated calculation procedure, analogous to that of Li and Thurber, which permits a quick and accurate evaluation of $\mu_{e(h)}$ over a wide range of T, $N_{d(a)}$ and $n(p)$. The proposed relations are well adapted to device simulation since they allow short computation times.
5. Minority-carrier recombination in heavily doped silicon [21.150]. A review of understanding of the recombination of minority carriers in heavily doped Si. A short phenomenological description of the carrier recombination process and lifetime was provided and the main theories of these were briefly reviewed with indications for their expected contributions in heavily doped Si. The various methods used for measuring the minority-carrier lifetime in heavily doped Si were described and critically examined. The insufficiency of existing theories to explain the patterns of lifetime versus doping was clearly demonstrated.
6. Minority-carrier transport modeling in heavily doped silicon emitters [21.110]. The experimental and theoretical efforts that addressed such important issues as: (i) the incomplete understanding of the minority-carrier physics in heavily doped Si, (ii) the lack of precise measurements for the minority-carrier parameters, (iii) the difficulties encountered with the modeling of transport and recombination in nonhomogeneously doped regions, and (iv) problems with the characterization of real emitters in bipolar devices, were reviewed with the goal of being able to achieve accurate modeling of the current injected into an arbitrarily heavily doped region in a silicon device.
7. Impact ionization in silicon: a review and update [21.147]. The multiplication factor and the ionization rate were revisited. The interrelationship between these parameters together with the multiplication and breakdown models for diodes and MOS transistors were discussed. Different models were compared and test structures were discussed to measure the multiplication factor accurately enough for reliable extraction of the ionization rates. Multiplication measurements at different T were performed on a BJT, and yielded new electron ionization rates at relatively low electric fields. An explanation for the spread of existed experimental data on ionization rate was given. A new implementation method for a local avalanche model into a device simulator was presented.

21.3 Optical Properties

21.3.1 Diversity of Silicon as an Optical Material

In *dielectric*-like material $n > k$ ($\varepsilon_1 > 0$), where for transparency and opacity it is necessary that $n \gg k$ ($\varepsilon_1 \gg \varepsilon_2$, Table 21.15) and $n \approx k$ ($\varepsilon_1 \ll \varepsilon_2$, Table 21.16), respectively. In *metallic*-like material $n < k$ ($\varepsilon_1 < 0$), where for good reflectivity and bad reflectivity it is necessary that $n \ll k(|\varepsilon_1| \gg \varepsilon_2)$ and $n \approx k(|\varepsilon_1| \ll \varepsilon_2)$, respectively, Table 21.16. Since in Si ε depends on the wavelength and carrier concentration, it may exhibit all these types of optical behavior ranging from dielectric-like to metallic-like. For example, $\langle 111 \rangle$ undoped Si ($n = 2.3 \times 10^{14}\,\mathrm{cm}^{-3}$) at $\lambda = 0.62\,\mu\mathrm{m}$ behaves as a transparent dielectric, as $\varepsilon_1 = 15.254$ and $\varepsilon_2 = 0.172$ at that wavelength, while it is an opaque dielectric at $\lambda = 0.295\,\mu\mathrm{m}$, where $\varepsilon_1 = 2.371$ and $\varepsilon_2 = 45.348$. For heavily doped Si ($n = 10^{20}\,\mathrm{cm}^{-3}$), $n = 1.911$ and $k = 8.63$ at $\lambda = 16.67\,\mu\mathrm{m}$, so it behaves as a good metallic reflector while at $\lambda = 2\,\mu\mathrm{m}$, where $n = 3.47$ and $k = 6.131 \times 10^{-3}$ it is a transparent dielectric.

21.3.2 Measurements of Optical Constants

Various methods are used to measure the dielectric constant of single-crystalline Si including transmission, reflection, and ellipsometric methods. For a smooth opaque sample, the quantity of interest is the complex reflection amplitude ρ_r (at normal incidence) defined by

$$\rho_\mathrm{r} = -\frac{\sqrt{\varepsilon}-1}{\sqrt{\varepsilon}+1}\,, \quad |\rho_\mathrm{r}|^2 = \frac{(n-1)^2+k^2}{(n+1)^2+k^2} = R_0\,,$$
$$\psi = \arctan\frac{2k}{n^2+k^2-1} = \arg(\rho_\mathrm{r})\,, \quad \mathrm{mod}\ (\pi)\,. \tag{21.34}$$

The Fresnel reflectance spectrum $R_0(\omega)$ is measured using reflectometry. The KKR analysis also applies to the causal function

$$\ln(\rho_\mathrm{r}) = \frac{1}{2}\ln R_0 + \mathrm{i}\psi$$

that gives

$$\psi(\omega) = \frac{1}{2\pi}\int_0^\infty \ln\left|\frac{\omega+\Omega}{\omega-\Omega}\right| \frac{\mathrm{d}}{\mathrm{d}\Omega}\ln R_0(\Omega)\mathrm{d}\Omega\,. \tag{21.35}$$

With the $\psi(\omega)$ retrieved in this way, the last two relations in (21.34) are simultaneously solved to yield the n and k spectra. The KKR method requires, in principle, data over an infinite ω range, which are supplied by measurements over a confined range and an appeal to simple models for high and low frequencies. This limitation is avoided using optical measurements on smooth slabs.

For a slab of thickness d, the reflectance and transmittance at normal incidence are given by

$$R = R_0\frac{1-2T_0\cos 2\varphi + T_0^2}{1-2R_0T_0\cos 2\chi + R_0^2T_0^2}\,,$$
$$T = T_0\frac{1-2R_0\cos 2\psi + R_0^2}{1-2R_0T_0\cos 2\chi + R_0^2T_0^2}\,,$$
$$T_0 = \mathrm{e}^{-\alpha d}\,, \tag{21.36}$$

where $\varphi = 2\pi nd/\lambda$ is the optical-path phase, $\chi = \varphi + \psi$ and $\alpha = 4\pi k/\lambda$ is the power attenuation coefficient. The parameter αd characterizes the slab's opacity. At $n > k$ ($\varepsilon_1 > 0$, (21.17)), i. e., in the optically dielectric range, α is referred to as *absorption coefficient*. In the optically metallic range that exists for doped Si in the IR, where $k > n$ ($\varepsilon_1 < 0$, (21.17)), α^{-1} is the skin depth multiplied by a factor ≈ 1. If the dispersion of the optical constants is disregarded, R and T are periodic in $1/\lambda$ (21.20). For nonopaque slabs, these interference effects show up at $nd/\lambda > 1$. For thick slabs ($nd/\lambda \gg 1$), even small fluctuation of n or/and λ causes changes in φ of the order 2π. Therefore, if such factors as bulk defects, source incoherence and low spectral resolution are present in an experiment, then the averages of R and T over φ, given by

$$\langle R\rangle = R_0\left(1 + T_0\langle T\rangle\right),$$
$$\langle T\rangle = T_0\frac{1-2R_0\cos 2\psi + R_0^2}{1-R_0^2T_0^2}\,, \tag{21.37}$$

are more appropriate than (21.36). This *multiple-reflection* approximation [21.151] has proved effective in countless studies. The apparent absorbance of the slab in the multiple-reflection approximation is calculated by

$$\langle A\rangle = 1 - \langle R\rangle - \langle T\rangle$$
$$= \frac{(1-R_0)(1-T_0) - 4R_0T_0\sin^2\psi}{1-R_0T_0}\,. \tag{21.38}$$

Equation (21.38) at $k \ll n$ is extensively used in silicon-wafer thermometry [21.152, 153]. Given $\langle R\rangle$, $\langle T\rangle$ and d, (21.37) builds up a system of two equations for the two unknowns n and k. Thus, measurement of

the reflection and transmission on the same slab of known thickness allows one to retrieve the optical constants. This R–T measurement method [21.154] is greatly simplified at $2k \ll n^2 + k^2 - 1$. Under this low-loss condition the above system can be solved analytically for T_0 and R_0. The calculated T_0 directly yields k, and n is then found using the calculated R_0. The R–T technique is the best method at $\alpha d \leq 1$, while at $\alpha d \gg 1$, where solving the aforementioned system becomes an ill-conditioned problem, the KKR analysis is more reliable. In two last decades, spectroscopic ellipsometry has gained wide recognition for being more precise than photometric methods. In ellipsometry, the ratio of reflectance for s- and p-polarized radiation, and the relative phase shift between the two, are both measured at large angles of incidence [21.155].

The measured results are affected by the structural atomic-scale properties of the samples. These properties are defined by polishing processes – mechanical or chemical – that affect the surface damage and roughness, the properties of the surface native oxide, the growth mechanism of the measured layer, grain boundaries, and the quality of the cleaved surface. Since Si samples may be optically inhomogeneous, retrieving the optical constants from measurements may become a complicated inverse electromagnetic problem [21.156, 157], which is why some of the reported data for ε disagree by up to 30%. A detailed list of publications on the subject can be found in [21.158, 159]. Emphasis on these effects should be especially considered when transmission measurement is done for a wavelength range in which the absorption coefficient is large and thin samples are therefore required.

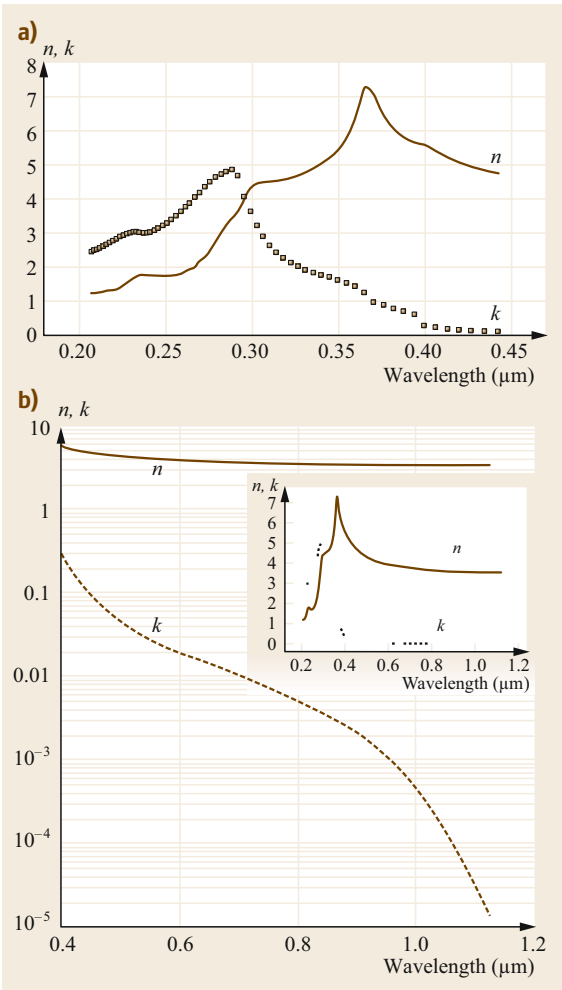


Fig. 21.19 Refractive index n and the extinction coefficient k versus wavelength, using Adachi and Geist models in the wavelength ranges: **(a)** 0.2–0.45 μm, **(b)** 0.4–1.127 μm. *Inset* – the complete range from 0.2 to 1.127 μm

21.3.3 Modeling of Optical Constants

A method for calculating ε_1 and ε_2 and then n, k and α for photon energies of 0–6 eV ($\lambda > 0.2\,\mu$m) has been reported [21.82]. The calculated data are in excellent agreement with experimental data for the wavelength range 0.2–4 μm [21.156, 160]. The model is based on the KKR (Sect. 21.1.3) and takes into account the dependence of ε on the energy-band structure. It considers the effect of indirect-band-gap and inter-band transitions as well as the electron (conduction bands) and hole (valence bands) density of states. The fundamental absorption (generation of electron–hole pair) edge energy of 1.12 eV corresponds to the indirect tran-

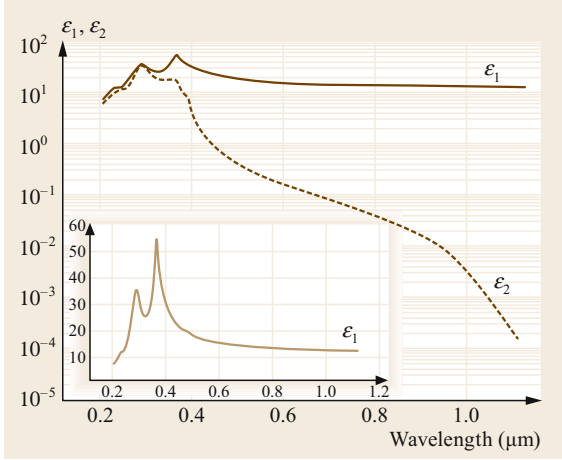


Fig. 21.20 Real ε_1 and imaginary ε_2 parts of dielectric constant versus wavelength, using Adachi and Geist models. *Inset*: real dielectric constant ε_1 versus wavelength

Table 21.15 The refractive index n and the extinction coefficient k of n-Si with electron concentration $N = 10^{16}\,\mathrm{cm}^{-3}$ at various wavelength

Energy (eV)	Wavenumber (cm^{-1})	Wavelength (μm)	n		k	
			HW[a]	GDA[b]	HW[a]	GDA[b]
0.6199	5000	2.000	3.453	3.453	1.160×10^{-7}	1.514×10^{-7}
0.5579	4500	2.222	3.447	3.447	1.594×10^{-7}	2.006×10^{-7}
0.4959	4000	2.500	3.441	3.441	2.273×10^{-7}	2.751×10^{-7}
0.4339	3500	2.857	3.435	3.435	3.398×10^{-7}	3.945×10^{-7}
0.3720	3000	3.333	3.431	3.431	5.403×10^{-7}	6.001×10^{-7}
0.3100	2500	4.000	3.427	3.427	9.347×10^{-7}	9.912×10^{-7}
0.2480	2000	5.000	3.424	3.424	1.827×10^{-6}	1.850×10^{-6}
0.1860	1500	6.667	3.421	3.421	4.334×10^{-6}	4.210×10^{-6}
0.1240	1000	10.00	3.419	3.419	1.463×10^{-5}	1.392×10^{-5}
0.1116	900	11.11	3.419	3.419	2.006×10^{-5}	1.910×10^{-5}
0.09919	800	12.50	3.419	3.419	2.856×10^{-5}	2.727×10^{-5}
0.08679	700	14.29	3.418	3.418	4.262×10^{-5}	4.092×10^{-5}
0.07439	600	16.67	3.418	3.418	6.765×10^{-5}	6.549×10^{-5}
0.06199	500	20.00	3.417	3.417	1.168×10^{-4}	1.143×10^{-4}
0.04959	400	25.00	3.416	3.416	2.278×10^{-4}	2.259×10^{-4}
0.03720	300	33.33	3.413	3.413	5.382×10^{-4}	5.413×10^{-4}
0.03472	280	35.71	3.412	3.412	6.612×10^{-4}	6.669×10^{-4}
0.03224	260	38.46	3.411	3.411	8.247×10^{-4}	8.341×10^{-4}
0.02976	240	41.67	3.410	3.410	1.047×10^{-3}	1.061×10^{-3}
0.02728	220	45.45	3.408	3.408	1.356×10^{-3}	1.379×10^{-3}
0.02480	200	50.00	3.406	3.406	1.799×10^{-3}	1.834×10^{-3}
0.02232	180	55.56	3.399	3.399	3.481×10^{-3}	3.563×10^{-3}
0.01984	160	62.50	3.403	3.403	2.458×10^{-3}	2.511×10^{-3}
0.01736	140	71.43	3.394	3.394	5.154×10^{-3}	5.284×10^{-3}
0.01488	120	83.33	3.385	3.385	8.085×10^{-3}	8.297×10^{-3}
0.01240	100	100.0	3.372	3.372	1.370×10^{-2}	1.405×10^{-2}

[a] Values calculated using an empiricial fit [21.78], [b] values calculated using GDA [21.79]

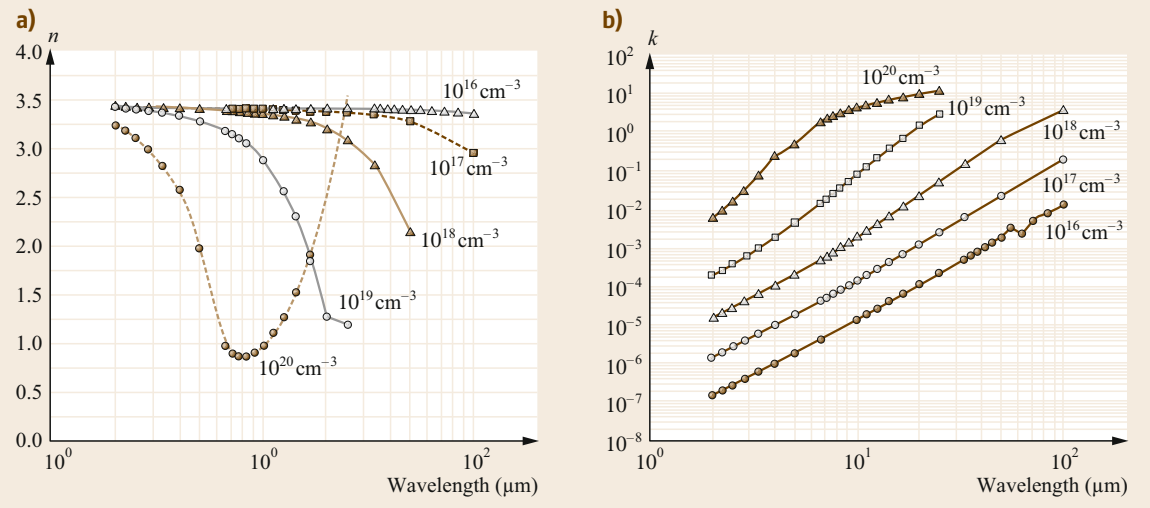


Fig. 21.21a,b Optical constants versus wavelength at various electron concentrations using the GDA model. **(a)** Refractive index n. **(b)** Extinction coefficient k

Table 21.16 The refractive index n and the extinction coefficient k of n-Si with electron concentration $N = 10^{20}\,\mathrm{cm}^{-3}$ at various wavelength

Energy (eV)	Wavenumber (cm^{-1})	Wavelength (μm)	n HW[a]	n GDA[b]	k HW[a]	k GDA[b]
0.6199	5000	2.000	3.270 [P]		1.834×10^{-2}	
			3.257 [As]	3.247	2.403×10^{-2}	6.131×10^{-3}
0.5579	4500	2.222	3.219		2.549×10^{-2}	
			3.203	3.190	3.341×10^{-2}	9.701×10^{-3}
0.4959	4000	2.500	3.151		3.698×10^{-2}	
			3.130	3.112	4.845×10^{-2}	1.664×10^{-2}
0.4339	3500	2.857	3.053		5.671×10^{-2}	
			3.027	3.000	7.432×10^{-2}	3.232×10^{-2}
0.3720	3000	3.333	2.902		9.410×10^{-2}	
			2.867	2.828	1.233×10^{-1}	8.053×10^{-2}
0.3100	2500	4.000	2.644		1.765×10^{-1}	
			2.597	2.577	2.314×10^{-1}	2.549×10^{-1}
0.2480	2000	5.000	2.138		4.176×10^{-1}	
			2.087	1.990	5.461×10^{-1}	5.336×10^{-1}
0.2356	1900	5.263	1.981		5.226×10^{-1}	
			1.939		6.796×10^{-1}	
0.2232	1800	5.556	1.800		6.717×10^{-1}	
			1.780		8.617×10^{-1}	
0.2108	1700	5.882	1.604		8.873×10^{-1}	
			1.626		1.107×10^{0}	
0.1984	1600	6.250	1.423		1.188×10^{0}	
			1.503		1.417×10^{0}	
0.1860	1500	6.667	1.295		1.566×10^{0}	
			1.429	0.971	1.777×10^{0}	1.902×10^{0}
0.1736	1400	7.143	1.237		1.989×10^{0}	
			1.411	0.893	2.172×10^{0}	2.369×10^{0}
0.1612	1300	7.692	1.239		2.437×10^{0}	
			1.442	0.873	2.591×10^{0}	2.859×10^{0}
0.1488	1200	8.333	1.291		2.911×10^{0}	
			1.518	0.872	3.037×10^{0}	3.387×10^{0}
0.1364	1100	9.091	1.390		3.418×10^{0}	
			1.642	0.913	3.514×10^{0}	3.960×10^{0}
0.1240	1000	10.00	1.540		3.971×10^{0}	
			1.817	0.985	4.034×10^{0}	4.603×10^{0}

[a] Values calculated using an empiricial fit [21.78], [b] values calculated using GDA [21.79]

sition from the highest valence band to the lowest conduction band. Sharp changes in the optical constants are obtained at wavelengths around 0.367, 0.29 and 0.233 µm, which correspond to the energy-band critical points of 3.38, 4.27 and 5.317 eV, respectively. An additional analytical model for calculating the n and k values for the wavelength range 0.4–1.127 µm has been developed [21.158]. The model is based on measured k and n data [21.161–163], where the calculated values are within ±10% of the measured values. Using Adachi and Geist models we have calculated and plotted n, k, ε_1 and ε_2 for the wavelength range 0.2–1.127 µm, as seen in Figs. 21.19 and 21.20.

In the IR range the lattice and free carriers, which arise from doping, give additive contributions. The free-carrier contribution to the dielectric function when the radiation quantum is much smaller than the mean electron energy, $\hbar\omega \ll \bar{E}$, is considered using the semiclassical Drude approximation (DA, Sect. 21.1.4), which leads to the simple formula

$$\varepsilon = \varepsilon_\mathrm{L} - \frac{n_\mathrm{L}^2\omega_\mathrm{p}^2}{\omega(\omega + \mathrm{i}\gamma)}\,,$$

$$\gamma = \frac{1}{\tau}\,, \qquad (21.39)$$

Table 21.16 (continued)

Energy (eV)	Wavenumber (cm^{-1})	Wavelength (μm)	n HW[a]	n GDA[b]	k HW[a]	k GDA[b]
0.1116	900	11.11	1.751 [P]		4.584×10^0	
			2.057 [As]	1.105	4.608×10^0	5.341×10^0
0.09919	800	12.50	2.042		5.278×10^0	
			2.378	1.278	5.251×10^0	6.216×10^0
0.08679	700	14.29	2.444		6.079×10^0	
			2.810	1.530	5.982×10^0	7.282×10^0
0.08431	680	14.71	2.542		6.255×10^0	
			2.913		6.141×10^0	
0.08183	660	15.15	2.646		6.437×10^0	
			3.022		6.305×10^0	
0.07935	640	15.63	2.758		6.625×10^0	
			3.139		6.474×10^0	
0.07690	620	16.13	2.877		6.820×10^0	
			3.262		6.649×10^0	
0.07439	600	16.67	3.006		7.022×10^0	
			3.395	1.911	6.829×10^0	8.630×10^0
0.07180	580	17.24	3.143		7.231×10^0	
			3.536		7.015×10^0	
0.06933	560	17.86	3.291		7.448×10^0	
			3.686		7.208×10^0	
0.06685	540	18.52	3.451		7.674×10^0	
			3.846		7.407×10^0	
0.06438	520	19.23	3.622		7.909×10^0	
			4.018		7.614×10^0	
0.06199	500	20.00	3.807		8.154×10^0	
			4.201	2.517	7.829×10^0	1.040×10^1
0.05951	480	20.83	4.006		8.409×10^0	
			4.398		8.052×10^0	
0.05703	460	21.74	4.222		8.675×10^0	
			4.608		8.285×10^0	
0.05455	440	22.73	4.455		8.954×10^0	
			4.835		8.528×10^0	
0.05207	420	23.81	4.707		9.245×10^0	
			5.078		8.782×10^0	
0.04959	400	25.00	4.982		9.551×10^0	
			5.340	3.551	9.048×10^0	1.285×10^1

[a] Values calculated using an empiricial fit [21.78], [b] values calculated using GDA [21.79]

where τ is the phenomenological relaxation time, ω_p is the electron plasma frequency given by

$$\omega_p = \sqrt{\frac{4\pi q^2 N}{\varepsilon_L m_c}} \,. \tag{21.40}$$

Here N denotes the carriers' concentration, $m_c \approx 0.26 m_0$ for electrons (Table 21.1) and $m_c = (0.33–0.39) m_0$ for holes (Sect. 21.1.4).

Advanced GDA, which extrapolates the free-carrier contribution to the range of $\hbar\omega > \bar{E}$ was also considered [21.79]. In Tables 21.15 and 21.16 we tabulated the values of n and k for practical use at two extreme electron concentrations: 10^{16} and $10^{20}\,cm^{-3}$. The results placed in the columns headed by the abbreviation HW were obtained using DA (21.22), (21.39) and (21.40) and the empirically adjusted τ versus N fit [21.78]. The columns of theoretical values calculated by us using GDA are headed by the corresponding abbreviation. A disagreement between the theory and the empirical fit notable at heavy doping has been discussed in detail [21.79]. The dependencies of n and k on wavelength in the range at carrier concentration of $10^{16}–10^{20}\,cm^{-3}$ in steps of $10\,cm^{-3}$ are shown in Fig. 21.21.

21.3.4 Electric-Field and Temperature Effects on Optical Constants

It has been known for many years that the optical absorption spectrum of Si can also be affected by external electric fields (the Franz–Keldysh effect). The Franz–Keldysh effect, which alters the α spectrum of crystalline Si, is field-induced tunneling between valence- and conduction-band states. In recent years, the generic term *electroabsorption* has been adopted for $\Delta\alpha$ versus E effects. The effect of electric field on the refractive index is shown in Fig. 21.22 [21.164]. Sharp changes occur around the wavelengths correspond to the band-gap transition.

The temperature dependence of the refractive index of high-purity damage-free Si, for photon energies less than 3 eV in the temperature range 300–500 K is given by [21.162]

$$\frac{\Delta n}{\Delta T} \approx 1.3 \times 10^{-4} n \quad (\mathrm{K}^{-1}) \,. \tag{21.41}$$

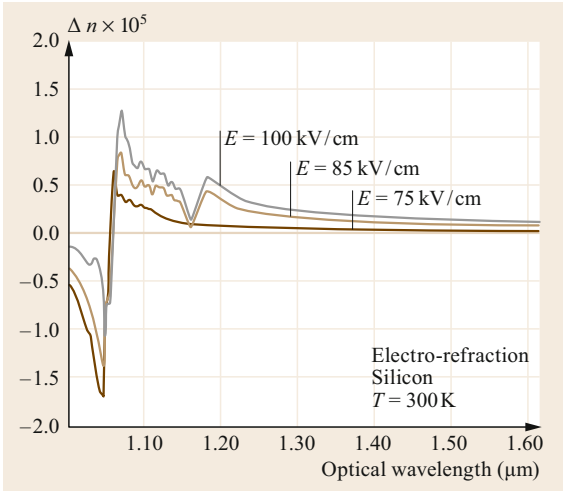


Fig. 21.22 Refractive index change versus wavelength at various electric fields (after [21.164])

References

21.1 C. Kittel: *Introduction to Solid State Physics*, 6th edn. (Wiley, New York 1986)
21.2 C. Kittel: *Quantum Theory of Solids*, 2nd edn. (Wiley, New York 1987)
21.3 K. Seeger: *Semiconductor Physics* (Springer, New York 1982)
21.4 T.S. Moss: *Optical Properties of Semiconductors* (Butterworths, London 1959)
21.5 S.M. Sze: *Physics of Semiconductor Devices* (Wiley, New York 1981)
21.6 H.M. van Driel: Appl. Phys. Lett. **44**, 617 (1984)
21.7 J.R. Chelikowsky, M.L. Cohen: Phys. Rev. B **14**, 556 (1976)
21.8 E.O. Kane: J. Phys. Chem. Solids **1**, 82 (1956)
21.9 G. Dresselhaus, A.F. Kip, C. Kittel: Phys. Rev. **98**, 368 (1955)
21.10 J.C. Hensel, H. Hasegawa, M. Nakayama: Phys. Rev. **138**, 225 (1965)
21.11 L. Kleinmann, J.C. Phillips: Phys. Rev. **118**, 1153 (1960)
21.12 M. Cardona, F.H. Pollak: Phys. Rev. **142**, 530 (1966)
21.13 J.C. Hensel, G. Feher: Phys. Rev. **129**, 1041 (1963)
21.14 P. Lawaetz: Phys. Rev. B **4**, 3460 (1971)
21.15 S. Zwerdling, K.J. Button, B. Lax, L.M. Roth: Phys. Rev. Lett. **4**, 173 (1960)
21.16 I. Balslev, P. Lawaetz: Phys. Lett. **19**, 3460 (1965)
21.17 B. Lax, J.G. Mavroides: Phys. Rev. **100**, 1650 (1955)
21.18 G. Ottaviani, L. Reggiani, C. Canali, F. Nava, A. Alberigi-Quaranta: Phys. Rev. B **12**, 3318 (1975)
21.19 C. Jacoboni, C. Canali, G. Ottaviani, A. Alberigi-Quaranta: Solid State Electron. **20**, 77 (1977)
21.20 H.D. Barber: Solid State Electron. **10**, 1039 (1967)
21.21 M.A. Green: J. Appl. Phys. **67**, 2944 (1990)
21.22 A.B. Sproul, M.A. Green: J. Appl. Phys. **70**, 846 (1991)
21.23 R.F. Pierret: In: *Advanced Semiconductor Fundamentals*, Modular Series on Solid State Devices, ed. by G.W. Neudeck, R.F. Pierret (Pearson Education, New York 2003)
21.24 W.R. Runyan: *Silicon Semiconductor Technology* (McGraw–Hill, New York 1965), Chap. 8
21.25 V.I. Fistul: *Heavily Doped Semiconductors* (Plenum, New York 1969)
21.26 W. Kohn, J.M. Luttinger: Phys. Rev. **97**, 1721 (1955)
21.27 W. Kohn, J.M. Luttinger: Phys. Rev. **98**, 915 (1955)
21.28 W. Kohn, D. Schechter: Phys. Rev. **99**, 1903 (1955)
21.29 E. Burstein, G. Picus, B. Henvis, R. Wallis: J. Phys. Chem. Solids **1**, 65 (1956)
21.30 G. Picus, E. Burstein, B. Henvis: J. Phys. Chem. Solids **1**, 75 (1956)
21.31 N.F. Mott: *Metal–Insulator Transitions*, 2nd edn. (Taylor Francis, London 1990) p. 2
21.32 R.H. Hall, J.H. Racette: J. Appl. Phys. **35**, 379 (1964)
21.33 W.M. Bullis: Solid State Electron. **9**, 143 (1966)
21.34 B. Jensen: *Handbook of Optical Constants of Solids*, Vol. 2 (Academic, Orlando 1985) p. 169
21.35 M. Kohler: Z. Phys. **124**, 777 (1948)
21.36 M. Kohler: Z. Phys. **125**, 679 (1949)
21.37 B.R. Nag: *Theory of Electrical Transport in Semiconductors* (Pergamon, Oxford 1972)
21.38 E.M. Conwell: *High Field Transport in Semiconductors* (Academic, New York 1967)
21.39 W. Fawcett, A.D. Boardman, S. Swain: J. Phys. Chem. Solids **31**, 1963 (1970)

21.40 W. Dumke: Phys. Rev. **124**, 1813 (1961)
21.41 I. Balslev: Phys. Rev. **143**, 636 (1966)
21.42 B.N. Brockhouse: Phys. Rev. Lett. **2**, 256 (1959)
21.43 W.A. Harrison: Phys. Rev. **104**, 1281 (1956)
21.44 H. Ehrenreich, A.W. Overhauser: Phys. Rev. **104**, 331 (1956)
21.45 J. Bardeen, W. Shockley: Phys. Rev. **80**, 72 (1950)
21.46 G.L. Bir, G.E. Pikus: Fiz. Tverd. Tela **22**, 2039 (1960), Sov. Phys. – Solid State 2 (1961) 2039
21.47 M. Tiersten: IBM J. Res. Dev. **5**, 122 (1961)
21.48 C. Herring, E. Vogt: Phys. Rev. **101**, 944 (1956)
21.49 M. Tiersten: J. Phys. Chem. Solids **25**, 1151 (1964)
21.50 H.W. Streitwolf: Phys. Status Solidi (b) **37**, K47 (1970)
21.51 D. Long: Phys. Rev. **120**, 2024 (1960)
21.52 D.L. Rode: Phys. Status Solidi (b) **53**, 245 (1972)
21.53 P. Norton, T. Braggins, H. Levinstein: Phys. Rev. B **8**, 5632 (1973)
21.54 C. Canali, C. Jacobini, F. Nava, G. Ottaviani, A. Alberigi-Quaranta: Phys. Rev. B **12**, 2265 (1975)
21.55 C. Erginsoy: Phys. Rev. **79**, 1013 (1950)
21.56 N. Sclar: Phys. Rev. **104**, 1559 (1956)
21.57 E.M. Conwell, V.F. Weisskopf: Phys. Rev. **77**, 338 (1950)
21.58 H. Brooks: Phys. Rev. **83**, 388 (1951)
21.59 C. Herring: Bell Syst. Tech. J. **36**, 237 (1955)
21.60 R. Dingle: Philos. Mag. **46**, 831 (1955)
21.61 H. Brooks: *Advances in Electronics and Electron Physics*, Vol. 7 (Academic, New York 1955) p. 85
21.62 D. Chattopadhyay, H.J. Queisser: Rev. Mod. Phys. **53**, 745 (1981)
21.63 A.G. Samoilovich, I.Y. Korenblit, I.V. Dakhovskii, V.D. Iskra: Fiz. Tverd. Tela **3**, 3285 (1961), Sov. Phys. – Solid State (1962) 2385
21.64 P.M. Eagles, D.M. Edwards: Phys. Rev. **138**, A1706 (1965)
21.65 I.V. Dakhovskii: Fiz. Tverd. Tela **55**, 2332 (1963), Sov. Phys. – Solid State 5 (1964) 1695
21.66 D. Long, J. Myers: Phys. Rev. **120**, 39 (1960)
21.67 J.R. Meyer, F.J. Bartoli: Phys. Rev. B **23**, 5413 (1981)
21.68 H.I. Ralph, G. Simpson, R.J. Elliot: Phys. Rev. B **11**, 2948 (1975)
21.69 H.K. Jung, H. Ohtsuka, K. Taniguchi, C. Hamaguchi: J. Appl. Phys. **79**, 2559 (1996)
21.70 G.L. Bir, E. Normantas, G.E. Pikus: Fiz. Tverd. Tela **4**, 1180 (1962)
21.71 P.P. Debye, E.M. Conwell: Phys. Rev. **93**, 693 (1954)
21.72 J. Appel: Phys. Rev. **122**, 1760 (1961)
21.73 M. Luong, A.W. Shaw: Phys. Rev. B **4**, 30 (1971)
21.74 J. Appel: Phys. Rev. **125**, 1815 (1962)
21.75 B.E. Sernelius: Phys. Rev. B **41**, 2436 (1990)
21.76 P.A. Shumann, R.P. Phillips: Solid State Electron. **10**, 943 (1967)
21.77 M.A. Saifi, R.H. Stolen: J. Appl. Phys. **43**, 1171 (1972)
21.78 J. Humlicek, K. Wojtechovsky: Czech. J. Phys. B **38**, 1033 (1988)
21.79 M. Auslender, S. Hava: In: *Handbook of Optical Constants of Solids*, Vol. 3, ed. by D.E. Palik (Academic, New York 1998) p. 155
21.80 W. Kaiser, P.H. Keck, C.F. Lange: Phys. Rev. **101**, 1264 (1956)
21.81 S. Adachi: Phys. Rev. B **38**, 12966 (1988)
21.82 S. Adachi: J. Appl. Phys. **66**, 3224 (1989)
21.83 T. Aoki, S. Adachi: J. Appl. Phys. **69**, 1574 (1991)
21.84 K.-F. Berggren, B.E. Sernelius: Phys. Rev. B **24**, 1971 (1981)
21.85 P.E. Schmid: Phys. Rev. B **23**, 5531 (1981)
21.86 L. Viña, M. Cardona: Phys. Rev. B **29**, 6739 (1984)
21.87 W.G. Spitzer, H.Y. Fan: Phys. Rev. **106**, 882 (1957)
21.88 M. Balkanski, A. Aziza, E. Amzallag: Phys. Status Solidi (b) **31**, 323 (1969)
21.89 M. Cardona, W. Paul, H. Brooks: Z. Naturforsch. **101**, 329 (1960)
21.90 L.M. Lambert: Phys. Status Solidi (a) **11**, 461 (1972)
21.91 L. Pearson, J. Bardeen: Phys. Rev. **75**, 865 (1949)
21.92 M.B. Prince: Phys. Rev. **93**, 1204 (1954)
21.93 G.W. Ludwig, R.L. Watters: Phys. Rev. **101**, 1699 (1956)
21.94 D.C. Cronemeyer: Phys. Rev. **105**, 522 (1957)
21.95 P.P. Debye, T. Kohane: Phys. Rev. **94**, 724 (1954)
21.96 F.J. Morin, J.P. Maita: Phys. Rev. **96**, 28 (1954)
21.97 G. Backenstoss: Phys. Rev. **108**, 579 (1957)
21.98 D. Long: Phys. Rev. **107**, 672 (1957)
21.99 D. Long, J. Myers: Phys. Rev. **109**, 1098 (1958)
21.100 R.A. Logan, A.J. Peters: J. Appl. Phys. **31**, 122 (1960)
21.101 D. Long, J. Myers: Phys. Rev. **115**, 1107 (1959)
21.102 L.J. Neuringer, D. Long: Phys. Rev. **135**, A788 (1964)
21.103 A.G. Samoilovich, I.Y. Korenblit, I.V. Dakhovskii, V.D. Iskra: Fiz. Tverd. Tela **3**, 2939 (1961), Sov. Phys. – Solid State 3 (1962) 2148
21.104 I.G. Kirnas, P.M. Kurilo, P.G. Litovchenko, V.S. Lutsyak, V.M. Nitsovich: Phys. Status Solidi (a) **23**, K123 (1974)
21.105 J.C. Irvin: Bell Syst. Tech. J. **41**, 387 (1962)
21.106 S.M. Sze, J.C. Irvin: Solid State Electron. **11**, 559 (1968)
21.107 F. Mousty, P. Ostoja, L. Passari: J. Appl. Phys. **45**, 4576 (1974)
21.108 D.M. Caughey, R.F. Thomas: Proc. IEEE **55**, 2192 (1967)
21.109 G. Baccarani, P. Ostoja: Solid State Electron. **18**, 1039 (1975)
21.110 J.A. del Alamo, R.M. Swanson: Solid State Electron. **30**, 1127 (1987)
21.111 C. Hilsum: Electron. Lett. **10**, 259 (1074)
21.112 Y. Furukawa: J. Phys. Soc. Japan **16**, 577 (1961)
21.113 C. Canali, C. Jacoboni, G. Ottaviani, A. Alberigi-Quaranta: Appl. Phys. Lett. **27**, 278 (1975)
21.114 E.H. Putley, W.H. Mitchell: Proc. Phys. Soc. **72**, 193 (1958)
21.115 C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, A. Alberigi-Quaranta: Phys. Rev. B **12**, 2265 (1975)
21.116 E.A. Davies, D.S. Gosling: J. Phys. Chem. Solids **23**, 413 (1962)
21.117 C. Canali, M. Costato, G. Ottaviani, L. Reggiani: Phys. Rev. Lett. **31**, 536 (1973)
21.118 C. Canali, G. Ottaviani, A. Alberigi-Quaranta: J. Phys. Chem. Solids **32**, 1707 (1971)
21.119 C.B. Norris, J.F. Gibbons: IEEE Trans. Electron. Dev. **14**, 30 (1967)
21.120 T.W. Sigmon, J.F. Gibbons: Appl. Phys. Lett. **15**, 320 (1969)
21.121 V. Rodriguez, H. Ruegg, M.-A. Nicolet: IEEE Trans. Electron. Dev. **14**, 44 (1967)

21.122 T.E. Seidel, D.L. Scharfetter: J. Phys. Chem. Solids **28**, 2563 (1967)
21.123 V. Rodriguez, M.-A. Nicolet: J. Appl. Phys. **40**, 496 (1969)
21.124 B.L. Boichenko, V.M. Vasetskii: Sov. Phys. Solid State **7**, 1631 (1966)
21.125 A.C. Prior: J. Phys. Chem. Solids **12**, 175 (1959)
21.126 C.Y. Duh, J.L. Moll: IEEE Trans. Electron. Dev. **14**, 46 (1967)
21.127 C.Y. Duh, J.L. Moll: Solid State Electron. **11**, 917 (1968)
21.128 M.H. Jorgensen, N.O. Gram, N.I. Meyer: Solid-State Commun. **10**, 337 (1972)
21.129 M. Asche, B.L. Boichenko, O.G. Sarbej: Phys. Status Solidi (b) **9**, 323 (1965)
21.130 J.G. Nash, J.W. Holm-Kennedy: Appl. Phys. Lett. **24**, 139 (1974)
21.131 J.G. Nash, J.W. Holm-Kennedy: Appl. Phys. Lett. **25**, 507 (1974)
21.132 M. Asche, O.G. Sarbej: Phys. Status Solidi (a) **38**, K61 (1971)
21.133 N.O. Gram: Phys. Lett. A **38**, 235 (1972)
21.134 C. Canali, A. Loria, F. Nava, G. Ottaviani: Solid-State Commun. **12**, 1017 (1973)
21.135 M. Asche, O.G. Sarbej: Phys. Status Solidi (a) **46**, K121 (1971)
21.136 J.P. Nougier, M. Rolland, O. Gasquet: Phys. Rev. B **11**, 1497 (1975)
21.137 E.J. Ryder, W. Shockley: Phys. Rev. **81**, 139 (1951)
21.138 W. Shockley: Bell. Syst. Tech. J. **30**, 990 (1951)
21.139 E.J. Ryder: Phys. Rev. **90**, 766 (1953)
21.140 M. Costato, L. Reggiani: Lett. Nuovo Cimento **3**, 728 (1970)
21.141 C. Jacoboni, R. Minder, G. Majni: J. Phys. Chem. Solids **36**, 1129 (1975)
21.142 D.L. Scharfetter, H.K. Gummel: IEEE Trans. Electron. Dev. **16**, 64 (1969)
21.143 C. Canali, G. Ottaviani: Phys. Lett. A **32**, 147 (1970)
21.144 J.G. Ruch, G.S. Kino: Phys. Rev. **174**, 921 (1968)
21.145 G. Persky, D.J. Bartelink: J. Appl. Phys. **42**, 4414 (1971)
21.146 J.P. Nougier, M. Rolland: Phys. Rev. B **8**, 5728 (1973)
21.147 W. Maes, K. de Meyer, R. van Overstraeten: Solid State Electron. **33**, 705 (1990)
21.148 S.S. Li, W.R. Thurber: Solid State Electron. **20**, 609 (1977)
21.149 J.M. Dorkel, P. Leturcq: Solid-State Electron. **24**, 821 (1981)
21.150 M.S. Tyagi, R. van Overstraeten: Solid State Electron. **10**, 1039 (1983)
21.151 H.O. McMachon: J. Opt. Soc. Am. **40**, 376 (1950)
21.152 J.C. Sturm, C.M. Reaves: IEEE Trans. Electron. Dev. **39**, 81 (1992)
21.153 P.J. Timans: J. Appl. Phys. **74**, 6353 (1993)
21.154 P.A. Shumann Jr., W.A. Keenan, A.H. Tong, H.H. Gegenwarth, C.P. Schneider: J. Electrochem. Soc. **118**, 145 (1971)
21.155 R.M.A. Azzam, N.M. Bashara: *Ellipsometry and Polarized Light*, 2nd edn. (Elsevier, Amsterdam 1987)
21.156 D.E. Aspnes, A.A. Studna: Phys. Rev. B **27**, 985 (1983)
21.157 E. Barta, G. Lux: J. Phys. D: Appl. Phys. **16**, 1543 (1983)
21.158 J. Geist: In: *Handbook of Optical Constants of Solids*, Vol. 3, ed. by D.E. Palik (Academic, New York 1998) p. 519
21.159 D.E. Aspnes, A.A. Studna, E. Kinsbron: Phys. Rev. B **29**, 768 (1984)
21.160 H.R. Philipp, E.A. Taft: Phys. Rev. B **120**, 37 (1960)
21.161 G.E. Jellison Jr.: Opt. Mater. **1**, 41 (1992)
21.162 H.A. Weaklien, D. Redfield: J. Appl. Phys. **50**, 1491 (1979)
21.163 D.F. Edward: *Handbook of Optical Constants of Solids*, Vol. 2 (Academic Press, Orlando 1985) p. 547
21.164 R.A. Soref, B.R. Bennett: IEEE J. Quantum Electron. **QE-23**, 123 (1987)

22. Silicon-Germanium: Properties, Growth and Applications

Yaser M. Haddara, Peter Ashburn, Darren M. Bagnall

Silicon-germanium is an important material that is used for the fabrication of SiGe heterojunction bipolar transistors and strained Si metal-oxide-semiconductor (MOS) transistors for advanced complementary metal-oxide-semiconductor (CMOS) and BiCMOS (bipolar CMOS) technologies. It also has interesting optical properties that are increasingly being applied in silicon-based photonic devices. The key benefit of silicon-germanium is its use in combination with silicon to produce a heterojunction. Strain is incorporated into the silicon-germanium or the silicon during growth, which also gives improved physical properties such as higher values of mobility. This chapter reviews the properties of silicon-germanium, beginning with the electronic properties and then progressing to the optical properties. The growth of silicon-germanium is considered, with particular emphasis on the chemical vapour deposition technique and selective epitaxy. Finally, the properties of polycrystalline silicon-germanium are discussed in the context of its use as a gate material for MOS transistors.

22.1 **Physical Properties of Silicon-Germanium** 524
22.1.1 Critical Thickness 524
22.1.2 Band Structure 526
22.1.3 Dielectric Constant 527
22.1.4 Density of States 527
22.1.5 Majority-Carrier Mobility in Strained $Si_{1-x}Ge_x$ 527
22.1.6 Majority-Carrier Mobility in Tensile-Strained Si on Relaxed $Si_{1-x}Ge_x$ 528
22.1.7 Minority-Carrier Mobility in Strained $Si_{1-x}Ge_x$ 529
22.1.8 Apparent Band-Gap Narrowing in $Si_{1-x}Ge_x$ HBTs 529
22.2 **Optical Properties of SiGe** 530
22.2.1 Dielectric Functions and Interband Transitions 530
22.2.2 Photoluminescence 530
22.2.3 SiGe Quantum Wells 532
22.3 **Growth of Silicon-Germanium** 534
22.3.1 In situ Hydrogen Bake 534
22.3.2 Hydrogen Passivation 534
22.3.3 Ultra-Clean Epitaxy Systems 534
22.3.4 $Si_{1-x}Ge_x$ Epitaxy 535
22.3.5 Selective $Si_{1-x}Ge_x$ Epitaxy 535
22.3.6 Heteroepitaxial Ge on Si 537
22.4 **Polycrystalline Silicon-Germanium** 537
22.4.1 Electrical Properties of Polycrystalline $Si_{1-x}Ge_x$ 538
References 539

Silicon-germanium ($Si_{1-x}Ge_x$) alloys have been researched since the late 1950s [22.1], but it is only in the past 30 years or so that these layers have been applied to new types of transistor technology. $Si_{1-x}Ge_x$ was first applied in bipolar technologies [22.2, 3], but more recently has been applied to metal-oxide-semiconductor (MOS) technologies [22.4–7]. This has been made possible by the development of new growth techniques, such as molecular-beam epitaxy (MBE), low-pressure chemical vapour deposition (LPCVD) and ultra-high-vacuum chemical vapour deposition (UHV-CVD). The key feature of these techniques that has led to the development of $Si_{1-x}Ge_x$ transistors is the growth of epitaxial layers at low temperatures (500–700 °C). This allows $Si_{1-x}Ge_x$ layers to be grown without disturbing the doping profiles of structures already present in the silicon wafer. $Si_{1-x}Ge_x$ layers can be successfully grown on silicon substrates even though there is a lattice mismatch between silicon and germanium of 4.2%.

The primary property of $Si_{1-x}Ge_x$ that is of interest for bipolar transistors is the band gap, which is smaller than that of silicon and controllable by varying the

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_22

germanium content. Band-gap engineering concepts, which were previously only possible in compound semiconductor technologies, have now become viable in silicon technology. These concepts have introduced new degrees of freedom in the design of bipolar transistors that have led to dramatic improvements in transistor performance. In $Si_{1-x}Ge_x$ heterojunction bipolar transistors (HBTs), the $Si_{1-x}Ge_x$ layer is incorporated into the base and the lower band gap of $Si_{1-x}Ge_x$ than Si is used to advantage to dramatically improve the high-frequency performance. $Si_{1-x}Ge_x$ HBTs have been produced with values of cut-off frequency, f_T, approaching 300 GHz [22.8], a value unimaginable in silicon bipolar transistors. Values of gate delay well below 10 ps can be achieved in properly optimised $Si_{1-x}Ge_x$ HBTs [22.9]. In $Si_{1-x}Ge_x$ MOS field-effect transistors (MOSFETs), the $Si_{1-x}Ge_x$ layer is incorporated in the channel and is used to give improved values of mobility. Initially, strained $Si_{1-x}Ge_x$ layers on silicon substrates were used to give improved hole mobility in p-channel transistors [22.7], but more recently thin, strained silicon layers on relaxed SiGe virtual substrates have been used to give improvements in both electron and hole mobility [22.4–6].

In this chapter, the properties of single-crystal silicon-germanium will first be outlined, followed by a description of the methods used for growing silicon-germanium layers. The properties and applications of polycrystalline silicon-germanium are also discussed later in the article.

22.1 Physical Properties of Silicon-Germanium

Silicon and germanium are completely miscible over the full range of compositions and hence can be combined to form $Si_{1-x}Ge_x$ alloys with the germanium content, x, ranging from 0 to 1 (0–100%). $Si_{1-x}Ge_x$ has a diamond-like lattice structure and the lattice constant is given by Vegard's rule:

$$a_{Si_{1-x}Ge_x} = a_{Si} + x\,(a_{Ge} - a_{Si})\ , \qquad (22.1)$$

where x is the germanium fraction and a is the lattice constant. The lattice constant of silicon, a_{Si}, is 0.543 nm, the lattice constant of germanium, a_{Ge}, is 0.566 nm and the lattice mismatch is 4.2%.

When a $Si_{1-x}Ge_x$ layer is grown on a silicon substrate, the lattice mismatch at the interface between the $Si_{1-x}Ge_x$ and the silicon has to be accommodated. This can either be done by compression of the $Si_{1-x}Ge_x$ layer so that it fits to the silicon lattice or by the creation of misfit dislocations at the interface. These two possibilities are illustrated schematically in Fig. 22.1. In the former case, the $Si_{1-x}Ge_x$ layer adopts the silicon lattice spacing in the plane of the growth and hence the normally cubic $Si_{1-x}Ge_x$ crystal is distorted. When $Si_{1-x}Ge_x$ growth occurs in this way, the $Si_{1-x}Ge_x$ layer is under compressive strain and the layer is described as *pseudomorphic*. In the second case, the $Si_{1-x}Ge_x$ layer is unstrained, or relaxed, and the lattice mismatch at the interface is accommodated by the formation of misfit dislocations. These misfit dislocations generally lie in the plane of the interface, as shown in Fig. 22.1, but dislocations can also thread vertically through the $Si_{1-x}Ge_x$ layer.

22.1.1 Critical Thickness

There is a maximum thickness of $Si_{1-x}Ge_x$ that can be grown before relaxation of the strain occurs through the formation of misfit dislocations. This is

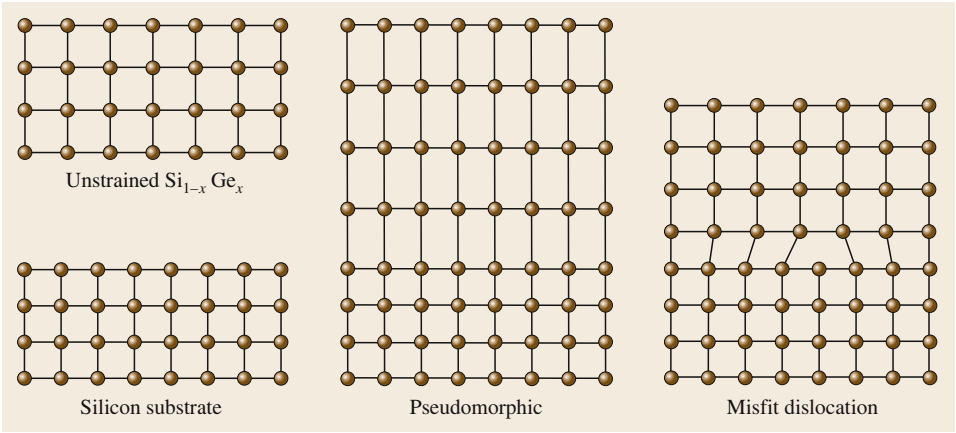


Fig. 22.1 Schematic illustration of pseudomorphic $Si_{1-x}Ge_x$ growth and misfit-dislocation formation

known as the critical thickness of the $Si_{1-x}Ge_x$ layer, and depends strongly on the germanium content, as shown in Fig. 22.2. The original calculations of critical layer thickness were made by *Matthews* and *Blakeslee* [22.10, 11] on the basis of the mechanical equilibrium of an existing threading dislocation. However, measurements of dislocation densities in $Si_{1-x}Ge_x$ showed, in many cases, no evidence of misfit dislocations for $Si_{1-x}Ge_x$ layers considerably thicker than the Matthews–Blakeslee limit. These results were explained by *People* and *Bean* [22.12] who calculated the critical thickness on the assumption that misfit-dislocation generation was determined solely by energy balance. The discrepancy between these two types of calculation can be explained by the observation that strain relaxation in $Si_{1-x}Ge_x$ layers occurs gradually. Layers above the People–Bean curve can be considered to be completely relaxed, whereas layers below the Matthews–Blakeslee curve can be considered to be fully strained. These fully strained layers are termed *stable* and will not relax during any subsequent high temperature processing. Layers lying between the two curves are termed *metastable*; these layers may be free of dislocations after growth, but are susceptible to relaxation during later high-temperature processing.

In practice, a number of additional factors influence the critical thickness of a $Si_{1-x}Ge_x$ layer. Of particular importance to both $Si_{1-x}Ge_x$ HBTs and $Si_{1-x}Ge_x$ MOSFETs, is the effect of a silicon cap layer, which has been shown to increase the critical thickness of the underlying $Si_{1-x}Ge_x$ layer. Figure 22.3 shows a comparison of the calculated critical thickness as a function of germanium percentage for stable $Si_{1-x}Ge_x$ layers with and without a silicon cap. It can be seen that the critical thickness is more than doubled by the presence of the silicon cap.

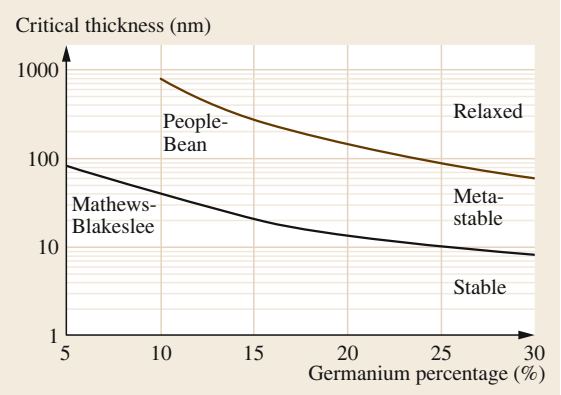


Fig. 22.2 Critical $Si_{1-x}Ge_x$ thickness as a function of germanium percentage. (After *Iyer* et al. [22.2], copyright 1989 IEEE)

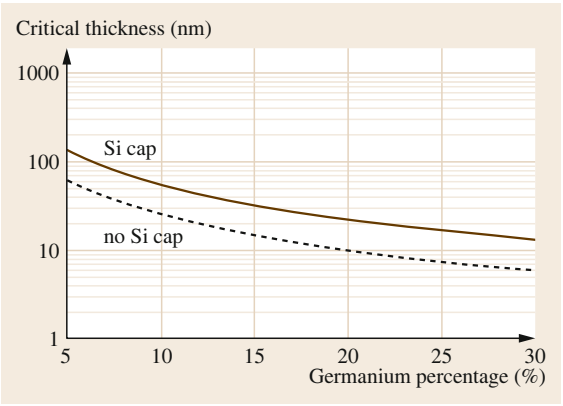


Fig. 22.3 Critical thickness as a function of germanium percentage for stable $Si_{1-x}Ge_x$ layers with and without a silicon cap. (After *Jain* et al. [22.13], copyright 1992 Elsevier)

The presence of misfit dislocations in devices is highly undesirable, since they create generation/recombination centres, which degrade leakage currents when they are present in the depletion regions of devices. Threading dislocations are also highly undesirable, as they can lead to the formation of emitter/collector pipes in $Si_{1-x}Ge_x$ HBTs. When designing $Si_{1-x}Ge_x$ devices, it is important that the $Si_{1-x}Ge_x$ thickness is chosen to give a stable layer, so that dislocation formation is avoided. A base thickness below the silicon cap curve in Fig. 22.3 will ensure a stable layer, which will withstand ion implantation and high-temperature annealing without encountering problems of relaxation and misfit-dislocation generation.

Considerable research has been done on the oxidation of $Si_{1-x}Ge_x$ [22.14, 15]. Both silicon and germanium oxidize so that it is possible to form a mixed oxide ($SiO_2 + GeO_2$). Si oxidation is more favorable [22.16] so that as long as the Si flux arriving at the oxidizing interface is sufficient Si will replace Ge in the oxide, no Ge oxide remains, and a Ge-rich layer forms at the oxide/$Si_{1-x}Ge_x$ interface. However, oxidation at low temperature [22.17], high pressure [22.16], or with high Ge content in the $Si_{1-x}Ge_x$ layer [22.18], produces a mixed oxide and the Ge pile-up is reduced or eliminated. *Rabie* et al. [22.19] have developed a kinetic model describing the details of the oxidation process. Given the process conditions, the model predicts oxide thickness and composition as well as the profile of Si/Ge in the SiGe layer.

The pile-up of germanium makes it difficult to achieve low values of interface state density in oxidised $Si_{1-x}Ge_x$ layers. Even when the pile-up is suppressed, the interface state density is still roughly two orders of magnitude higher than it is for Si. This is likely due to

the higher fraction of Si atoms in suboxide states and strain relaxation during the oxidation process [22.20]. In some experiments, good oxide quality and low interface trap and fixed charge densisties were achieved and the Ge pile-up suppressed by oxidation at very low temperatures [22.21]. Nonetheless, most practical process flows do not involve direct oxidation of $Si_{1-x}Ge_x$ layers, particularly in $Si_{1-x}Ge_x$ MOS technologies. In $Si_{1-x}Ge_x$ MOSFETs, a silicon cap is often included above the $Si_{1-x}Ge_x$ layer that can be oxidised to create the gate oxide. Alternatively, there has been considerable recent work exploring the deliberate use of the Ge-rich layer by oxidizing SiGe on SOI. This has been proposed for applications in fully depleted MOSFETs and for photonics applications (see [22.22] and references therein).

22.1.2 Band Structure

$Si_{1-x}Ge_x$ alloys have a smaller band gap than silicon partly because of the larger lattice constant and partly because of the strain. Figure 22.4 shows the variation of band gap with germanium percentage for strained and unstrained $Si_{1-x}Ge_x$. The bandgap of strained Si depends on the type as well as magnitude of strain and is therefore strongly affected by the substrate. The data shown in Fig. 22.4 are for SiGe on (100) Si. For strained SiGe on Ge the bandgap would be an increasing function of Ge fraction over almost the entire range from Si to Ge. It can be seen that the effect of strain is considerable. For 10% germanium, the reduction in band gap compared with silicon is 92 meV for strained $Si_{1-x}Ge_x$, compared with 50 meV for unstrained $Si_{1-x}Ge_x$. The variation of band gap E_G with germanium content for unstrained $Si_{1-x}Ge_x$ can be described by the following empirical equation

$$E_G(x) = \begin{cases} 1.17 - 0.47x + 0.24x^2\,, & x < 0.85 \\ 5.88 - 9.58x + 4.43x^2\,, & x > 0.85 \end{cases} \quad (22.2)$$

The transition at $x = 0.85$ is due to the conduction band minima switching from the [100] direction associated with Si to the [111] direction associated with Ge. For strained $Si_{1-x}Ge_x$ the appropriate equation is

$$E_G(x) = 1.17 - 0.94x + 0.34x^2 \quad (22.3)$$

This equation is valid over the entire range of Ge fraction.

In addition to strain, heavy doping will also result in significant bandgap narrowing. Figure 22.4 contains data showing the bandgap of SiGe layers doped with

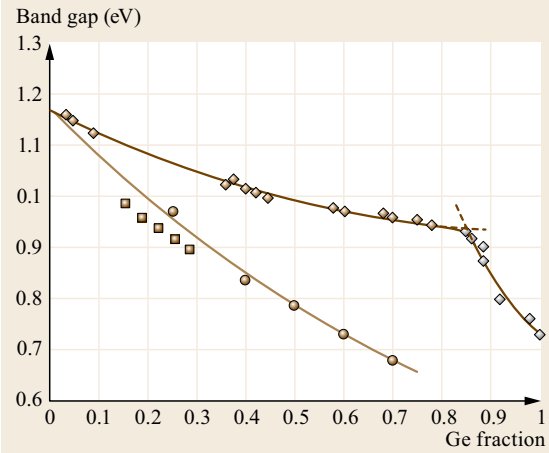


Fig. 22.4 Band gap as a function of germanium percentage for unstrained (*diamonds* [22.1]), strained undoped (*circles* [22.23]), and strained heavily doped $Si_{1-x}Ge_x$ (*squares* [22.24]). The *lines* are empirical best fits corresponding to (22.2) and (22.3). Data from [22.1] and [22.24] were corrected to 90 K using the methodology of [22.25]

B in the range of $4.3-5.8\times10^{19}$ cm^{-3}. These data were obtained by fitting measured device characteristics from a Si/SiGe HBT.

From the graph, we note that the effect of strain is significantly more than that of doping. For example, at 25% Ge the bandgap is reduced by approximately 97 meV due to strain and a further 50 meV due to heavy doping. The narrowing due to strain increases with increasing Ge fraction. The narrowing due to doping does not appear to be significantly affected by Ge fraction. A theoretical model for bandgap narrowing is discussed below (see Sect. 22.1.8).

The band alignment for compressively strained $Si_{1-x}Ge_x$ on unstrained silicon is illustrated schematically in Fig. 22.5. This band alignment is referred to as *type I*, and the majority of the band offset at the heterojunction interface occurs in the valence band, with only a small offset in the conduction band. Different band alignments can be obtained by engineering the strain in the substrate and the grown layer in different ways. For example, *type II* band alignments can be obtained by growing tensile-strained silicon on top of unstrained $Si_{1-x}Ge_x$. This arrangement gives large conduction- and valence-band offsets and is used in strained silicon MOSFETs.

Figure 22.6 shows the variation of valence-band offset, ΔE_V, conduction-band offset, ΔE_C, and band-gap narrowing, ΔE_G, with germanium content. Once again, it should be emphasized that this is for SiGe on (100)

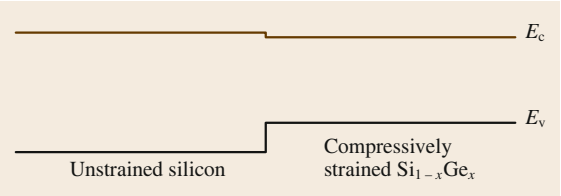


Fig. 22.5 Schematic illustration of the band alignment obtained for a compressively strained $Si_{1-x}Ge_x$ layer grown on an unstrained silicon substrate

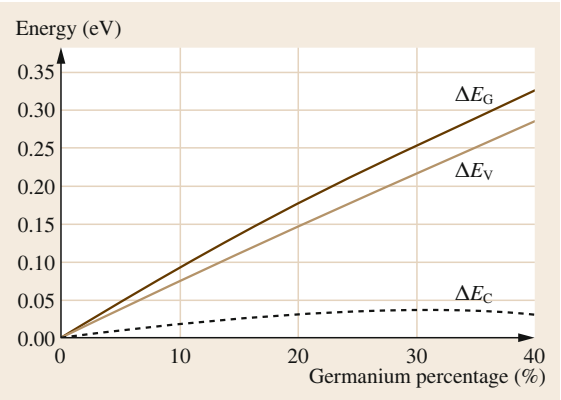


Fig. 22.6 Valence- and conduction-band offsets as a function of germanium percentage for strained $Si_{1-x}Ge_x$ grown on unstrained silicon. (After *Poortmans* et al. [22.26], copyright 1993 Elsevier)

Si and that the behavior would be different for other substrates. It can be seen that the majority of the band offset occurs in the valence band. For example for 10% germanium, the valence-band offset is 0.073 eV, compared with 0.019 eV for the conduction-band offset. The conduction-band offset can therefore be neglected for most practical purposes.

22.1.3 Dielectric Constant

The dielectric constant of $Si_{1-x}Ge_x$ can be obtained by linear interpolation between the known values for silicon and germanium [22.26] using the following equation

$$\varepsilon(x) = 11.9(1 + 0.35x) \,. \qquad (22.4)$$

22.1.4 Density of States

While the density of states in the conduction band in $Si_{1-x}Ge_x$ is generally assumed to be the same as that in silicon, there is some evidence in the literature to suggest that the density of states in the valence band is considerably smaller. *Manku* and *Nathan* [22.27, 28] have calculated the *E–k* diagram for strained $Si_{1-x}Ge_x$ and shown that the density-of-states hole mass is significantly lower, by a factor of approximately three at 30% germanium. There is some experimental evidence to support this calculation. For example, freeze-out of holes in p-type $Si_{1-x}Ge_x$ has been reported to occur at higher temperatures than in p-type silicon [22.29] and enhancements in the majority-carrier, hole mobility have been reported for p-type $Si_{1-x}Ge_x$ [22.30]. Using the calculated values of hole density of states of *Manku* and *Nathan* [22.27, 28], the hole concentration can be calculated as a function of Fermi-level position. These results are shown in Fig. 22.7 for $Si_{1-x}Ge_x$ with four different germanium contents. It can be seen that the Fermi level moves deeper into the valence band as the germanium concentration increases. Figure 22.8 shows the ratio of the calculated density of states in the valence band for $Si_{1-x}Ge_x$ to that for silicon as a function of germanium content. It is clear that the density of states in the valence band for $Si_{1-x}Ge_x$ is significantly lower than that for silicon at germanium contents of practical interest.

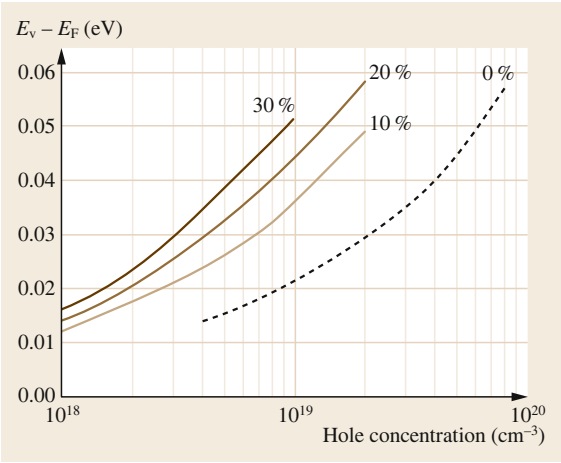


Fig. 22.7 Fermi-level position as a function of hole concentration for $Si_{1-x}Ge_x$ with four different germanium concentrations. (After *Iyer* et al. [22.2], copyright 1989 IEEE)

22.1.5 Majority-Carrier Mobility in Strained $Si_{1-x}Ge_x$

Values of in-plane and out-of-plane low-field mobility in strained $Si_{1-x}Ge_x$ grown on (100) Si have been calculated by *Fischetti* et al. [22.31], and are shown in Fig. 22.9. There is some uncertainty in the chosen values of alloy scattering parameters used in the calculations, but nevertheless the results are representative of current understanding. These results show a large

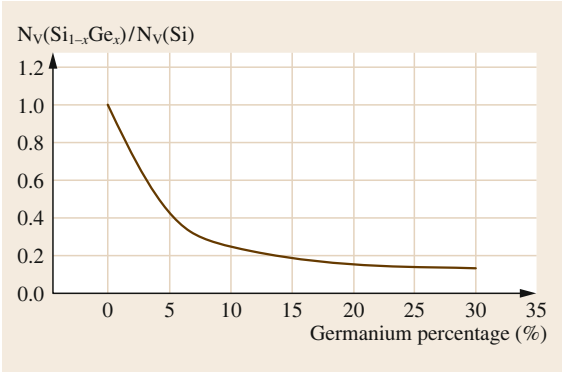


Fig. 22.8 Ratio of density of states in the valence band for $Si_{1-x}Ge_x$ to that for Si as a function of germanium percentage. (After *Poortmans* [22.29], copyright 1993 University of Leuven)

enhancement of low-field hole mobility for $Si_{1-x}Ge_x$ compared with unstrained silicon, but only a modest enhancement of low-field electron mobility. These results indicate that strained $Si_{1-x}Ge_x$ channels can be used to significantly improve the mobility of p-channel MOSFETs, but little benefit is obtained for n-channel MOSFETs. For this reason, industry focus has moved away from channels realised in $Si_{1-x}Ge_x$ to channels realised in tensile-strained silicon, as discussed below.

22.1.6 Majority-Carrier Mobility in Tensile-Strained Si on Relaxed $Si_{1-x}Ge_x$

Tensile-strained silicon can be produced by growing a thin silicon layer on top of a relaxed $Si_{1-x}Ge_x$ virtual substrate. Figure 22.10 shows a typical virtual substrate for a surface-channel MOS transistor. A graded $Si_{1-x}Ge_x$ layer is grown on top of the silicon substrate with the Ge content varying from 0 to 30%. Misfit dislocations will form in this layer, but the majority of dislocations will be in the plane of the $Si_{1-x}Ge_x$ layer and only a small percentage will propagate vertically to the surface of the layer. A 30%-relaxed $Si_{1-x}Ge_x$ buffer is then grown followed by a thin tensile-strained $Si_{1-x}Ge_x$ layer in which the channel is fabricated. The key to any virtual substrate growth process is the minimisation of dislocation propagation to the surface of the wafer.

Figure 22.11 shows typical values of effective electron mobility obtained from measurements on n-channel MOS transistors for a $Si_{1-x}Ge_x$ virtual substrate with 30% Ge [22.6]. For the surface-channel strained Si device, the effective mobility is enhanced by 80% compared with the Si control transistor due to the tensile strain in the $Si_{1-x}Ge_x$ layer.

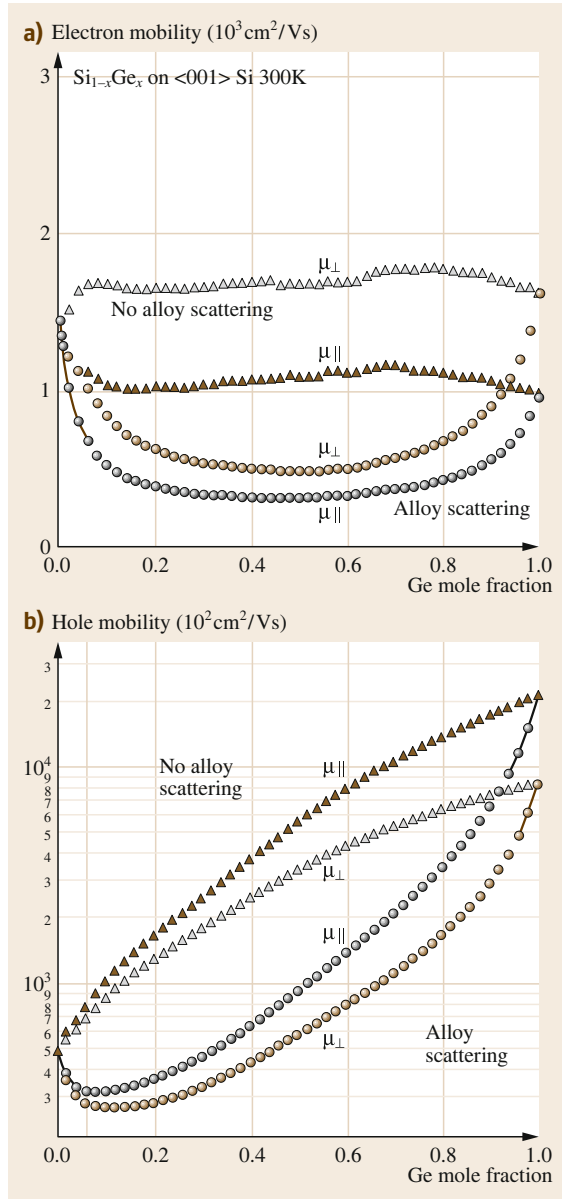


Fig. 22.9a,b Calculated 300 K electron **(a)** and hole **(b)** in-plane and out-of-plane low-field mobilities in strained $Si_{1-x}Ge_x$ grown on (100) Si. (After *Fischetti* and *Laux* [22.31], copyright 1996 American Institute of Physics)

Enhanced hole mobility can also be obtained in tensile-strained Si grown on a $Si_{1-x}Ge_x$ virtual substrate, though higher germanium contents are needed to obtain a significant mobility enhancement. Figure 22.12 shows typical values of effective hole mobility in strained Si for Ge contents between 35 and 50% [22.32]. The effective mobility of the strained Si

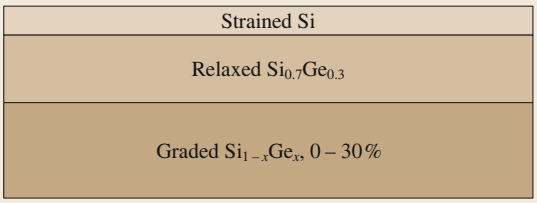


Fig. 22.10 Schematic illustration of a typical tensile-strained Si layer grown on top of a $Si_{1-x}Ge_x$ virtual substrate

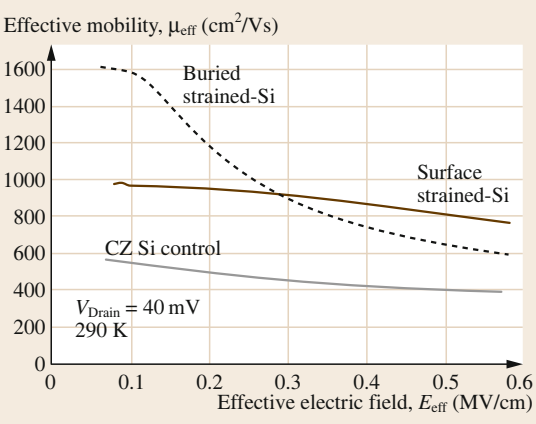


Fig. 22.11 Effective electron mobility as a function of effective electric field for strained Si MOSFETs fabricated on a 30% $Si_{1-x}Ge_x$ virtual substrate. (After *Welser* et al. [22.6], copyright 1994 IEEE)

device is enhanced by 100% compared with the Si control device.

22.1.7 Minority-Carrier Mobility in Strained $Si_{1-x}Ge_x$

$Si_{1-x}Ge_x$ HBTs are minority-carrier devices and hence values of the minority-carrier mobility of more interest than the majority-carrier mobility. Unfortunately very few measurements of minority-carrier mobility have been made in $Si_{1-x}Ge_x$. *Poortmans* [22.29] inferred values of minority-carrier mobility from measurements on $Si_{1-x}Ge_x$ HBTs and found an enhancement in mobility compared with silicon by a factor of 1.2–1.4 for base doping concentrations in the range 5×10^{18}–5×10^{19} cm^{-3}. Given the scarcity of measured data on minority-carrier mobility and density of states in $Si_{1-x}Ge_x$, the most reliable way of calculating the expected gain improvement in a $Si_{1-x}Ge_x$ HBT is to use data directly obtained from measurements on $Si_{1-x}Ge_x$ HBTs. The gain enhancement in a $Si_{1-x}Ge_x$ HBT is determined by the ratio of the product $N_C N_V D_{nb}$ in

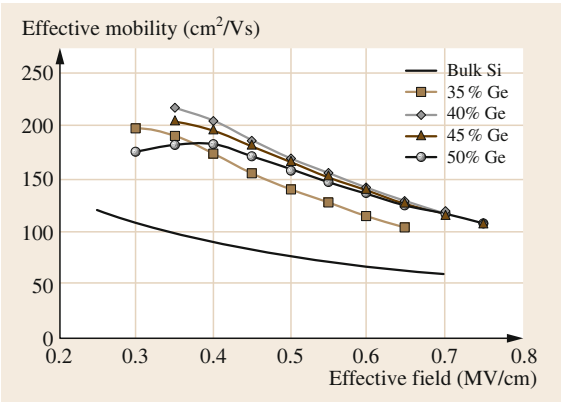


Fig. 22.12 Effective hole mobility as a function of effective electric field for strained Si MOSFETs fabricated on a $Si_{1-x}Ge_x$ virtual substrate with Ge contents in the range 35–50%. (After *Leitz* et al. [22.32], copyright 2002 IEEE)

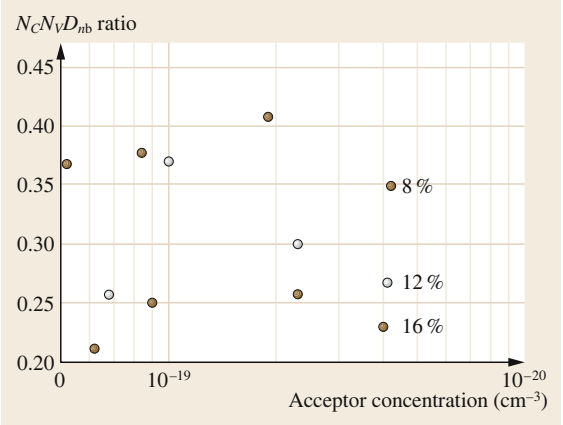


Fig. 22.13 Measured values of the ratio of $N_C N_V D_{nb}$ in $Si_{1-x}Ge_x$ to that in Si as a function of acceptor concentration. (After *Poortmans* [22.29], copyright 1993 University of Leuven)

$Si_{1-x}Ge_x$ and Si, together with the band-gap narrowing due to the strained $Si_{1-x}Ge_x$. Figure 22.13 shows a graph of this $N_C N_V D_{nb}$ ratio as a function of acceptor concentration for three values of germanium content. It can be seen that for germanium contents of practical interest, in the range 11–16%, this ratio has a value of around 0.25.

22.1.8 Apparent Band-Gap Narrowing in $Si_{1-x}Ge_x$ HBTs

In $Si_{1-x}Ge_x$ HBTs, the apparent band-gap narrowing is often quoted, which combines the effect of the band-gap reduction and the effect of high doping. *Poortmans* et al. [22.26] have developed a theoretical approach that

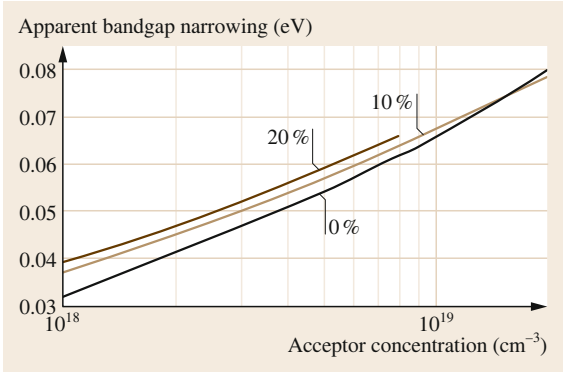


Fig. 22.14 Apparent band-gap narrowing as a function of acceptor concentration for $Si_{1-x}Ge_x$ with three different germanium concentrations. (After *Poortmans* et al. [22.26], copyright 1993 Elsevier) ▶

has been shown to be in reasonable agreement with experiment. Figure 22.14 shows the apparent band-gap narrowing in $Si_{1-x}Ge_x$ as a function of acceptor concentration for three values of germanium content. At low acceptor concentrations, the apparent band-gap narrowing in $Si_{1-x}Ge_x$ is slightly higher than that in silicon, but at acceptor concentrations in the range $1-2\times 10^{19}\,cm^{-3}$, the apparent band-gap narrowing is approximately the same. This latter doping range is the base doping range that is of practical interest for $Si_{1-x}Ge_x$ HBTs. This is reasonably close to experimental data such as that reported by [22.24] (see Sect. 22.1.2 and Fig. 22.4).

22.2 Optical Properties of SiGe

Interest in the optical properties of SiGe stems from the desire to design silicon-based optoelectronic devices as well as the usefulness of many optical techniques in material analysis. The optical properties of bulk SiGe provides an important starting point for any attempt at in-depth understanding, however, nearly all real applications of SiGe involve the use of thin, strained layers. Beyond this, the formation of SiGe quantum structures within silicon devices has remained one of the most promising methods by which device engineers hope to improve the largely unimpressive optical behaviour of silicon that is brought about by its indirect band gap. The growth of quantum wells, quantum wires and quantum dots in the Si/SiGe system has been extensively explored [22.33–38]. However, there has as yet been little success at using Si/SiGe quantum structures to produce efficient silicon-based light emitters, although there have been impressive attempts [22.38]. Perhaps the most promising devices based on SiGe quantum wells are near-infra-red photodetectors in which the SiGe can be used to enhance sensitivity at the optical-communications wavelengths [22.39–43]. More futuristic applications of SiGe quantum wells include devices based on transitions between the confined energy levels in quantum wells. Devices based on these *inter-subband* transitions include quantum-well infrared photodetectors (QWIPs) [22.44, 45] and quantum cascade lasers [22.46–48].

22.2.1 Dielectric Functions and Interband Transitions

A range of 1 μm-thick $Si_{1-x}Ge_x$ films grown by MBE on Si(100) substrates have been studied by spectroscopic ellipsometry to yield their complex dielectric functions at room temperature [22.49]. Both the real (ε_1) and imaginary (ε_2) parts of the measured dielectric function for Ge compositions of 0, 0.2, 0.4, 0.6, 0.8 and 1.0 are shown in Figs. 22.15a and 22.15b, respectively.

In Fig. 22.15b the absorption structures observed at 3.4 and 4.2 eV in the spectra shown for silicon originate from direct band-to-band transitions at various regions in the Brillouin zone of silicon. The structure seen around 3.4 eV is due to E_0', E_1 and $E_1+\Delta_1$ interband transition edges, whereas the structure at 4.2 eV is due to $E_2(X)$ and $E_2(\Sigma)$ edges [22.49]. The evolution of each of these transition edges for the full range of SiGe compositions is shown in Fig. 22.16.

The quadratic fits shown in Fig. 22.16 are as follows [22.49],

$$E_0'(x) = 3.337 - 0.348x + 0.222x^2\,, \qquad (22.5)$$

$$E_1(x) = 3.398 - 1.586x + 0.27x^2\,, \qquad (22.6)$$

$$E_1+\Delta_1(x) = 3.432 - 1.185x + 0.065x^2\,, \qquad (22.7)$$

$$E_2(X)(x) = 4.259 - 0.052x + 0.084x^2\,, \qquad (22.8)$$

$$E_2(\Sigma)(x) = 4.473 - 0.139x + 0.072x^2\,. \qquad (22.9)$$

22.2.2 Photoluminescence

The emission properties of semiconductor structures are of fundamental interest to scientists as well as being an important analytical technique for engineers. In general, the features of low-temperature photoluminescence spectra are very dependent on the specific conditions under which materials are grown and treated. This is because emission energies and emission rates are of-

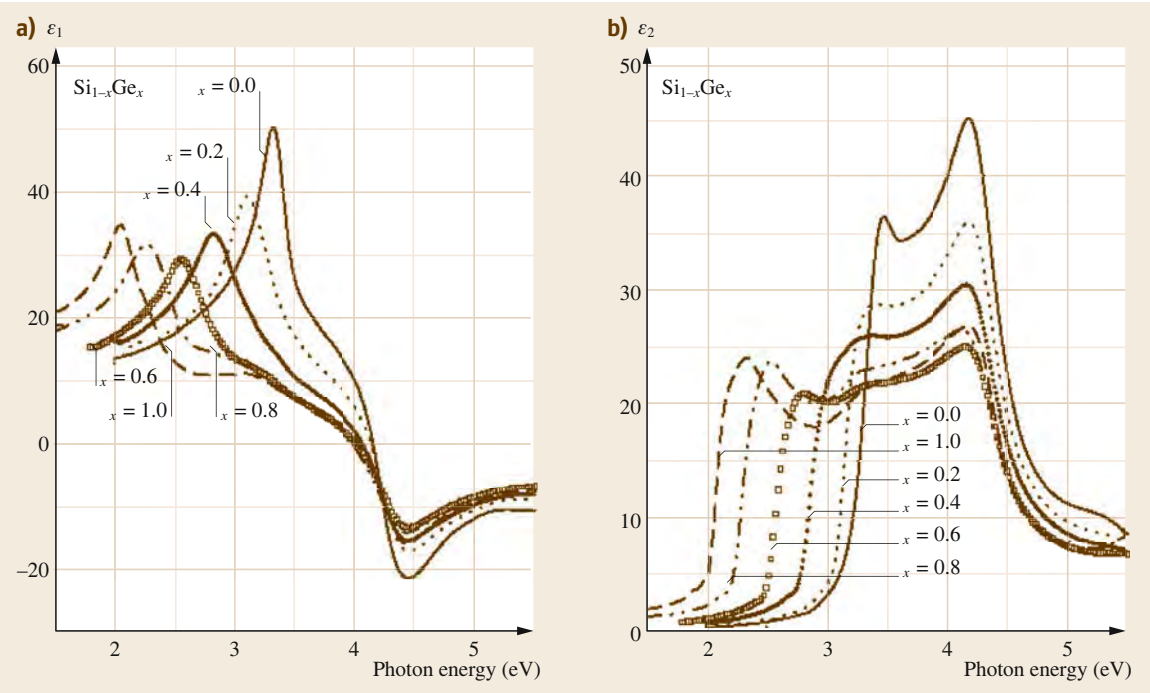


Fig. 22.15a,b Real (a) and imaginary (b) parts of the dielectric functions of relaxed $Si_{1-x}Ge_x$ alloys with composition x indicated in the legend. (After *Bahng* et al. [22.49], copyright 2001, American Physical Society)

ten sensitive to even small variations of impurity or defect densities, as well as variations in strain or composition. A brief examination of low-temperature photoluminescence spectra is nearly always sufficient to allow a simple qualitative assessment of material quality; alternatively, detailed analysis can permit a broad range of material or structural parameters to be assessed or determined. No two photoluminescence spectra are the same. In this section and the section that follows on photoluminescence studies of Si/SiSe quantum wells, we will present a range of spectra that represent most of the key features that have been observed.

Weber and *Alonso* [22.50] have provided a very useful study of the near-band-gap photoluminescence of bulk SiGe alloys. In their study bulk SiGe samples are cut from nominally undoped polycrystalline ingots prepared by a zone-levelling technique. Figure 22.17 shows the photoluminescence spectra for a range of compositions. Samples were excited using the 514 nm line of an argon ion laser and the sample temperature was 4.2 K.

Excitonic emission lines are a strong feature of photoluminescence spectra when the thermal energy of the semiconductor is less than the exciton binding energy. Each spectrum featured in Fig. 22.17, across the full range of SiGe compositions, show similar *excitonic* features.

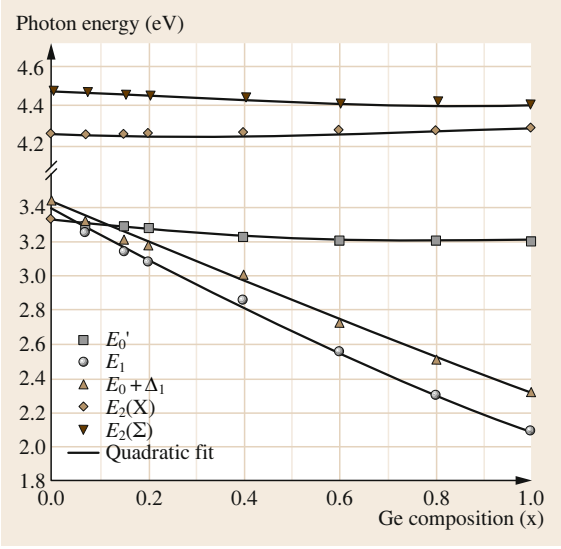


Fig. 22.16 Evolution of E_0', E_1, $E_1 + \Delta_1$, $E_2(X)$ and $E_2(\Sigma)$ transition energies for relaxed $Si_{1-x}Ge_x$ with composition x. (After *Bahng* et al. [22.49], copyright 2001, American Physical Society)

In most spectra the most pronounced peak is the no-phonon (X^{NP}) line caused by the optical recombination of excitons bound to shallow impurities. In the

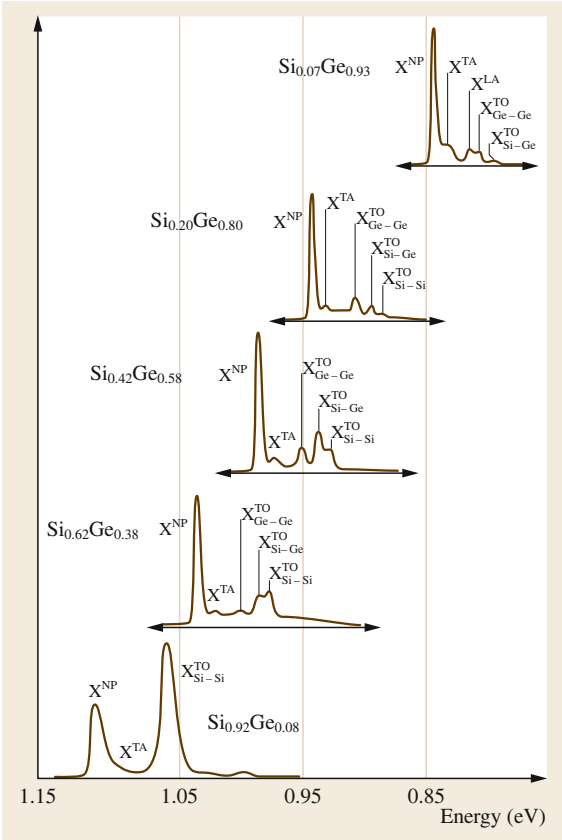


Fig. 22.17 Near-band-gap photoluminescence spectra for several bulk SiGe samples. (After *Weber* et al. [22.50], copyright 1989, American Physical Society)

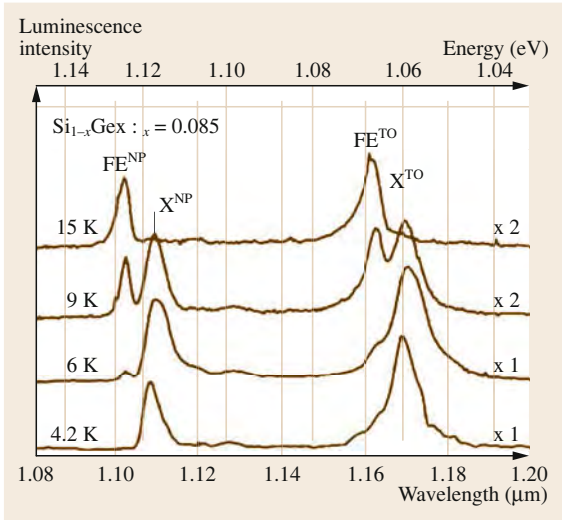


Fig. 22.18 Photoluminescence spectra of a bulk $Si_{0.915}Ge_{0.085}$ sample at different temperatures. (After *Weber* et al. [22.50], copyright 1989, American Physical Society)

case of the no-phonon line, momentum is conserved through interaction with the binding impurity. There are many candidate atoms for these shallow impurities and with B, P and As having binding energies of 4.2, 5.0 and 5.6 meV, respectively [22.51]. These bound energy states will tend to dominate luminescence spectra for doped samples and will always tend to be present in nominally undoped samples. The no-phonon line is accompanied by transverse-optical (X^{TO}) or transverse acoustic (X^{TA}) phonon replicas that are created as photon emission is accompanied by the momentum-conserving creation of lattice vibrations in Si–Si, Si–Ge or Ge–Ge bonds.

Figure 22.18 shows how the spectrum from a bulk $Si_{0.915}Ge_{0.085}$ sample develops with increasing temperature [22.50]. With increasing temperature the X^{NP} line thermalises to leave the free-exciton (FE^{NP}) line. Here, emission from nominally free excitons is greatly enhanced for the alloy samples by local fluctuations in composition that provide momentum-conserving scattering centres [22.52].

At higher temperatures the free-exciton line is also thermalised and all fine structure is lost, at temperatures around 25 K broad luminescence bands are commonly observed (Fig. 22.19) [22.50]. These deep luminescence bands are difficult to assign and have been ascribed to impurities, structural defects and, as in the case of the line presented in Fig. 22.19, potential wells formed by alloy fluctuations [22.50].

Weber and *Alonso* [22.50] use their data to provide analytical expressions for both the X^{NP} and L bands for bulk $Si_{1-x}Ge_x$ in the range $0 \le x \le 0.85$ as follows,

$$E_{gx}^{(x)}(x) = 1.155 - 0.43x + 0.206x^2 \text{ eV} , \quad (22.10)$$

$$E_{gx}^{L}(x) = 2.010 - 1.270x \text{ eV} . \quad (22.11)$$

At low temperatures narrow excitonic luminescence features are indicative of defect-free material, and in this way low-temperature photoluminescence becomes a good qualitative tool with which material quality can be assessed.

22.2.3 SiGe Quantum Wells

Figure 22.20 shows the first excitonic luminescence spectra from a Si/SiGe multiple quantum well grown by atmospheric-pressure CVD [22.35]. As we can see, many of the features seen in the photoluminescence spectra of quantum-well samples are similar to those seen from the bulk samples described in the previous section. Again, the most pronounced peak is the no-phonon (NP) line and this is accompanied by phonon

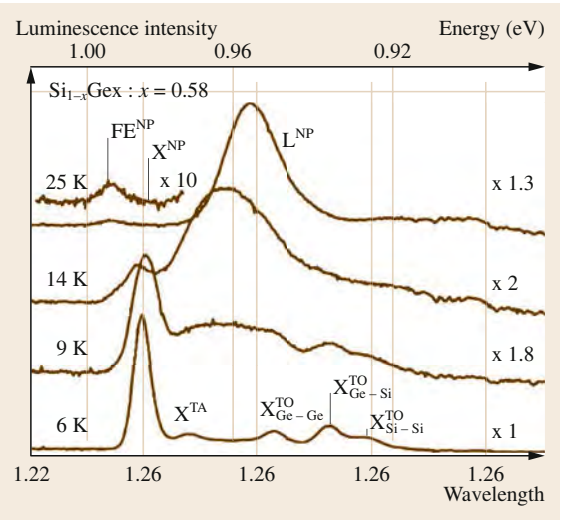


Fig. 22.19 Photoluminescence spectra of a bulk $Si_{0.42}Ge_{0.58}$ sample at different temperatures. (After *Weber* and *Alonso* [22.50], copyright 1989, American Physical Society)

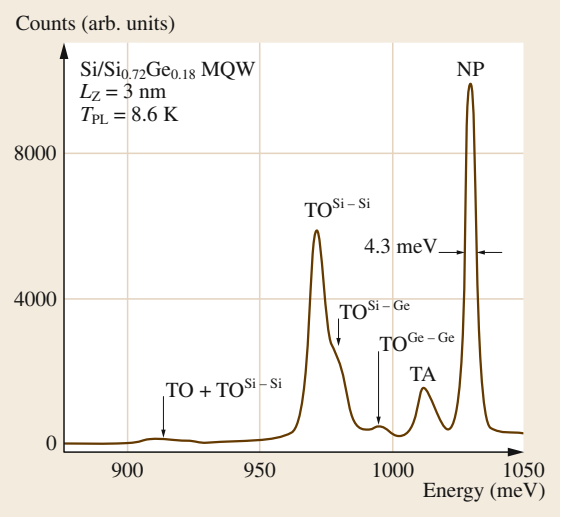


Fig. 22.20 Excitonic photoluminescence spectrum of a SiGe quantum well. (After *Grutzmacher* et al. [22.35], copyright 1993, AVS)

replicas, including, impressively, a two-phonon replica of the NP line $\left(\mathrm{TO} + \mathrm{TO}^{\mathrm{Si-Si}}\right)$. The most significant difference between bulk and quantum well spectra is the energy positions of the excitonic features as these are shifted by quantum confinement effects.

Robbins et al. [22.33], have provided one of the most detailed studies of near-band-gap photoluminescence from pseudomorphic SiGe layers and provide analytical expressions for all factors pertaining to the energy positions of the excitonic energy gap for $Si_{1-x}Ge_x$ quantum wells in the range $0 < x < 0.24$. The effects of alloying, confinement, band offsets, alignment type and exciton binding energy are all taken into account. Samples used in the study were grown by low-pressure CVD at 920 °C; a typical set of photoluminescence spectra are shown in Fig. 22.21.

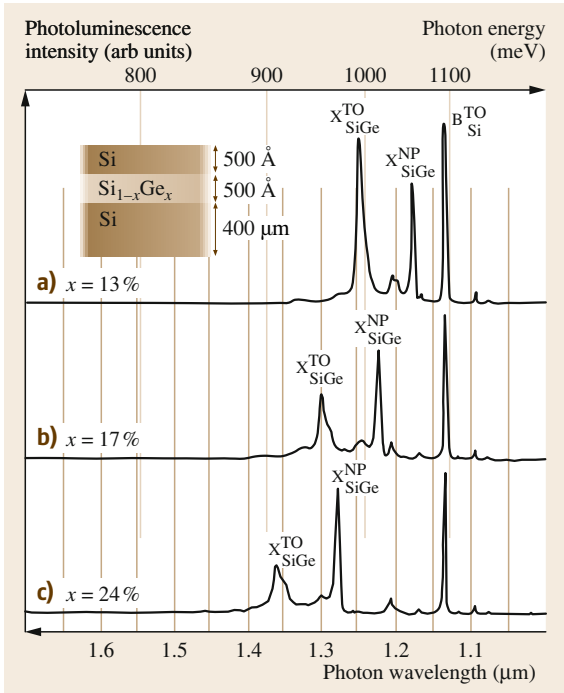


Fig. 22.21 4.2 K photoluminescence spectra from layers with the nominal structure shown in the *inset* (514 nm Ar-ion laser excitation). (After *Robbins* et al. [22.33], copyright 1992, American Physical Society)

The exciton band gap at 4.2 K is considered for thick (50 nm) strained layers (E_X^S) where the energies are not affected by quantum shifts. The following expression is derived

$$E_X^S(x) = 1.155 - 0.874x + 0.376x^2 \text{ eV}, \ (x < 0.25) \,. \tag{22.12}$$

Here the presence of strain is responsible for the differences from the expression obtained for bulk samples (22.10). An expression for the exciton binding energy $E_B^C(x)$ is theoretically derived for the cubic alloy and the following quadratic expression is fitted

$$E_B^C(x) \approx 0.0145 - 0.022x + 0.020x^2 \text{ eV}, \ (x < 0.25) . \quad (22.13)$$

A strain-corrected expression is also provided but this is found to modify (22.13) only slightly. Thus by adding (22.12) and (22.13) an expression for the band gap can be obtained

$$E_C - E_V \approx 1.17 - 0.896x + 0.39x^2 \text{ eV}, \ (x < 0.25) . \quad (22.14)$$

22.3 Growth of Silicon-Germanium

Over the past ten years and more there have been rapid developments in techniques for the growth of Si and $Si_{1-x}Ge_x$ epitaxial layers at low temperatures. This has been made possible by a number of changes in the design of epitaxy equipment and by improvements to growth processes. There are two main prerequisites for the growth of epitaxial layers at low temperature:

- Establishment of a clean surface prior to growth [22.53–56]
- Growth in an ultra-clean environment [22.57–59].

The removal of oxygen and carbon is the main problem in establishing a clean surface prior to growth. A clean silicon surface is highly reactive and oxidises in air even at room temperature. The secret of low-temperature epitaxial growth is therefore the removal of this native oxide layer and the maintenance of a clean surface until epitaxy can begin. Two alternative approaches to pre-epitaxy surface cleaning have been developed, as described below.

22.3.1 In situ Hydrogen Bake

The concept that underlies this surface clean is the controlled growth of a thin surface oxide layer, followed by its removal in the epitaxy reactor using a hydrogen bake. The controlled growth of the surface oxide layer is generally achieved using a Radio Corporation of America (RCA) clean [22.53] or a variant [22.54]. The oxide created by the RCA clean is removed in the reactor using an in-situ bake in hydrogen for around 15 min at a temperature in the range 900–950 °C. The temperature required to remove the native oxide depends on the thickness of the oxide, which is determined by the severity of the surface clean.

22.3.2 Hydrogen Passivation

An alternative approach to pre-epitaxy cleaning is to create an oxide-free surface using an ex-situ clean and then move quickly to epitaxial growth before the native oxide can grow. The aim of the ex-situ clean is to produce a surface that is passivated by hydrogen atoms bonded to dangling bonds from silicon atoms on the surface. When the wafers are transferred in the epitaxy reactor, the hydrogen can be released from the surface of the silicon very quickly using a low-temperature bake or even in the early stages of epitaxy without any bake. *Meyerson* [22.55] has reported that hydrogen desorbs at 600 °C at a rate of a few monolayers per second, so the hydrogen passivation approach allows epitaxial layers to be grown at low temperatures without the need for a high-temperature bake. The hydrogen-passivated surface is stable for typically 30 min after completion of the ex-situ cleaning [22.56].

22.3.3 Ultra-Clean Epitaxy Systems

Having produced a clean hydrogen-passivated silicon surface, it is clearly important to maintain the state of this surface in the epitaxy system. This necessitates the use of low-pressure epitaxy systems if epitaxial growth at low temperatures is required. Figure 22.22 summarises the partial pressures of oxygen and water vapour that need to be achieved in an epitaxy system if an oxide-free surface is to be maintained at a given temperature [22.57, 58]. This figure shows that epitaxial growth at low temperature requires low partial pres-

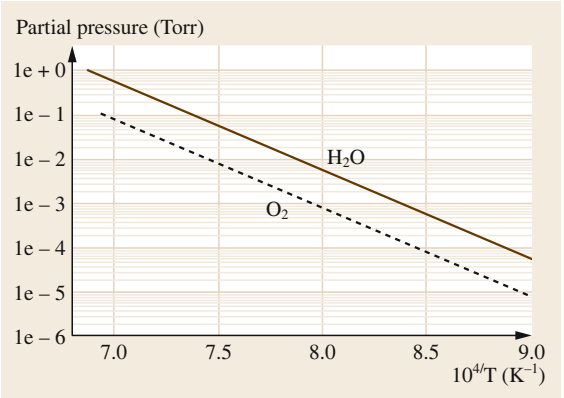


Fig. 22.22 Conditions for oxide formation in an epitaxy system. Note that 1 atm = 1.113 bar = 760 Torr = 1.113×10^5 Pa. (After *Smith* and *Ghidini* [22.57, 58], copyright Electrochemical Society 1984)

sures of oxygen and water vapour, which of course can be achieved by reducing the pressure in the epitaxy system. Research [22.59] has shown that a pressure below 30 Torr is needed to achieve silicon epitaxial growth below 900 °C.

22.3.4 $Si_{1-x}Ge_x$ Epitaxy

The growth of $Si_{1-x}Ge_x$ epitaxial layers can be achieved over a wide range of temperatures using low-pressure chemical vapour deposition (LPCVD) [22.59] or ultra-high-vacuum chemical vapour deposition (UHV-CVD) [22.61, 62]. The gas used to introduce the germanium into the layers is germane, GeH_4. The influence of germanium on the growth rate is complex, as illustrated in Fig. 22.23. At temperatures in the range 577–650 °C a peak in the growth rate is seen. At low germanium contents, the growth rate increases with germanium content, whereas at high germanium content, the growth rate decreases with germanium content. In the low-temperature regime it has been proposed that hydrogen desorption from the surface is the rate-limiting step. In $Si_{1-x}Ge_x$ this occurs more easily at germanium sites than at silicon sites and hence the growth rate increases with germanium content [22.44]. As the germanium content increases, the surface contains more and more germanium and less and less hydrogen. The rate-limiting step then becomes the adsorption of germane or silane. *Robbins* et al. [22.63] proposed that the sticking coefficient for germane or silane was lower at germanium sites. This would slow the adsorption rate as the germanium content increased and hence slow the growth rate.

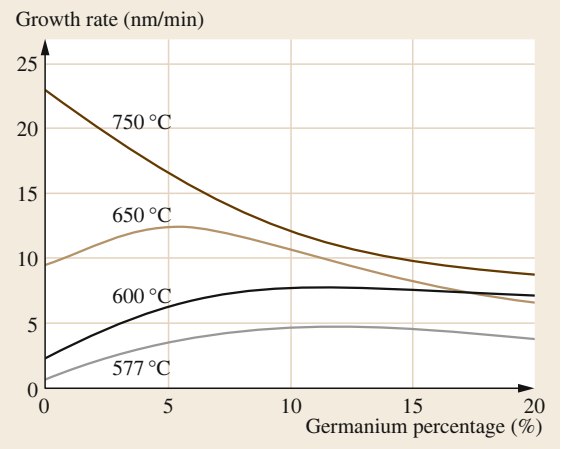


Fig. 22.23 Growth rate of $Si_{1-x}Ge_x$ as a function of germanium percentage for temperatures in the range 577–750 °C. (After *Racanelli* and *Greve* [22.60], copyright 1990, American Institute of Physics)

22.3.5 Selective $Si_{1-x}Ge_x$ Epitaxy

Selective epitaxy is the growth of a single-crystal layer in a window, with complete suppression of growth elsewhere, and can be achieved in a number of different ways. The most common method of achieving both selective Si and $Si_{1-x}Ge_x$ epitaxy is by introducing chlorine or HCl into the growth chamber. This can either be done by adding chlorine or HCl as a separate gas or by using a growth gas that contains chlorine, for example dichlorosilane, SiH_2Cl_2. With chlorine chemistry, selective growth of silicon and $Si_{1-x}Ge_x$ can be achieved to both silicon dioxide and silicon nitride.

Chlorine is reported to have two effects that lead to selective growth. First it increases the surface mobility of silicon and germanium atoms, so that atoms deposited on the oxide or nitride layer are able to diffuse across the surface to the window where the growth is occurring. Second it acts as an etch [22.59] and hence can remove silicon or germanium atoms deposited on the oxide or nitride. The strength of the etching action increases with chlorine content and, if the chlorine content is too high, etching of the substrate will occur instead of epitaxial growth.

A typical growth process for selective silicon epitaxy would use silane and a few percent of HCl [22.59]. The growth rate for this process is shown in Fig. 22.24, and compared with the growth rate for dichlorosilane and silane epitaxy. It can be seen that the activation energy for the silane-plus-HCl process is very similar to that for the dichlorosilane process, indicating that the growth mechanisms are similar. One disadvantage of chlorine-based growth processes over the silane

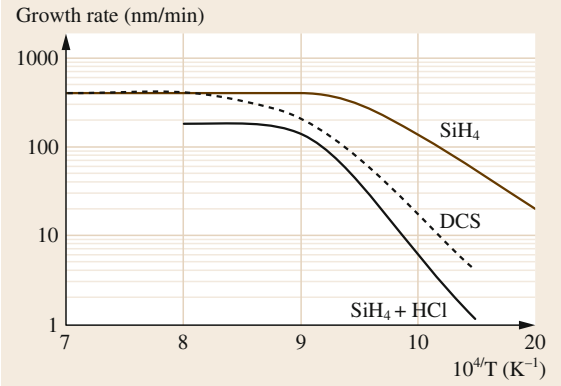


Fig. 22.24 Silicon growth rate as a function of reciprocal temperature for three different growth gases: 40 sccm of SiH_4, 80 sccm of dichlorosilane and 20 sccm of SiH_4 with 2 sccm of HCl. The hydrogen flow was 2 slm. (After *Regolini* et al. [22.59], copyright 1989 American Institute of Physics)

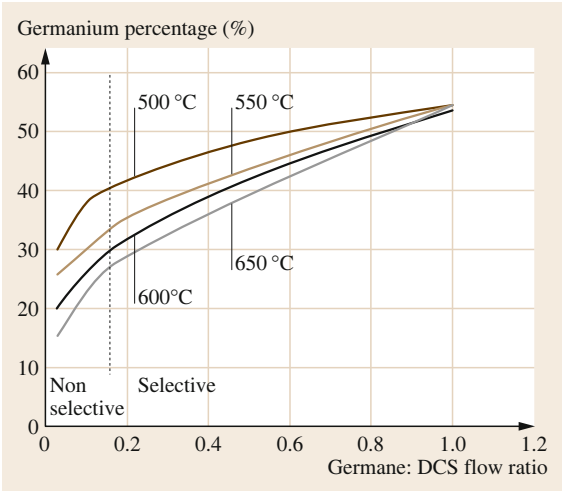


Fig. 22.25 Germanium percentage as a function of germane: dichlorosilane (DCS) flow ratio for temperatures in the range 500–650 °C showing the move from nonselective to selective growth as the proportion of germane in the gas flow increases. (After *Zhong* et al. [22.65], copyright 1990, American Institute of Physics)

process is a lower growth rate at low temperatures, as can clearly be seen in Fig. 22.24. It is also possible to grow silicon selectively using dichlorosilane and HCl [22.64].

Selective $Si_{1-x}Ge_x$ growth is generally easier to achieve than selective silicon growth, as illustrated in Fig. 22.25 [22.65] for $Si_{1-x}Ge_x$ growth using germane and dichlorosilane. The growth moves from nonselective to selective as the proportion of germane in the gas flow increases.

Arrhenius plots for $Si_{1-x}Ge_x$ growth using germane and dichlorosilane are shown in Fig. 22.26 for $Si_{1-x}Ge_x$ layers grown using germane and dichlorosilane and for two different HCl flows. It can be seen that the growth rate decreases and the activation energy increases with increasing HCl flow. The explanation proposed for this behaviour is that the limiting growth mechanism changes from hydrogen desorption from the growing surface to chlorine or HCl desorption from the surface [22.66]. This decrease in growth rate at high HCl flows is a disadvantage because it leads to increased growth times. High HCl flows can also cause surface roughening when the $Si_{1-x}Ge_x$ layer is heavily boron-doped [22.66]. These considerations demonstrate that the HCl flow should be chosen to be to the smallest value that is consistent with good selective epitaxy.

Silane can be used for selective silicon epitaxy if the growth is performed at a high temperature. This approach relies on the fact that nucleation on oxide is

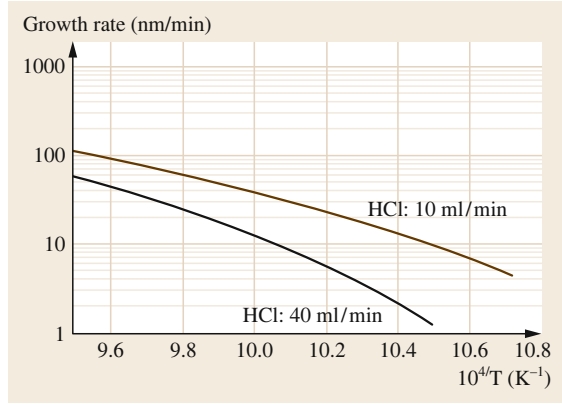


Fig. 22.26 Arrhenius plots for $Si_{1-x}Ge_x$ growth at two different HCl flow rates. The dichlorosilane and germane flow rates were fixed at 100 and 8 ml/min, respectively. (After *Kiyota* et al. [22.66], copyright 2002, IEEE)

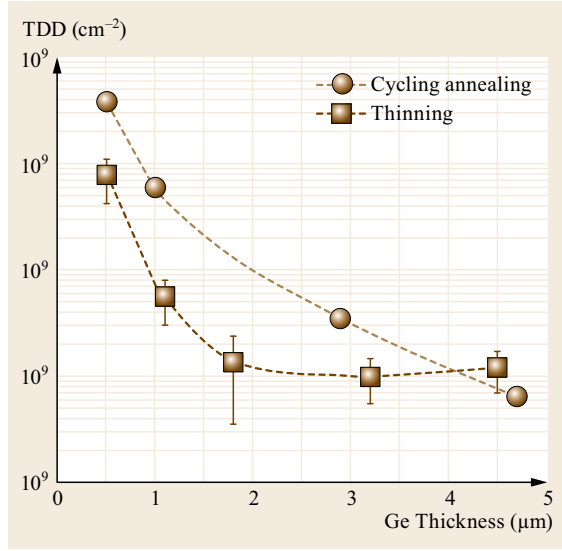


Fig. 22.27 Variation of TDD on Ge thickness for Ge grown on Si using cycling annealing at 800 °C versus Ge grown to 4.5 μm followed by thinning to the desired thickness. (Adapted from [22.67])

more difficult than that on silicon. This incubation time for growth on an oxide layer is relatively long at high temperatures but much shorter for growth at low temperatures. Selective silicon layers 1 μm thick can be grown using silane at a temperature of 960 °C [22.68], but the achievable layer thickness decreases with decreasing temperature. At 800 °C the maximum selective silicon layer thickness is around 130 nm, at 700 °C it is around 60 nm, and at 620 °C it is around 40 nm. Selective growth to silicon dioxide can be achieved using silane only, but not to silicon nitride.

22.3.6 Heteroepitaxial Ge on Si

The heteroepitaxial growth of Ge on Si has also garnered much recent interest for electronic and photonic applications. In almost all cases, a low temperature buffer layer is deposited to relieve misfit strain and allow two-dimensional (2-D) film growth. In the absence of such a buffer layer islands begin to form beyond thicknesses of several nm [22.69]. The LT buffer layer acts as a seed and allows the growth of thick Ge flat films. However, the density of threading dislocations becomes a problem for these layers. Several techniques have been explored to reduce the threading dislocation density (TDD). Cyclic annealing and thick film growth followed by etching to thin the layers have both produced good results bringing TDD to the order of $10^6\,\mathrm{cm}^{-2}$. Figure 22.27 shows the dependence on film thickness from these two techniques [22.67].

Ge layers on Si can be strain engineered for electronic and photonic applications. As-grown layers are compressively strained because of the lattice mismatch to the underlying Si substrate. However, cooling from a high growth temperature or a postanneal results in the Ge being tensile strained because of the difference in the thermal expansion coefficients of Ge and Si [22.70]. Tensile strained Ge is attractive because the strain reduces the direct bandgap of the Ge. N-type doping can then be used to fill the L conduction band valley to the level of the Γ valley, effectively making the Ge a direct bandgap material [22.67, 69, 70].

22.4 Polycrystalline Silicon-Germanium

In the past ten years there has been increasing interest in polycrystalline silicon-germanium for a number of applications that require polycrystalline material deposition at low temperature (around 600 °C). Examples of potential applications are thin film transistors, gates of MOS transistors and poly-SiGe emitters for SiGe HBTs.

In thin-film transistor technologies [22.71–73], polycrystalline silicon-germanium is compatible with the low-thermal-budget processing that is needed to produce thin-film devices for large-area electronics. The key physical property of polycrystalline silicon-germanium that makes it attractive is its lower melting point compared with silicon. This means that processes such as deposition, crystallisation, grain growth, and dopant activation will occur at a lower temperature than in silicon. Thus lower temperature processes can be used for poly-SiGe devices and hence it is preferable to poly-Si in applications with tight thermal-budget requirements.

In MOS transistors, polycrystalline silicon-germanium is attractive as a gate material for future generations of MOS transistor, since the germanium content in the silicon-germanium layer can be varied by 200–300 mV in the direction of a mid-gap gate [22.74–76]. This can be understood from Fig. 22.28, which compares the conduction- and valence-band energy levels in single-crystal silicon, silicon-germanium and germanium. Silicon and germanium have similar electron affinities (4.05 and 4.00 eV respectively), but germanium has a much smaller band gap (0.66 eV compared with 1.12 eV). The energy difference between the valence band and the vacuum level is therefore about 0.5 eV smaller in germanium than in silicon. In silicon-germanium, this energy difference can be varied by varying the germanium content. This allows the threshold voltage of p-channel MOS transistors to be tuned by varying the germanium content in the poly-$Si_{1-x}Ge_x$ gate.

Figure 22.29 shows values of work-function difference between the poly-$Si_{1-x}Ge_x$ gate and the n-type silicon substrate as a function of germanium content in a poly-$Si_{1-x}Ge_x$ gate [22.72]. The work function is defined as the difference in energy between the vacuum level and the Fermi level. In p^+ poly-$Si_{1-x}Ge_x$ the Fermi level is near the valence band and hence the work-function difference varies strongly with germanium content. In n^+ poly-$Si_{1-x}Ge_x$ the Fermi level is near the conduction band and hence the work-function difference varies little with germanium content.

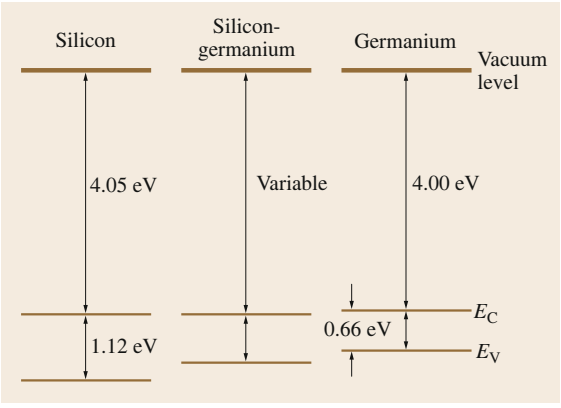


Fig. 22.28 Band-energy levels in silicon, silicon-germanium and germanium

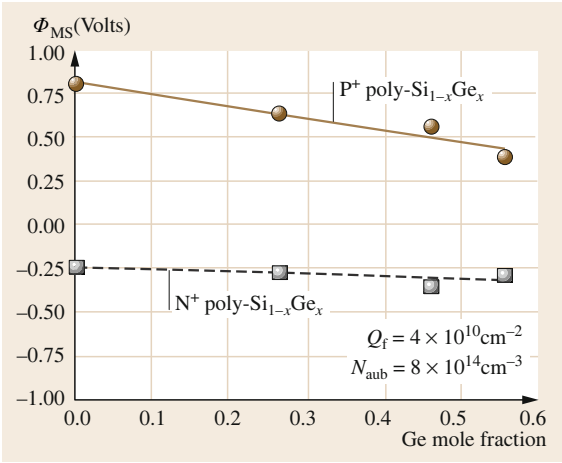


Fig. 22.29 Work-function difference between a poly-$Si_{1-x}Ge_x$ gate and an n-type substrate as a function of germanium content. (After *King* and *Saraswat* [22.72], copyright 1994, IEEE)

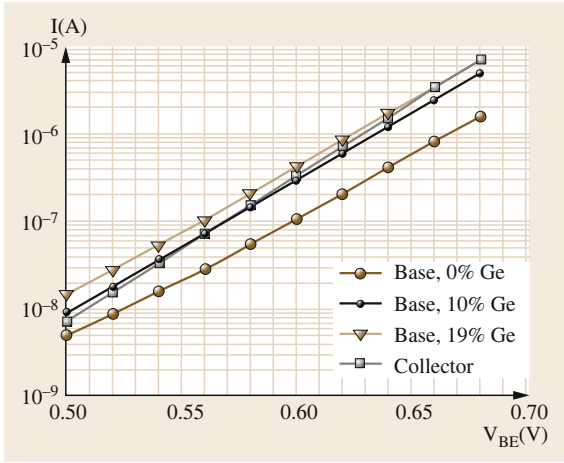


Fig. 22.30 Use of a poly-$Si_{1-x}Ge_x$ emitter to vary the base current of a $Si_{1-x}Ge_x$ HBT and hence give the best trade-off between gain and breakdown voltage BV_{CEO}. (After *Kunz* et al. [22.77], copyright 2003, IEEE)

In $Si_{1-x}Ge_x$ HBTs, poly-$Si_{1-x}Ge_x$ has potential as an emitter of a SiGe HBT [22.77]. In bipolar transistors, the breakdown voltage, BV_{CEO}, is inversely proportional to the gain [22.78] and hence transistors with a high gain have lower values of breakdown voltage. $Si_{1-x}Ge_x$ HBTs inherently have high values of the gain because the reduced band gap of the $Si_{1-x}Ge_x$ base enhances the collector current. The base current is unchanged by the $Si_{1-x}Ge_x$ base and hence the gain, which is the ratio of collector current to base current, is increased. The use of a poly-$Si_{1-x}Ge_x$ emitter instead of a poly-Si emitter provides a reduced band gap in the emitter, which enhances the base current, and thereby reduces the gain. Typical measured values of base current in a poly-$Si_{1-x}Ge_x$ emitter are shown in Fig. 22.30, where it can be seen that 19% germanium gives a factor of approximately four reduction in gain. A poly-$Si_{1-x}Ge_x$ emitter therefore allows the gain to be tuned to give the best trade-off between gain and breakdown voltage BV_{CEO}.

22.4.1 Electrical Properties of Polycrystalline $Si_{1-x}Ge_x$

Figure 22.31 shows the sheet resistance as a function of anneal temperature for boron- and phosphorus-doped $Si_{1-x}Ge_x$ for different germanium contents. For boron-doped poly-$Si_{1-x}Ge_x$ the sheet resistance decreases with increasing germanium content, with the decrease being large between 0 and 25% germanium and smaller between 25 and 50% germanium. In contrast, for phosphorus-doped poly-$Si_{1-x}Ge_x$ the sheet resistance increases with increasing germanium content, with the increase being small between 0 and 25% germanium and large between 25 and 50%. Similar behaviour is seen for arsenic-doped poly-$Si_{1-x}Ge_x$ where higher values of sheet resistance have been reported for poly-$Si_{1-x}Ge_x$ than for poly-Si [22.75].

The explanation for the sheet-resistance results in Fig. 22.31 can be found in Fig. 22.32, which shows the results of Hall measurements [22.72]. For boron-doped poly-$Si_{1-x}Ge_x$ both the activation and the Hall mobility increase with increasing germanium content, thereby explaining the decrease in sheet resistance with increasing germanium content. For phosphorus-doped poly-$Si_{1-x}Ge_x$ there is little change in activation and electron mobility at low germanium contents, but a sharp decrease in activation at germanium contents above 35%. This explains the sharp increase in sheet resistance seen in Fig. 22.31 for germanium contents between 25 and 50%. The decrease in activation at high germanium contents in the phosphorus-doped poly-$Si_{1-x}Ge_x$ may be due to increased segregation at grain boundaries. Boron does not generally segregate to grain boundaries [22.80], which may explain the different behaviour in boron- and phosphorus-doped poly-$Si_{1-x}Ge_x$.

Asafa et al. experimented with Ge-rich polycrystalline films grown by CVD and demonstrated that the electrical properties are very sensitive to the Ge fraction [22.81]. An increase in Ge fraction from 0.83 to 0.88 resulted in a sixfold increase in Hall mobility and decrease in resistivity. This was accompanied by a doubling of grain size resutling in more columnar grains.

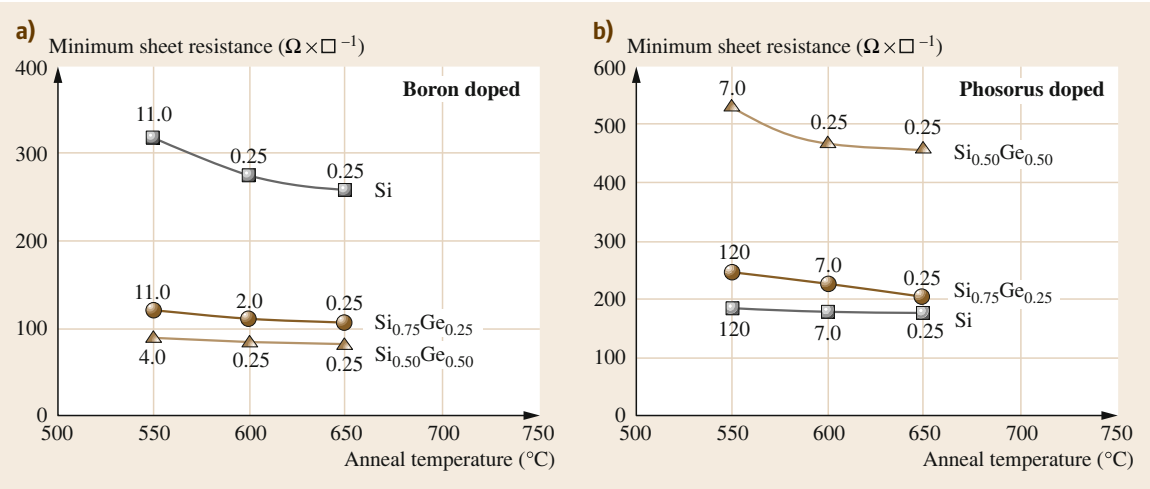


Fig. 22.31a,b Sheet resistance as a function of anneal temperature for **(a)** boron- and **(b)** phosphorus-doped polycrystalline $Si_{1-x}Ge_x$ with various germanium contents. (After *Bang* et al. [22.79], copyright 1995, American Institute of Physics)

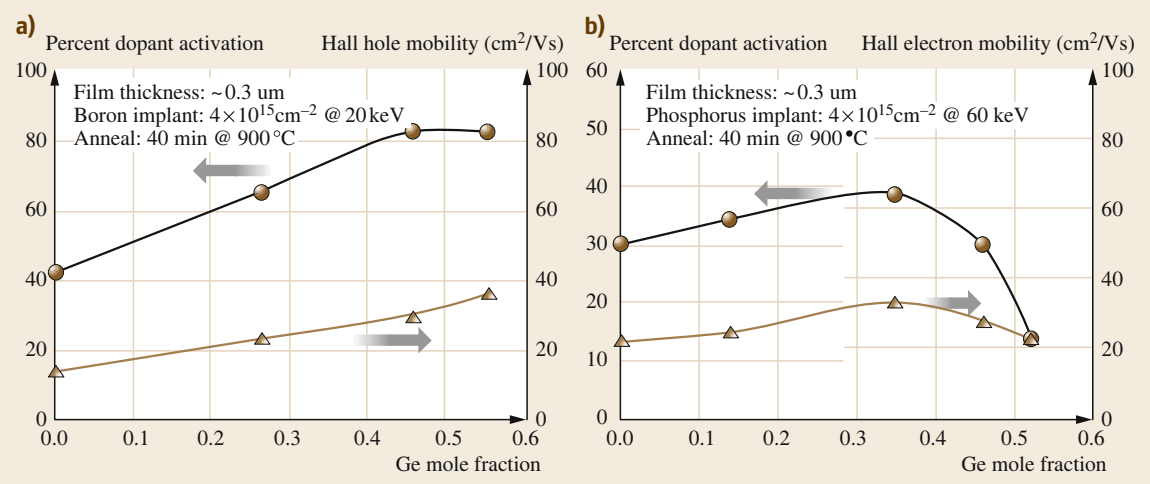


Fig. 22.32a,b Percentage dopant activation and Hall hole mobility as a function of germanium content for **(a)** boron- and **(b)** phosphorus-doped poly-$Si_{1-x}Ge_x$. (After *King* and *Saraswat* [22.72], copyright 1994, IEEE)

References

22.1 R. Braunstein, A.R. Moore, F. Herman: Phys. Rev. **109**, 695 (1958)

22.2 S.S. Iyer, G.L. Patton, J.M.C. Stork, B.S. Meyerson, D.L. Harame: IEEE Trans. Electron Devices **36**, 2043 (1989)

22.3 C.A. King, J.L. Hoyt, J.F. Gibbons: IEEE Trans. Electron Devices **36**, 2093 (1989)

22.4 H. Miyata, T. Yamada, D.K. Ferry: Appl. Phys. Lett. **62**, 2661 (1993)

22.5 T. Vogelsang, K.R. Hofmann: Appl. Phys. Lett. **63**, 186 (1993)

22.6 J. Welser, J.L. Hoyt, J.F. Gibbons: IEEE Electron Device Lett. **15**, 100 (1994)

22.7 A. Sadak, K. Ismile, M.A. Armstrong, D.A. Antoniadis, F. Stern: IEEE Trans. Electron Devices **43**, 1224 (1996)

22.8 B. Jagannathan, M. Khater, F. Pagette, J.-S. Rieh, D. Angell, H. Chen, J. Florkey, F. Golan, D.R. Greenberg, R. Groves, S.J. Jeng, J. Johnson, E. Mengistu, K.T. Schonenberger, C.M. Schnabel, P. Smith, A. Stricker, D. Ahlgren, G. Freeman, K. Stein, S. Subbanna: IEEE Electron Device Lett. **23**, 258 (2002)

22.9 Z.A. Shafi, P. Ashburn, G.J. Parker: IEEE J. Solid State Circuits **25**, 1268 (1990)
22.10 J.M. Matthews, A.E. Blakeslee: J. Cryst. Growth **27**, 118 (1974)
22.11 J.M. Matthews, A.E. Blakeslee: J. Cryst. Growth **32**, 265 (1975)
22.12 R. People, J.C. Bean: Appl. Phys. Lett. **47**, 322 (1985)
22.13 S.C. Jain, T.J. Gosling, J.R. Willis, R. Bullough, P. Balk: Solid State Electron. **35**, 1073 (1992)
22.14 S. Margalit, A. Bar-lev, A.B. Kuper, H. Aharoni, A. Neugroschel: J. Cryst. Growth **17**, 288 (1972)
22.15 O.W. Holland, C.W. White, D. Fathy: Appl. Phys. Lett. **51**, 520 (1987)
22.16 D.C. Paine, C. Caragianis, A.F. Schwartzman: J. Appl. Phys. **70**, 5076 (1991)
22.17 W.S. Liu, E.W. Lee, M.A. Nicolet, V. Arbet-Engels, K.L. Wang, N.M. Abuhadba, C.R. Aita: J. Appl. Phys. **71**, 4015 (1992)
22.18 J.P. Zhang, P.L.F. Hemment, S.M. Newstead, A.R. Powell, T.E. Whall, E.H.C. Parker: Thin Solid Films **222**, 141 (1992)
22.19 M.A. Rabie, Y.M. Haddara, J. Carette: J. Appl. Phys. **98**, 074904 (2005)
22.20 D. Nayak, K. Kamjoo, J.C.S. Woo, J.S. Park, K.L. Wang: Appl. Phys. Lett. **56**, 66 (1990)
22.21 M. Mukhopadhyay, S.K. Ray, C.K. Maiti, D.K. Nayak, Y. Shiraki: Appl. Phys. Lett. **65**, 895 (1994)
22.22 T. David, A. Benkouider, J.-N. Aqua, M. Cabie, L. Favre, T. Neisius, M. Abbarchi, M. Naffouti, A. Ronda, K. Liu, I. Berbezier: J. Phys. Chem. C **119**, 24606 (2015)
22.23 D.V. Lang, R. People, J.C. Bean, A.M. Sergent: Appl. Phys. Lett. **47**, 1333 (1985)
22.24 J. Eberhardt, E. Kasper: Mater. Sci. Eng. B **89**, 93 (2002)
22.25 C.D. Thurmond: J. Electrochem. Soc. **122**, 1133 (1975)
22.26 J. Poortmans, S.C. Jain, D.H.J. Totterdell, M. Caymax, J.F. Nijs, R.P. Mertens, R. Van Overstraeten: Solid State Electron. **36**, 1763 (1993)
22.27 T. Manku, A. Nathan: J. Appl. Phys. **69**, 8414 (1991)
22.28 T. Manku, A. Nathan: Phys. Rev. B **43**, 12634 (1991)
22.29 J. Poortmans: Low Temperature Epitaxial Growth of Silicon and Strained $Si_{1-x}Ge_x$ Layers and Their Application in Bipolar Transistors, PhD Thesis (University of Leuven, Leuven 1993)
22.30 J.M. McGregor, T. Manku, J.-P. Noël, D.J. Roulston, A. Nathan, D.C. Houghton: J. Electron. Mater. **22**(3), 319 (1993)
22.31 M.V. Fischetti, S.E. Laux: J. Appl. Phys. **80**, 2234 (1996)
22.32 C.W. Leitz, M.T. Currie, M.L. Lee, Z.-Y. Cheng, D.A. Antoniadis, E.A. Fitzgerald: J. Appl. Phys. **92**, 3745 (2002)
22.33 D.J. Robbins, L.T. Canham, S.J. Barnett, A.D. Pitt, P. Calcott: J. Appl. Phys. **71**, 1407 (1992)
22.34 N.L. Rowell, J.-P. Noel, D.C. Houghton, A. Wang, D.D. Perovic: J. Vac. Sci. Technol. B **11**, 1101 (1993)
22.35 D.A. Grutzmacher, T.O. Sedgwick, G.A. Northrop, A. Zaslavsky, A.R. Powell, V.P. Kesan: J. Vac. Sci. Technol. B **11**, 1083 (1993)
22.36 J. Brunner, J. Nutzel, M. Gail, U. Menczigar, G. Abstreiter: J. Vac. Sci. Technol. B **11**, 1097 (1993)
22.37 K. Terashima, M. Tajima, T. Tatsumi: J. Vac. Sci. Technol. B **11**, 1089 (1993)
22.38 H. Prestling, T. Zinke, A. Splett, H. Kibbel, M. Jaros: Appl. Phys. Lett. **69**, 2376 (1996)
22.39 L. Masarotto, J.M. Hartmann, G. Bremond, G. Rolland, A.M. Papon, M.N. Semeria: J. Cryst. Growth **255**, 8 (2003)
22.40 J.S. Park, T.L. Lin, E.W. Jones, H.M. Del Castillo, S.D. Gunapall: Appl. Phys. Lett. **64**, 2370 (1994)
22.41 S.S. Murtaza, J.C. Cambell, J.C. Bean, L.J. Peticolas: IEEE Photon. Tech. Lett. **8**, 927 (1996)
22.42 D.J. Robbins, M.B. Stanaway, W.Y. Leong, R.T. Carline, N.T. Gordon: Appl. Phys. Lett. **66**, 1512 (1995)
22.43 A. Chin, T.Y. Chang: Lightwave Technol. **9**, 321 (1991)
22.44 R. People, J.C. Bean, C.G. Bethia, S.K. Sputz, L.J. Peticolas: Appl. Phys. Lett. **61**, 1122 (1992)
22.45 P. Kruck, M. Helm, T. Fromherz, G. Bauer, J.F. Nutzel, G. Abstreiter: Appl. Phys. Lett. **69**, 3372 (1996)
22.46 R.A. Soref, L. Friedman, G. Sun: Superlattices Microstruct. **23**, 427 (1998)
22.47 G. Sun, L. Friedman, R.A. Soref: Superlattices Microstruct. **22**, 3 (1998)
22.48 J. Halpin, S. Rhead, A.M. Sanchez, M. Myronov, D.R. Leadley: Semicond. Sci. Technol. **30**, 114009 (2015)
22.49 J.H. Bahng, K.J. Kim, H. Ihm, J.Y. Kim, H.L. Park: J. Phys. Condens. Matter **13**, 777 (2001)
22.50 J. Weber, M.I. Alonso: Phys. Rev. B **40**, 5684 (1989)
22.51 H. Landolt, R. Bornstein: *Numerical Data and Functional Relationships in Science and Technology*, Vol. III/17a, ed. by O. Madelung (Springer, Berlin, Heidelberg, New York 1982)
22.52 G.S. Mitchard, T.C. McGill: Phys. Rev. B **25**, 5351 (1982)
22.53 M. Meuris, S. Verhaverbeke, P.W. Mertens, M.M. Heyns, L. Hellemans, Y. Bruynseraede, A. Philipessian: Jpn. J. Appl. Phys. **31**, L1514 (1992)
22.54 A. Ishizaki, Y. Shiraki: J. Electrochem. Soc. **129**, 666 (1986)
22.55 B.S. Meyerson, F.J. Himpsel, K.J. Uram: Appl. Phys. Lett. **57**, 1034 (1990)
22.56 G.S. Higashi, Y.T. Chabal, G.W. Trucks, K. Raghavachari: Appl. Phys. Lett. **56**, 656 (1990)
22.57 F.W. Smith, G. Ghidini: J. Electrochem. Soc. **129**, 1300 (1982)
22.58 G. Ghidini, F.W. Smith: J. Electrochem. Soc. **131**, 2924 (1984)
22.59 J.L. Regolini, D. Bensahel, E. Scheid, J. Mercier: Appl. Phys. Lett. **54**, 658 (1989)
22.60 M. Racanelli, D.W. Greve: Appl. Phys. Lett. **56**, 2524 (1990)
22.61 G.R. Srinivasan, B.S. Meyerson: J. Electrochem. Soc. **134**, 1518 (1987)
22.62 M. Racanelli, D.W. Greve, M.K. Hatalis, L.J. van Yzendoorn: J. Electrochem. Soc. **138**, 3783 (1991)
22.63 D.J. Robbins, J.L. Glasper, A.G. Cullis, W.Y. Leong: J. Appl. Phys. **69**, 3729 (1991)
22.64 A. Ishitani, H. Kitajima, N. Endo, N. Kasai: Jpn. J. Appl. Phys. **28**, 841 (1989)
22.65 Y. Zhong, M.C. Ozturk, D.T. Grider, J.J. Wortman, M.A. Littlejohn: Appl. Phys. Lett. **57**, 2092 (1990)

22.66 Y. Kiyota, T. Udo, T. Hashimoto, A. Kodama, H. Shimamoto, R. Hayami, E. Ohue, K. Washio: IEEE Trans. Electron Devices **49**, 739 (2002)
22.67 Y. Yamamoto, S. Lischke, L. Zimmerman, D. Knoll, J. Murota, B. Tillack: ECS Trans. **67**, 123 (2015)
22.68 J.M. Bonar: Process Development and Characterisation of Silicon and Silicon–Germanium Grown in a Novel Single-Wafer LPCVD System, PhD Thesis (University of Southampton, Southampton 1996)
22.69 H. Ye, J. Yu: Sci. Technol. Adv. Mater. **15**, 024601 (2014)
22.70 C. Zhao, S. Wen, Q. Hou, W. Qiu, Y. Xing, S. Su, B. Cheng: J. Phys. Chem. Solids **90**, 87 (2016)
22.71 T.-J. King, K. C. Saraswat: IEDM Tech. Dig. (IEEE, 1991) p. 567
22.72 T.-J. King, K.C. Saraswat: IEEE Trans. Electron Devices **41**, 1581 (1994)
22.73 J.A. Tsai, A.J. Tang, T. Noguchi, R. Reif: J. Electrochem. Soc. **142**, 3220 (1995)
22.74 T.-J. King, J.R. Pfiester, K.C. Saraswat: IEEE Electron Device Lett. **12**, 533 (1991)
22.75 C. Salm, D.T. van Veen, D.J. Gravesteijn, J. Holleman, P.H. Woerlee: J. Electrochem. Soc. **144**, 3665 (1997)
22.76 Y.V. Ponomarev, P.A. Stolk, C.J.J. Dachs, A.H. Montree: IEEE Trans. Electron Devices **47**, 1507 (2000)
22.77 V.D. Kunz, C.H. de Groot, S. Hall, P. Ashburn: IEEE Trans. Electron Devices **50**, 1480 (2003)
22.78 P. Ashburn: *Silicon-Germanium Heterojunction Bipolar Transistors* (Wiley, Chichester 2003)
22.79 D.S. Bang, M. Cao, A. Wang, K.C. Saraswat, T.-J. King: Appl. Phys. Lett. **66**, 195 (1995)
22.80 I.R.C. Post, P. Ashburn: IEEE Trans. Electron Devices **38**, 2442 (1991)
22.81 T.B. Asafa, A. Witvrouw, B.S. Morcos, K. Vanstreels, S.A.M. Said: Appl. Phys. A **116**, 751 (2014)

23. Temperature-Insensitive Band-Gap III-V Semiconductors: Tl-III-V and III-V-Bi

Hajime Asahi

Thallium-containing III-V (Tl-III-V) and bismuth-containing III-V (III-V-Bi) alloy semiconductors were first proposed as novel functional semiconductors. They are alloys consisting of semiconductors (III-V) and semimetals (Tl-V, III-Bi), and are important materials for the fabrication of temperature-insensitive lasing wavelength laser diodes as well as long wavelength infrared (LWIR) optical devices. However, the growth conditions for alloys containing Tl and Bi are very strict and the growth windows are narrow. In this chapter, the expected properties of these semiconductors and the experimental results for the growth, characterization, and device applications are described.

23.1 **Tl-III-V Alloy Semiconductors** 544
23.1.1 Expected Properties of Tl-Based III-V Alloys 544
23.1.2 Growth of Tl-III-V Alloys 546
23.1.3 Temperature Dependence of Physical Properties 547
23.1.4 Tl-III-V LD Application 549
23.2 **III-V-Bi Alloy Semiconductors** 550
23.2.1 Expected Properties of Bi-Based III-V Alloys 550
23.2.2 Growth of III-V-Bi Alloys 551
23.2.3 Optical Properties of III-V-Bi Alloys 552
23.2.4 III-V-Bi LD Application 553
23.3 **Summary** 554
References 554

Wavelength division multiplexing (WDM) technology is very important for an optical fiber communication system for increasing transport capacity and obtaining flexible network management. However, one of the problems encountered when using InGaAsP/InP laser diodes (LDs) in the WDM system is that the lasing wavelength fluctuates with ambient temperature variation mainly due to the temperature dependence of the band-gap energy. Therefore, LDs in WDM systems must be equipped with Peltier elements to stabilize the LD temperature.

To solve this problem, *Oe* et al. proposed the use of temperature-insensitive band-gap semiconductors as an active layer of LDs [23.1]. Such a temperature-insensitive band-gap property was observed in the II-VI semiconductor, HgCdTe, as shown in Fig. 23.1 [23.2]. HgCdTe is usually used for the long wavelength infrared (LWIR) detectors. CdTe is a semiconductor and HgTe is a semimetal. With the increase of HgTe composition, their band-gap energy decreases. At the CdTe-rich side of HgCdTe, the band-gap energy decreases with the increase of temperature, while at the HgTe-rich side of HgCdTe, the band-gap energy increases with the increase of temperature. Therefore, at a certain HgTe composition (0.48) region, the temperature-insensitive behavior is expected [23.3]. Similar behavior was also observed for HgCdSe, where the very small temperature dependence of refractive index was also confirmed [23.4].

The origin of this temperature-insensitive band-gap energy is considered as follows. Figure 23.2 shows the alloy composition dependence of the band structure for HgCdTe alloys [23.5]. At the CdTe side of HgCdTe, the Γ_6 band behaves as a conduction band and the Γ_8 band behaves as a valence band. Also in this composition region, the band-gap energy decreases with increasing temperature similar to usual semiconductors. With increasing HgTe composition, the Γ_6 band goes down and the band-gap energy decreases. By further increasing HgTe composition, the Γ_6 and Γ_8 bands cross and the band inversion occurs. After this band inversion, the Γ_6 band becomes a valence band and the Γ_8 light hole band forms a conduction band. The Γ_8 heavy hole band still behaves as a valence band and HgCdTe becomes a semimetal. Therefore, the energy difference between the two bands increases with increasing temperature so that the band-gap energy for HgCdTe semiconductors at the HgTe side composition increases with increasing temperature. At some intermediate alloy composition, we can observe the

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_23

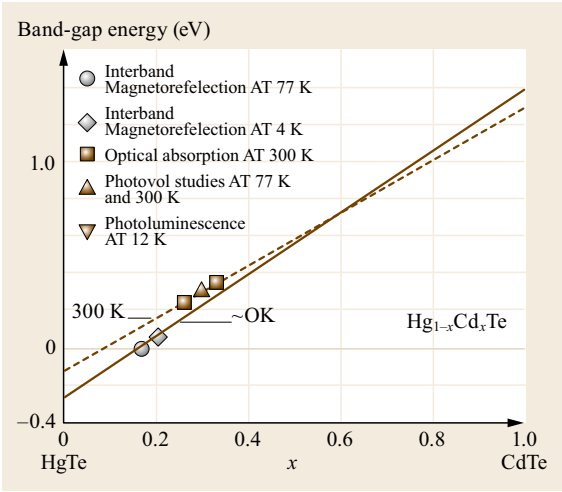


Fig. 23.1 Band-gap energy versus alloy composition for HgCdTe as a function of temperature. (After [23.2])

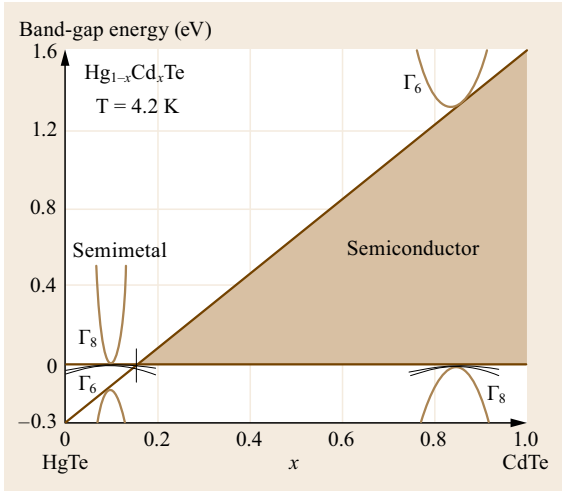


Fig. 23.2 Alloy composition dependence of the band structure for HgCdTe alloys. (After [23.5])

temperature-insensitive band-gap energy as shown in Fig. 23.1 [23.2].

From the above-mentioned discussion, the alloy semiconductor consisting of a semiconductor and a semimetal is expected to show the temperature-insensitive band-gap behavior. *Oe* et al. suggested GaAsBi and InGaAsBi alloys as new semiconductors showing temperature-insensitive band-gap energy for possible application to the temperature-stable lasing wavelength LDs [23.1]. *Asahi* et al. also proposed TlInGaAs and TlInGaP as new semiconductors showing temperature-insensitive band-gap energy [23.6, 7]. Similar to HgCdTe, it is also expected that these alloy semiconductors can be used for the LWIR detectors.

This chapter describes the expected properties of the Tl-III-V and III-V-Bi alloy semiconductors and the present status of their growth, characterization, and device applications.

23.1 Tl-III-V Alloy Semiconductors

Alloy semiconductor HgCdTe is widely used for LWIR focal-plane arrays (FPA). However, HgCdTe is a weakly bonded II-VI compound with material growth and processing problems dominated by native defects that limit FRA performance and yield [23.8]. *Van Schilfgaarde* et al. proposed thallium (TI)-based III-V ternary alloy semiconductors, TlInSb, TlInAs, and TlInP as the possible candidates for future LWIR detector materials [23.8–11]. Their theoretical calculations of TlSb, TlAs, and TlP showed that they have negative band gaps similar to HgTe (semimetal). Therefore, it is expected that TlInSb, TlInAs, and TlInP should exhibit LWIR band gaps. Furthermore, they are III-V materials and are compatible with the conventional III-V device fabrication technologies. Therefore, it was expected that these Tl-based alloys could overcome the difficulties encountered in the widely used LWIR semiconductor HgCdTe.

Interesting extensions of TlInSb, TlInAs, and TlInP alloys are quaternary alloys TlInGaP and TlInGaAs [23.6, 7, 12]. These quaternary alloys can be lattice matched to GaAs, InP, and InAs substrates. Lattice matching is very important to obtain good quality crystals. Furthermore, they are expected to have temperature-insensitive band-gap energies at some alloy compositions and they can be applied to the fabrication of temperature-insensitive lasing wavelength LDs [23.6]. Further extension is a quinary alloy semiconductor TlInGaAsN, which consists of a semiconductor GaInNAs and a semimetal TlAs and is also expected to be applicable to fabricate the temperature-stable lasing wavelength and temperature-stable threshold LDs [23.13] as discussed below.

23.1.1 Expected Properties of Tl-Based III-V Alloys

Figure 23.3 shows the expected relationships between the band-gap energy (E_g) and the lattice constant for

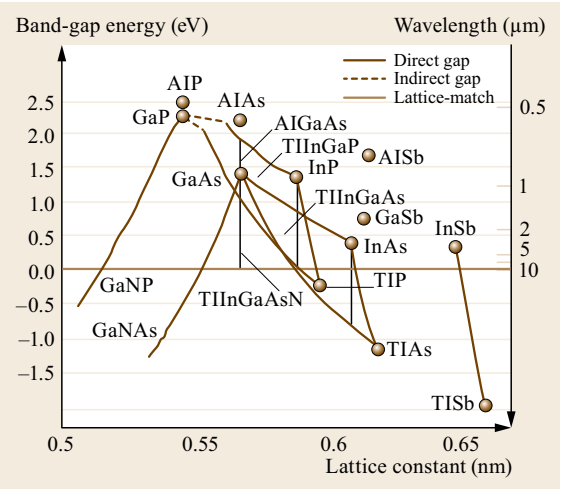


Fig. 23.3 Relationship for the band-gap energy versus alloy composition for Tl-III-V alloys as well as usual III-V semiconductors. (After [23.6, 12, 13])

Tl-based III-V alloys [23.6, 12, 13] based on the theoretical results for TlInSb, TlInAs, and TlInP by *van Schilfgaarde* et al. [23.8, 11]. The lines for TlGaAs and TlGaP were predicted by *Asahi* et al. [23.6]. The data for other III-V compound materials are also plotted in the figure for reference.

Van Schilfgaarde et al. [23.8–11] have theoretically calculated the lattice constant, cohesive energy, elastic constant, band-gap energy, effective mass, and electron mobility by the local density approximation first principle theory. They obtained the following results: (1) TlAs and TlP are formed in the zincblende structure, while TlSb is formed in the CsCl structure; (2) the lattice constants of TlSb (0.659 nm), TlAs (0.618 nm), and TlP (0.596 nm) are only 1.5−2% larger than those of InSb (0.648 nm), InAs (0.606 nm), and InP (0.587 nm), respectively; (3) the cohesive energies are about twice that of HgTe; therefore, bond strengths are higher; (4) TlSb, TlAs, and TlP are semimetals with negative band gaps of −1.60, −1.34, and −0.27 eV, respectively, similar to HgTe (−0.3 eV); (5) TlInSb, TlInAs, and TlInP exhibit a direct band gap of 0.1 eV (12.4 μm) at the Tl compositions of 0.09, 0.15, and 0.67, respectively. These properties are promising for application to LWIR detectors.

Thermodynamical properties of TlInP and TlInAs were theoretically estimated [23.14]. It was predicted that only low concentrations of Tl can be achieved in TlInP (< 5% at 350 °C) using gas-source molecular beam epitaxy (GSMBE). Although much less than 1% Tl was expected to be soluble in TlInAs, the error estimates in calculation indicated that obtaining 15% Tl is possible.

Quaternary alloys of TlInGaP and TlInGaAs proposed by *Asahi* et al. [23.6, 7] can be easily lattice matched to InP, GaAs, and InAs substrates as shown in Fig. 23.3. TlInGaP lattice matched to InP can cover the band-gap energy from 1.24 eV (0.92 μm at 0 K, InP) to less than 0.1 eV (12.4 μm, TlGaP), and over the whole lattice-matched composition the direct band-gap is expected to form. The addition of Al and As can further extend the energy (wavelength) range. TlInGaAs lattice matched to InP can cover from 0.75 eV (1.65 μm, InGaAs) to about 0 eV (TlGaAs). These quaternary alloys can be applied to LWIR detectors and emitters. They are expected to exhibit the temperature-insensitive band-gap energy behavior at some alloy compositions similar to HgCdTe [23.6]. This is because these semiconductors are alloys consisting of semiconductors (InGaP, InGaAs) and semimetals (TlP, TlAs). These temperature-insensitive band-gap energy characteristics are very promising for the fabrication of LDs whose wavelength does not change with or is insensitive to ambient temperature variation. The band lineup of lattice-matched TlInGaP/InP heterostructures was suggested to be type-I and the conduction band discontinuities are larger than the valence band offsets [23.6], also proved by the theoretical calculation [23.15].

Quinary alloy semiconductor TlInGaAsN was proposed by *Asahi* et al. to fabricate the temperature-stable lasing wavelength and temperature-stable threshold current TlInGaAsN/AlGaAs LDs [23.13]. TlInGaAsN is an alloy of a semiconductor InGaNAs and a semimetal TlAs as shown in Fig. 23.3. InGaNAs/AlGaAs heterostructures are widely studied to fabricate temperature-stable threshold current LDs in the wavelength range of 1 μm because of their large conduction band discontinuity [23.16]. However, there exists a long wavelength limit of 1.3 μm in this material system. This is because of the degradation of the crystal quality as well as the reduction of valence band offset with increasing N composition [23.17]. To overcome this problem, new alloy semiconductor TlInGaAsN/AlGaAs heterostructures were proposed [23.13].

As can be clearly seen in Fig. 23.3, much lower band-gap energy, therefore, longer wavelength, can be easily obtained for this new semiconductor compared with GaInNAs. TlInGaAsN/AlGaAs heterostructures also have a large conduction band discontinuity similar to GaInNAs/AlGaAs ones. Furthermore, TlInGaAsN is expected to have temperature-insensitive band-gap energy because it is an alloy of a semiconductor and a semimetal, similar to TlInGaAs. Therefore, the fabrication of temperature-stable lasing wavelength and temperature-stable threshold current LDs is expected by using TlInGaAsN/AlGaAs heterostructures.

23.1.2 Growth of Tl-III-V Alloys

The growth of Tl-based III-V alloys has been attempted by metal-organic vapor phase epitaxy (MOVPE) [23.18–20], gas source MBE (GSMBE) [23.6, 21–26], and solid source MBE (SSMBE) [23.27, 28].

In the MOVPE growth of TlInSb and TlInP, cyclopentadiethylthallium (CPTl) and trimethylindium (TMIn) were used as Tl and In precursors, and trimethylantimony (TMSb) or tertiarybutyldemethylantimony (TBDMSb) and phosphine (PH_3) as Sb and P precursors, respectively [23.18–20, 29]. The vapor pressure of CPTl is as low as about 3 Torr even at 120 °C. Therefore, high bubbler temperature and heated valves, and stainless delivery lines are needed.

In GSMBE or SSMBE growth, elemental Tl, In, and Ga were used as group III sources [23.6, 21–28]. Thermally cracked arsine (AsH_3), phosphine (PH_3), or elemental As and Sb were used as group V sources. Tl vaporizes from the melt state, which means that the controllability of the Tl flux is much better than that of Sb. The problem in using Tl metal is its ease of oxidation. The toxicity of Tl is known to be higher than that of As, so extreme care must be taken during treatment [23.6]. Tl-condensed surfaces should not be touched with bare hands and Tl flakes or vapors should not be inhaled: legal intake is documented to be at 600 mg. From this toxicity point of view, growth by MBE is much safer than that by MOPVE. In MBE, the solid Tl is used and is stored in the growth chamber, whereas in MOVPE, the Tl is supplied and vented as a form of metal organic (MO) gas. It is considered that there is no essential difference between GSMBE and SSMBE both in safety and growth mechanism, though growth conditions may be slightly different.

Growth of Tl-III-Sb Ternary Alloys

The growth of TlInSb was first conducted by *Choi* et al. [23.18]. They grew TlInSb layers with good surface morphology on GaAs and InSb substrates at 455 °C by low-pressure MOVPE. X-ray diffraction measurements showed resolved peaks of InSb and TlInSb. However, in contrast to the theoretical prediction, an increase in Tl incorporation led to a decrease in the lattice constant (higher diffraction angle). This was also observed by *Karam* et al. [23.20]. The latter obtained TlInSb with Tl composition of 0.1 (10%). Incorporation of Tl was also confirmed with Auger electron spectroscopy and Rutherford backscattering (RBS) measurements. Infrared transmission measurements showed a shift of the absorption edge toward longer wavelengths with increasing Tl composition for TlInSb [23.18, 20, 29].

Infrared photoconductors were fabricated from TlInSb/InSb grown on (100) semi-insulating GaAs substrates [23.20, 30, 31]. A clear shift of response to longer wavelength was observed. The cutoff wavelength extended to 11 μm at 300 K for the TlInSb (Tl composition of 6%) photoconductor [23.30]. The maximum detectivity was 7.6×10^8 cm $Hz^{1/2}$ W^{-1} at 77 K.

Growth of Tl-III-As Ternary Alloys

The growth experiments of TlInAs were conducted by GSMBE [23.25, 26, 32] and SSMBE [23.27, 28]. *Takenaka* et al. [23.25] confirmed the successful growth of TlInAs on InAs substrate at 430 °C by double crystal x-ray diffraction measurement (40 arcsec shift) as well as by reflection high-energy electron diffraction (RHEED) intensity oscillation measurement, where the oscillation period was reduced by the addition of Tl flux during the growth of InAs (1.3% reduction). *Lubyshev* et al. [23.28] also reported the growth of TlInAs with Tl composition of 0.6%. However, *Antonell* et al. [23.26, 32] and *Lange* et al. [23.27] reported that the incorporation of Tl into InAs was difficult. The growth of high Tl composition TlInAs has not been accomplished. The window for the appropriate growth conditions, especially substrate temperature, was narrow.

The growth of TlGaAs was conducted on GaAs substrates by SSMBE by *Lubyshev* et al. [23.28]. They reported the successful growth of TlGaAs with a Tl composition of 0.05 (diffraction peak separation of 1260 arcsec between TlGaAs and GaAs in the (400) double crystal x-ray diffraction (DCXRD)) and a second metal phase (Tl+Ga+As) on the surface. They observed that the growth of TlGaAs is easier than that of TlInAs.

Growth of Tl-III-P Ternary Alloys

TlInP growth was conducted on InP substrates by both MOVPE and GSMBE. *Razeghi* et al. [23.29] reported the successful growth of $(TlP_3)_x(InP)_{1-x}$ layers at 520 °C by MOVPE and observed clear photoresponse with onset wavelengths of 5.5 and 8.0 μm in infrared photoconductivity measurements for two samples (estimated compositions: $x = 0.6$ and 0.64), respectively. *Yamamoto* et al. confirmed the GSMBE growth of TlInP by the DCXRD measurement [23.21] and by the RHEED intensity oscillation measurement [23.22]. The growth temperature was about 450 °C and (2×4) RHEED patterns and specular surface morphology with SEM were observed [23.21]. With increasing Tl/(Tl+In) flux ratio, the XRD peak was shifted to lower diffraction angles, indicating the increase in Tl incorporation. On the other hand, *Antonell* et al. [23.26]

reported the difficulty of the growth of TlInP similar to TlInAs. Narrow growth window for the growth of these materials is probably the reason for this discrepancy.

TlGaP layers closely lattice matched to GaAs were successfully grown on GaAs substrates at about 450 °C by GSMBE [23.23]. However, TlGaP on GaAs showed the phase separation into three stable compositions: TlGaP nearly lattice matched to GaAs, GaP-like TlGaP and TlP-like TlGaP. However, by adjusting Tl and Ga fluxes, TlGaP layers without phase separation and nearly lattice matched to GaAs were grown. Photoconductance measurement on TlGaP showed the absorption in the 1.3 μm wavelength region, agreeing with the expected band-gap energies for the TlGaP lattice matched to GaAs. Furthermore, a small temperature variation of band-gap energy was observed as expected.

Growth of Tl-III-V Quaternary Alloys

Quaternary TlInGaP layers were successfully grown on InP substrates by GSMBE [23.22, 33, 34]. During growth, the RHEED pattern showed (2×4) reconstructions [23.22]. The DCXRD measurement showed a lower angle shift for the TlInGaP peak with increasing Tl flux [23.33, 34], agreeing with the increase of Tl composition. No phase separation was observed. The Tl composition obtained was less than 0.1. The increase of photoconductivity (PC) was observed in the longer wavelength region compared with the PC of InP [23.24], clearly indicating the formation of narrow band-gap semiconductors.

Quaternary TlInGaA layers were also successfully grown on InP substrates by GSMBE [23.25, 33, 34]. The growth temperature was 450 °C. The growth condition for alloys containing Tl was very strict. For example, the growth below 400 °C resulted in segregation of Tl, and the growth above 460 °C failed to incorporate Tl due to desorption of Tl atoms from the growing surface. During the growth at 450 °C, the RHEED pattern showed (2×2) reconstructions. The grown samples showed mirror surface and incorporation of Tl. With increasing Tl flux, the x-ray diffraction peak shifted toward lower angles and the Tl composition was increased (Fig. 23.4) [23.33]. It is considered that during the growth of TlInGaAs, the large part of incident Tl atoms is re-evaporated from the surface and only a small part of them is incorporated into films, very similar to the case of MBE growth of GaN under usual growth conditions [23.35].

Growth of TlInGaAsN Quinary Alloys

The growth of TlInGaAsN was conducted by GSMBE with thermally cracked AsH_3 and plasma-enhanced N_2 as As and N sources [23.36]. TlInGaAsN/GaAs quantum well (QW) structures were grown on GaAs (100) substrates. High Tl flux was needed for the incorporation of Tl into the films.

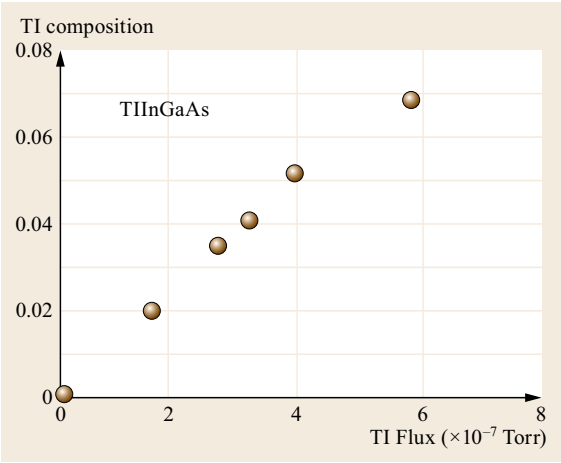


Fig. 23.4 Dependence of Tl composition in TlInGaAs on Tl flux during MBE growth. (After [23.33])

The TlInGaAsN/TlGaAsN QW samples having higher N concentration in the QWs showed higher Tl incorporation [23.37]. The addition of N species during growth induced the increase of Tl incorporation into TlInGaAs. This is considered due to the strong bonds of Tl–N.

23.1.3 Temperature Dependence of Physical Properties

TlInGaAs: Photoluminescence

Photoluminescence (PL) emission was observed from the grown TlInGaAs/InP samples in the temperature range of 10–300 K. With increasing Tl composition, the PL peak energy was shifted toward lower energy [23.38]. As the Tl composition increased, the temperature variation of the PL peak energy decreased and was much smaller than that of InAs, as shown in Fig. 23.5. It is noteworthy that the band-gap energy (≈ 0.8 eV) of TlInGaAs is larger than that (0.356 eV) of InAs. In general, the wider the band gap a material has, the larger is the variation of band-gap energy with temperature.

The temperature variation of as small as −0.03 meV/K was observed in Fig. 23.5, which corresponds to the wavelength variation of 0.04 nm/K. This value is much smaller than those of lasing wavelengths for the InGaAsP/InP Fabry–Pérot (FP) LDs (0.4 nm/K) as well as the InGaAsP/InP-distributed feedback (DFB) LDs (0.1 nm/K). The InGaAsP/InP

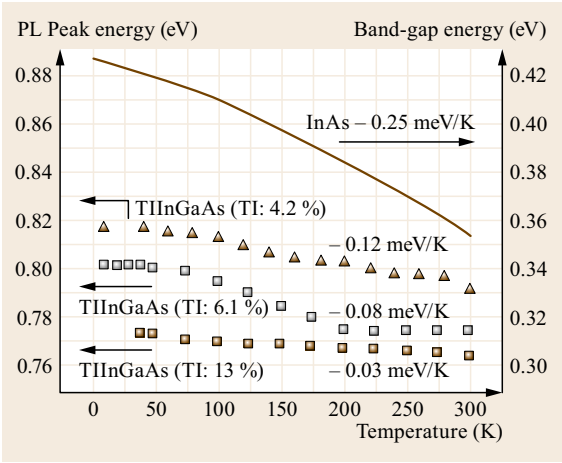


Fig. 23.5 Temperature variation of the PL peak energy for the TlInGaAs/InP DH samples as a function of Tl composition. (After [23.38])

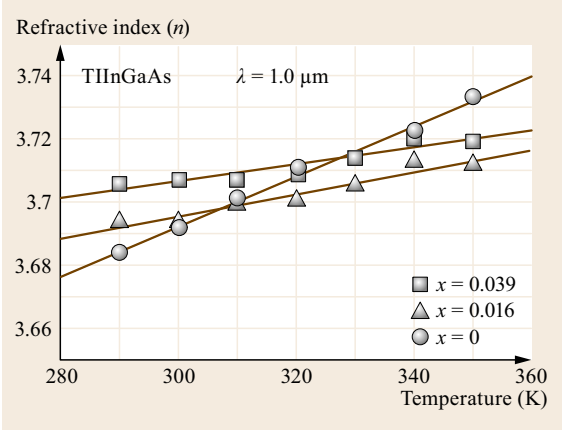


Fig. 23.6 Temperature variation of refractive indices for TlInGaAs (with different Tl composition x) and InGaAs at a wavelength of 1.0 μm. (After [23.39])

DFB LDs are presently used in the optical fiber communication systems. With further increase of Tl composition, real temperature-insensitive band-gap energy temperature dependence is expected.

TlInGaAs: Refractive Index and Band-Gap Energy

A spectroscopic ellipsometer (SE) was used to obtain the refractive indices of TlInGaAs in the photon energy range of 1.2–2.0 eV (above band-gap energy region) [23.39]. Optical reflectance and absorption measurements were carried out with a double-beam spectrometer in the wavelength range of 1700–2500 nm (below band-gap energy region) to determine the refractive indices n and the band-gap energies E_0 (E_g) of TlInGaAs, respectively [23.40].

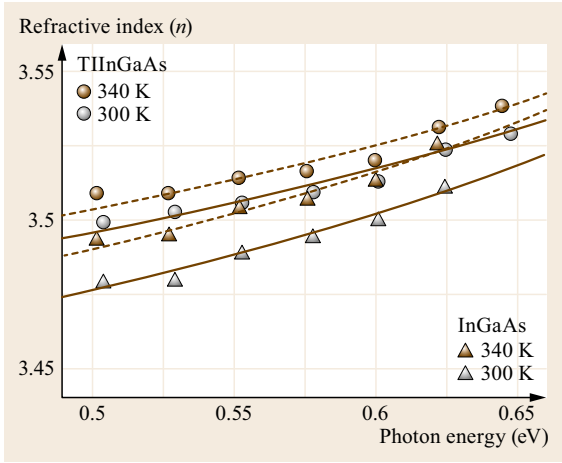


Fig. 23.7 Refractive index dispersions at $T = 300$ and 340 K. *Gray-colored circles* and *triangles* represent experimental refractive indices at 300 K and *brown-colored circles* and *triangles* represent those at 340 K for TlInGaAs (Tl = 0.043) and InGaAs, respectively. Solid and dashed lines represent the calculated results with the first-order Sellmeier equation. (After [23.40])

Table 23.1 Temperature coefficients of the direct-band edge E_0 and the refractive index $[n(\lambda \to \infty)]$ of InGaAs and TlInGaAs. (After [23.40])

Material	dE_0/dT (10^{-4} eV/K)	$n^{-1}(dn/dT)$ (10^{-5} K^{-1})	Ref.
InGaAs	−3.48		[23.41]
	−3.5	9.8	[23.40]
TlInGaAs	−2.1	5.0	[23.40]

At the energies above band gap, the refractive index n increases with photon energy [23.39]. However, the temperature variation of n for TlInGaAs (Tl = 0.039) greatly decreases as compared to InGaAs, as can be seen in Fig. 23.6 at the wavelength of 1.0 μm.

The photon energy dependencies of n at the energies below band gap for TlInGaAs (Tl = 0.043) and InGaAs at 300 and 340 K together with the calculated results by the first-order Sellmeier equation are shown in Fig. 23.7 [23.40]. The value of n of TlInGaAs is greater than that of InGaAs. The n value increases with increasing temperature. However, the temperature variation of n of TlInGaAs is remarkably smaller than that of InGaAs. Therefore, the temperature variation of n decreases with the increase of Tl composition at the photon energy both above and below E_g.

Table 23.1 [23.40] summarizes the temperature coefficients of the E_0 edge, dE_0/dT, and those of the

refractive index, $n^{-1}(dn/dT)$, for TlInGaAs (Tl = 0 and 0.043), together with those from [23.41]. Table 23.1 shows that the TlInGaAs alloy, although the Tl composition is limited to small value, exhibits temperature stability not only in the E_0 edge but also in n.

TlInGaAsN: Photoluminescence and Electroluminescence

TlInGaAsN/GaAs and InGaAsN/GaAs double quantum well (DQW) LEDs with GaAs layers as barrier layers were fabricated [23.36]. Reduction in the temperature variation of electroluminescence (EL) peak energy was observed in the temperature range of 288–328 K by the addition of Tl into InGaAsN/GaAs DQW layers; −0.62 meV/K for the InGaAsN/GaAs DQW and −0.53 meV/K for the TlInGaAsN/GaAs DQW. By replacing GaAs barrier layers with TlGaAs barrier layers, further reduction was obtained; −0.35 meV/K for TlInGaAsN/TlGaAs DQW LEDs [23.36]. SIMS measurements indicated that this improvement is caused by the increased incorporation of Tl into the DQW layers by using TlGaAs barrier layers.

TlInGaAsN/TlGaAsN triple QWs (TQWs) with TlGaAsN layers as barrier layers were investigated [23.37]. The TQW samples having higher N concentration in the TQWs have the highest Tl incorporation without deterioration of the crystalline quality. The temperature dependence of the PL peak energy was found to be the least for the highest Tl containing TQW sample.

23.1.4 Tl-III-V LD Application

TlInGaAs/InP double heterostructure (DH) LD structures were grown by GSMBE on (100) InP substrates [23.42]. Temperature variation of the EL peak wavelength as small as 0.06 nm/K was confirmed. This is much smaller than the 0.4 nm/K observed for InGaAsP/InP FP LDs as well as the 0.1 nm/K observed for InGaAsP/InP DFB LDs.

Laser operation was achieved under the pulsed condition from 77 to 310 K [23.42]. The threshold current density at room temperature was approximately 7 kA/cm^2, and the lasing wavelength was about 1660 nm. The characteristic temperature T_0 value for the temperature dependence of the threshold current in the temperature range of 77–300 K was approximately 94 K [23.42], which is similar to that of InGaAsP/InP LDs.

Several longitudinal-mode peaks were observed in the lasing spectra. Each longitudinal mode peak wavelength is determined by the effective refractive index of the TlInGaAs/InP waveguide and the cavity length of the LD. The temperature variation of each longitudinal-mode peak wavelength was as small as 0.06 nm/K [23.42]. However, in the wide temperature range, the main peak of the lasing spectrum moved to another peak of longer wavelength with increasing temperature. The average temperature variation of the main peak wavelength was 0.3 nm/K [23.42]. The temperature variation of the main peak for the FP LDs is mainly determined by the temperature variation of the gain peak. The observed average temperature variation of 0.3 nm/K is due to the temperature variation of the gain peak.

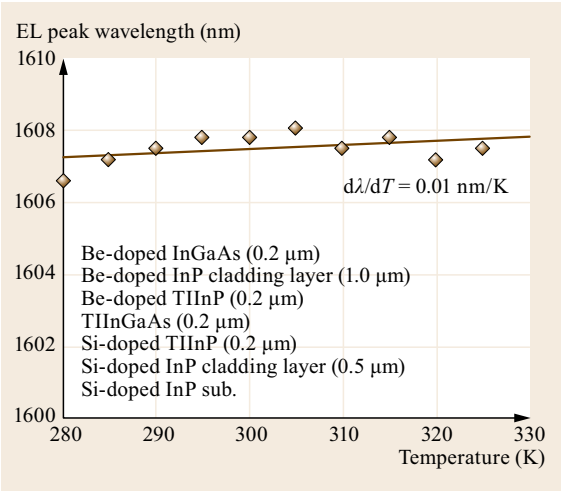


Fig. 23.8 Temperature dependence of EL peak wavelength for the TlInGaAs/TlInP/InP SCH LED. (After [23.43])

Laser light propagates along both TlInGaAs active layer and InP cladding layer; hence the temperature variation of lasing wavelength is influenced by their temperature variations. The addition of Tl into InP cladding layer results in the smaller temperature-dependent refractive index of cladding layer. The refractive index for TlInP was measured with the SE and it was confirmed that the refractive index of TlInP is larger than that of InP [23.43].

The InP/TlInP/TlInGaAs/TlInP/InP separate confinement heterostructure (SCH) LDs were fabricated and the improved temperature dependence of EL peak wavelength was obtained as shown in Fig. 23.8 [23.43]. Temperature dependence of the EL peak wavelength was reduced to 0.01 nm/K. The insertion of TlInP layer also increased the incorporated Tl composition in the TlInGaAs active layer.

The lasing spectra as a function of temperature and the temperature dependence of main peak wavelength for the SCH LDs under pulsed operation are shown in Fig. 23.9 [23.43]. The temperature

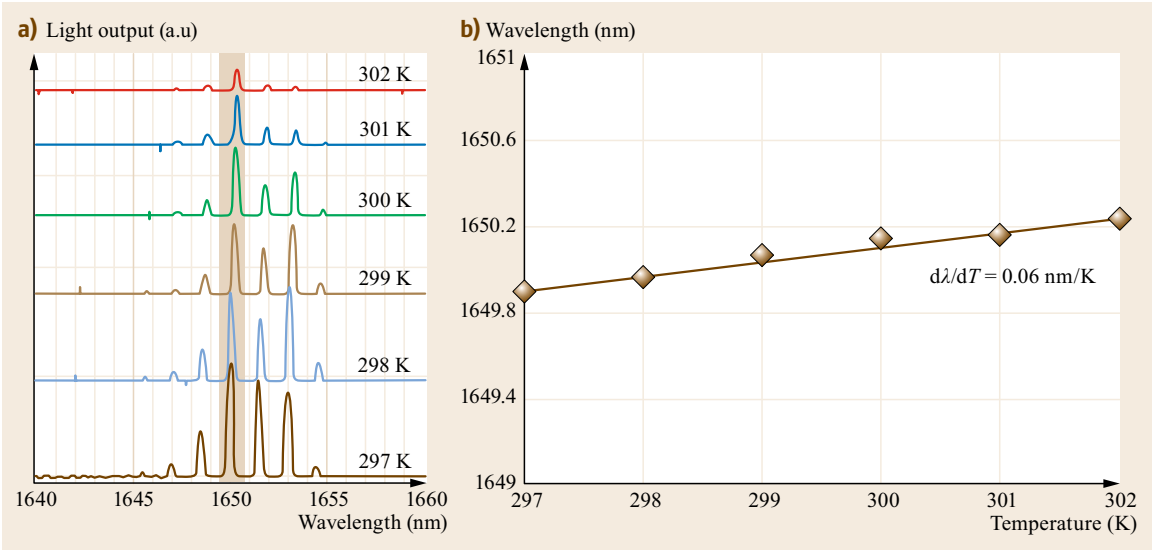


Fig. 23.9 (a) Lasing spectra as a function of temperature and (b) temperature dependence of lasing peak wavelength for the TlInGaAs/ TlInP/InP SCH LD under pulsed operation. (After [23.43])

variation of main peak wavelength is as small as 0.06 nm/K within the experimental temperature range (297–302 K). This value is much smaller than those of InGaAsP/InP FP LDs (0.4 nm/K) and InGaAsP/InP DFB LDs (0.1 nm/K). This reduced temperature variation is considered to be due to the reduced temperature dependence of refractive indices for TlInGaAs and TlInP compared with those for InGaAs and InP [23.43].

The observed temperature variation of the main peak wavelength for the TlInGaAs/TlInP/InP SCH LD is larger than that of the EL peak wavelength, which corresponds to that of gain peak or that of band-gap energy [23.43]. The main peak in the FP LD is one of the longitudinal-mode peaks, which are determined by the effective waveguide refractive index, that is, refractive indices of TlInGaAs active layers, TlInP cladding layers, and InP cladding layers, as well as the cavity (waveguide) length. Therefore, the temperature variation of the main peak cannot be smaller than that of the longitudinal-mode peak [23.43].

23.2 III-V-Bi Alloy Semiconductors

Bismuth-containing III-V (III-V-Bi) semiconductors were first studied for the applications to mid-infrared and far-infrared optical devices [23.44]. InSbBi [23.45–48], InAsBi [23.49, 50], and InAsSbBi [23.49] were grown by various growth methods, such as the Czochralski bulk growth method, RF sputtering, MBE, and MOVPE [23.44]. The solid solubility of Bi in InSb was reported to be as low as 2.6% in the bulk growth [23.45]. Low temperature MBE growth increased the Bi incorporation up to 5% [23.47].

Oe et al. [23.1] suggested the unique possible properties (temperature-independent band-gap energy characteristics) of III-V-Bi semiconductors and proposed the application to the temperature-insensitive wavelength LDs.

23.2.1 Expected Properties of Bi-Based III-V Alloys

Figure 23.10 shows the expected relationships between the band-gap energy and the lattice constant for the III-V-Bi alloy semiconductors [23.44]. As already described, the alloy semiconductors consisting of a semiconductor and a semimetal are expected to exhibit the temperature-insensitive band-gap energy similar to HgCdTe. The temperature-insensitive band gap is caused by the reverse temperature dependence of band gap (overlap) energy between the semiconductor and semimetal [23.1]. InGaAsBi was first suggested as a candidate for the temperature-insensitive band-gap energy semiconductor. GaAs and InAs are

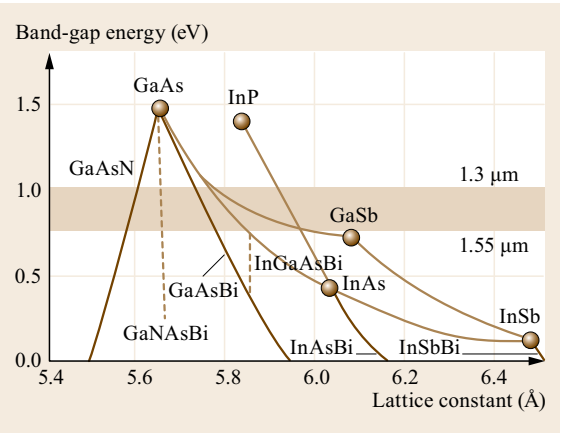


Fig. 23.10 Expected relationships between the band-gap energy and lattice constant for the III-V-Bi alloy semiconductors. (After [23.44])

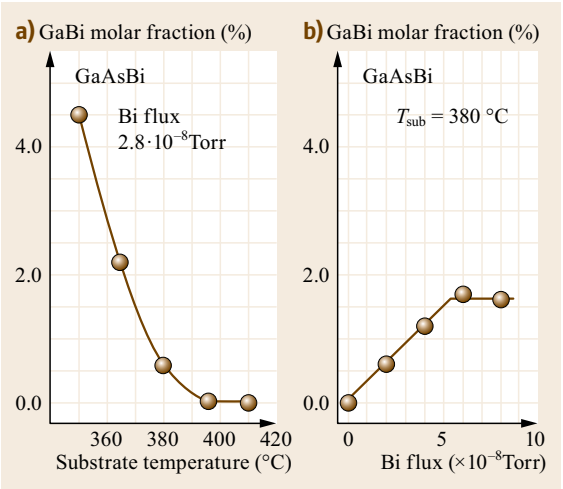


Fig. 23.11 **(a)** GaBi molar fraction in GaAsBi against substrate temperature for the constant Bi flux. **(b)** GaBi molar fraction in GaAsBi against Bi flux for the same substrate temperature. (After [23.52])

semiconductors and GaBi and InBi are semimetals.

To adjust the lasing wavelength of GaAsBi and related alloys to the waveband of optical fiber communication, a quaternary alloy of GaNAsBi was proposed which can be lattice matched to GaAs [23.44]. On the other hand, InGaAsBi is lattice matched to InP, but an additional element such as Al has to be added to InGaAsBi to adjust its band gap to the waveband of optical fiber communication [23.44]. GaAsBi and GaNAsBi have a characteristic of large spin–orbit interaction, which helps in reducing the Auger recombination in LDs [23.51].

23.2.2 Growth of III-V-Bi Alloys

Growth of III-V-Bi Ternary Alloys

In MBE growth of InSbBi, it is imperative to control the V/III ratio (Sb/In flux ratio) in the range of near unity to achieve Bi incorporation into InSb [23.44, 47]. Bi incorporation was suppressed at Sb/In ratio greater than unity despite the existence of Bi atoms on the growing surface. The strength of the In–Sb bond is considered to be stronger than that of In–Bi. Therefore, the Sb atoms are preferentially incorporated and the Bi atoms are difficult to be incorporated when the Sb/In flux ratio is greater than unity. Furthermore, the In–Bi bond is easily replaced by the In–Sb bond. Another important condition is that the growth temperature has to be set lower than 400 °C to avoid re-evaporation of Bi atoms from the growing surface [23.44, 47]. The InBi molar fraction increases by decreasing the substrate temperature (T_{sub}) [23.47].

After the proposal of GaAsBi alloys as a temperature-insensitive band-gap semiconductor [23.1], GaAsBi was grown on GaAs substrates by MOVPE [23.53]. To obtain the GaAsBi alloys, growth at relatively low temperatures was needed because Bi has high vapor pressure. However, in the MOVPE growth, the insufficient decomposition of MOs at low temperatures induced the incorporation of carbon (C) impurity species into GaAsBi films. On the other hand, in the MBE growth, elemental Ga, As, and Bi are used as sources [23.52, 54], and the incorporation of C impurity species is expected to be avoided.

GaAsBi was grown on GaAs (001) substrates at T_{sub} less than 400 °C by MBE [23.52, 54]. During the growth of GaAsBi, the RHEED pattern revealed the surface reconstruction of (2 × 1), suggesting the existence of Bi–Bi or Bi–As dimers [23.44, 55]. The unusual metallic nature of Bi-induced (2 × 1) was proposed as the origin of the surfactant nature of Bi [23.44, 56]. RHEED intensity oscillations observed during growth suggest a layer-by-layer growth [23.57].

For the growth of GaAsBi, the GaBi molar fraction decreased with increasing T_{sub} under constant Ga and As fluxes as shown in Fig. 23.11a [23.52]. At T_{sub} higher than 400 °C, Bi atoms are not incorporated into GaAs. Therefore, the GaAsBi alloys have to be grown at low temperature. However, the GaBi molar fraction increases with Bi flux, followed by the saturation for the constant Ga and As fluxes at the constant T_{sub} as shown in Fig. 23.11b. The unincorporated Bi atoms are desorbed from the surface without the formation of Bi droplets below critical Bi supply, although excess Bi supply forms Bi-containing droplets on the growth surface [23.52]. The droplet

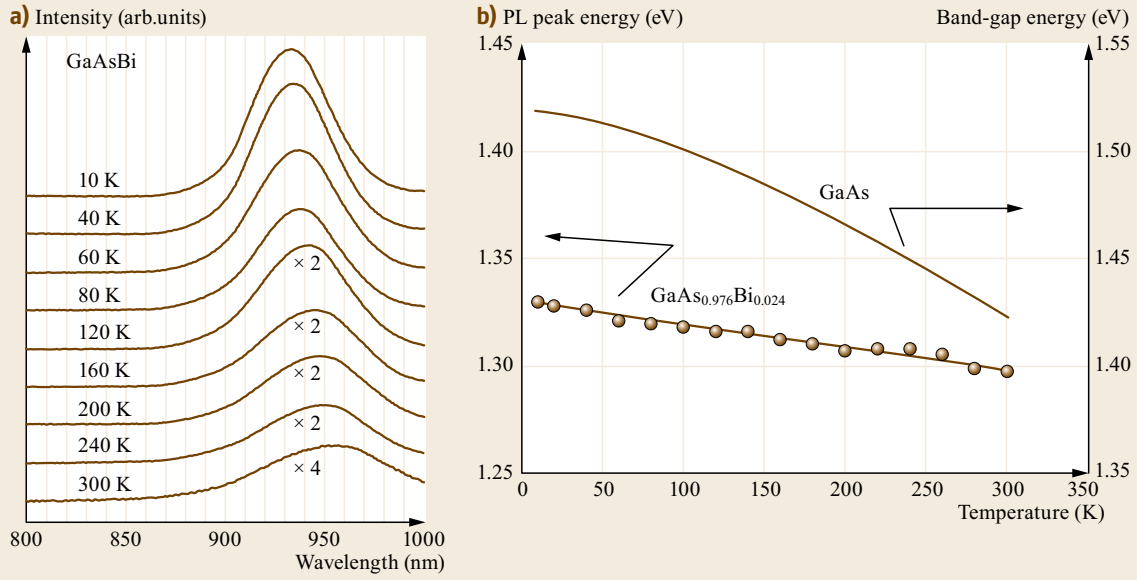


Fig. 23.12 **(a)** PL spectra as a function of temperature and **(b)** temperature variation of PL peak energy for the GaAsBi (GaBi = 0.024). For comparison, the band gap of GaAs is also shown. (After [23.53])

formation causes severe surface roughness and inhomogeneity.

The range of the As flux is also very narrow to obtain GaAsBi films [23.52]. For the growth at T_{sub} = 380 °C with Ga flux = 3×10^{-7} Torr and Bi flux = 2×10^{-8} Torr, only GaAs without Bi incorporation was grown above As flux of 10^{-5} Torr, while below 10^{-6} Torr the surface of the film becomes rough and only heavily degraded film was grown. Therefore, to grow the device quality GaAsBi, careful adjustment of the source supply and T_{sub} is needed.

Growth of III-V-Bi Quaternary Alloys

GaNAsBi quaternary alloys can be lattice matched to GaAs substrate with the band gap of optical fiber communication wavelengths. They were grown on GaAs (100) substrate by MBE using solid sources of Ga, As, and Bi and RF-plasma-enhanced N_2 gas at T_{sub} of 350–400 °C [23.58]. Precise control of As flux within a limited range and low temperature growth was crucial as in the case of GaAsBi. During growth, the (2×1) RHEED pattern was observed also similar to GaAsBi, implying Bi-stabilized reconstruction. The GaBi molar fraction decreased with the increase of T_{sub} for the constant Ga, As, and Bi fluxes, which is similar to the case of GaAsBi growth. Lattice matching between GaNAsBi film and GaAs substrate was achieved by adjusting GaN molar fraction [23.59].

The InGaAsBi quaternary alloy is another extension of GaAsBi and can be lattice matched to InP [23.60]. Considering the band gaps of InAs (0.36 eV) and GaAs (1.42 eV), the band gap of InGaAsBi is expected to cover the whole wavelength range used in the optical fiber communication systems, and InGaAsBi is a promising semiconductor to fabricate the temperature-insensitive wavelength LDs [23.60].

InGaAsBi was grown on InP (100) substrate by MBE using solid sources of In, Ga, As, and Bi [23.60]. T_{sub} was ranging from 270 to 350 °C, which is lower by 50 °C than that for GaAsBi growth, corresponding to the lower desorption temperature of Bi atoms from the InGaAs surface than that from the GaAs surface. During the growth of InGaAsBi, the (1×3) streaky RHEED pattern was observed. The XRD pattern showed a well-defined diffraction peak with distinct Pendellösung fringes, indicating the homogeneous alloy composition and high epitaxial quality with smooth surface. RBS channeling angular scans showed that Bi atoms are located exactly on the substitutional sites in InGaAs.

23.2.3 Optical Properties of III-V-Bi Alloys

Although the GaAsBi films were grown at low temperatures below 400 °C by both MOVPE [23.53] and MBE [23.58, 61, 62], intense PL emission was observed at room temperature. This is considered to be due to the surfactant effect of Bi atoms during growth. Figure 23.12 shows the PL spectra as a function of temperature and the temperature variation of the PL peak energy for the MOVPE-grown GaAsBi

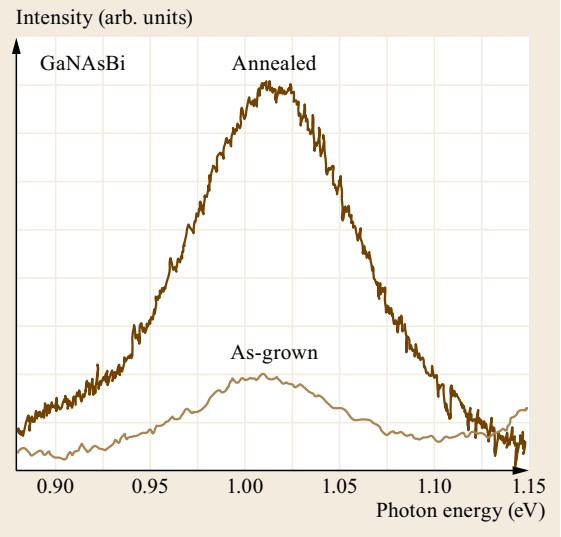


Fig. 23.13 PL spectra of as-grown and annealed GaNAsBi (GaBi = 0.02). (After [23.58])

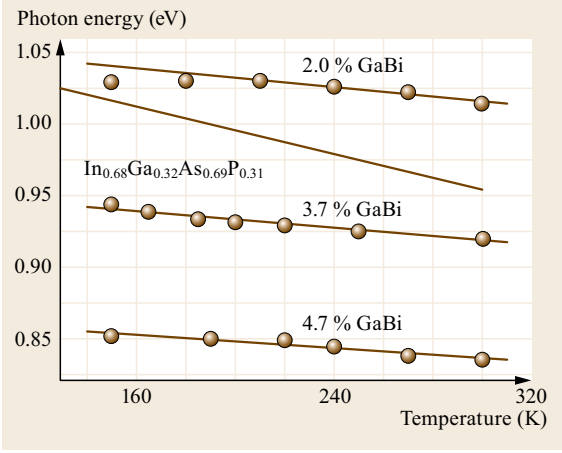


Fig. 23.14 Temperature variation of PL peak energy for GaNAsBi (GaBi = 0.02, 0.037, 0.047) and InGaAsP. (After [23.58])

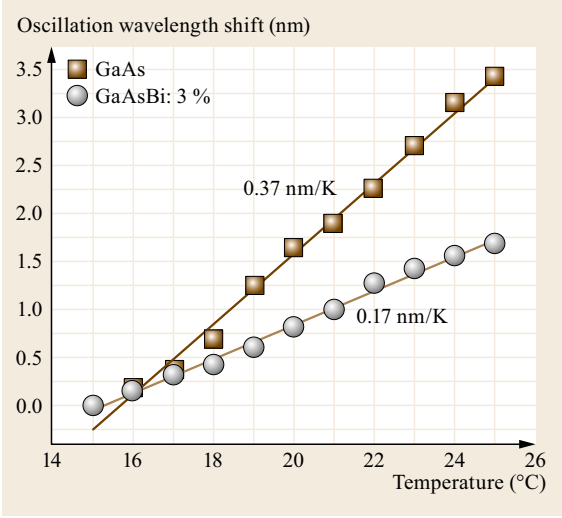


Fig. 23.15 Lasing wavelength shift for the GaAsBi and GaAs FP-LDs from the lasing wavelength at 15 °C in the temperature range of 15–25 °C. (After [23.64])

with GaBi molar fraction of 0.024 [23.53]. The temperature variation of the PL peak energy is as small as 0.1 meV/K, which is much smaller than that of GaAs as can be seen in Fig. 23.12b. The results suggest that the band gap of GaAsBi is less sensitive to temperature than the usual III-V semiconductors [23.53]. The PL peak energy decreased with the increase of GaBi molar fraction [23.58]. Reduction in the temperature variation of PL peak energy was also observed for the MBE-grown GaAsBi samples [23.61, 62]. The temperature coefficient of PL peak energy was 0.23 meV/K [23.61] and 0.15 meV/K [23.62] for the GaBi molar fraction of 0.013 and 0.025, respectively. The temperature coefficients of GaAsBi band gaps obtained by photoreflectance spectroscopy were 0.24, 0.23, and 0.15 meV/K between 150 and 300 K for samples with GaBi = 0.005, 0.013, and 0.026, respectively [23.63]; that is, with the increase of GaBi molar fraction the temperature coefficient of band gap decreased.

To obtain intense PL emission for the grown GaNAsBi, post-growth thermal annealing was needed, similar to GaInNAs. PL intensity of the GaNAsBi (GaBi = 0.02) film was increased by approximately fivefold after thermal annealing as shown in Fig. 23.13 [23.58]. PL peak energy of GaNAsBi showed no notable variation even after thermal annealing. The PL peak energy was decreased with the increase of GaBi molar fraction [23.58]. The temperature coefficient of the PL peak energy for GaNAsBi was very close to that for GaAsBi with the same GaBi molar fraction. Figure 23.14 shows the temperature variation of PL peak energy for three samples [23.58]. The temperature variation of PL peak energy for GaNAsBi is smaller than that of InGaAsP with similar band-gap energy (near 1300 nm). It can be said that the temperature coefficients of GaAsBi and GaNAsBi are governed by GaBi molar fraction and not by GaN molar fraction [23.58].

23.2.4 III-V-Bi LD Application

Photo-pumped lasing oscillation at 983 nm was achieved in the MBE-grown GaAs/GaAsBi/GaAs

(GaBi = 0.025) DH structure sample [23.62]. The temperature coefficient of the lasing wavelength decreased to 40% of that of the band gap of GaAs in the temperature range of 150–240 K. By using the GaAs/GaAsBi/AlGaAs DH structure sample to improve the carrier confinement, the photo-pumped lasing oscillation at the longer wavelength of up to 1200 nm was achieved [23.65]. The temperature coefficient of the lasing wavelength was about 40% to that of the 1300 nm InGaAsP LDs in the temperature range of 20–80 °C.

Electrically pumped lasing oscillation at 947 nm was demonstrated in the GaAsBi (GaBi = 0.022) FP-LD grown by MOVPE [23.66], though the temperature coefficient of the lasing wavelength is unclear. GaAsBi (GaBi = 0.03 or 0.04)/AlGaAs SCH FP-LD was grown by MBE [23.64]. Lasing oscillation was observed in the temperature range of 15–40 °C. The characteristic temperature T_0 for the GaAsBi (GaBi = 0.03)/AlGaAs SCH FP-LD was calculated to be 125 K, which is higher than that of 1300 nm InGaAsP FP-LD (T_0 = 60 K) and is similar to that for the 980 nm InGaAs/GaAs LDs. High T_0 is because of the stronger carrier confinement of the GaAsBi/AlGaAs system than that of the InGaAsP/InP system [23.64]. The temperature coefficient of lasing wavelength was reduced to 0.17 nm/K, which is 45% of that (0.37 nm/K) for the GaAs FP-LD as shown in Fig. 23.15 [23.64].

23.3 Summary

Tl-III-V and III-V-Bi alloy semiconductors are important materials for the fabrication of temperature-insensitive lasing wavelength LDs as well as LWIR optical devices. This chapter described the expected interesting properties of these semiconductors and the present status of their growth, characterization, and device applications. It was shown that these alloy semiconductors have a potential to have temperature-insensitive band gap and to fabricate temperature-insensitive lasing wavelength LDs.

References

23.1 K. Oe, H. Asai: IEICE Trans. Electron. **E79-C**, 1751 (1996)
23.2 D. Long: *Energy Band in Semiconductors* (Wiley, New York 1968)
23.3 W.S. Pelouch, L.A. Schline: Appl. Phys. Lett. **68**, 1389 (1996)
23.4 C.J. Summers, J.G. Broerman: Phys. Rev. **B21**, 559 (1980)
23.5 R. Dornhaus, G. Nimtz: *Springer Tracts in Modern Physics*, Vol. 78 (Spring, Berlin 1976)
23.6 H. Asahi, K. Yamamoto, K. Iwata, S. Gonda, K. Oe: Jpn. J. Appl. Phys. **35**, L876 (1996)
23.7 H. Asahi: Compound Semicond. **2**, 34 (1996)
23.8 M. van Schilfgaarde, A. Sher, A.B. Chen: Appl. Phys. Lett. **62**, 1857 (1994)
23.9 A. Sher, M. van Schilfgaarde, S. Krishnamurthy, M.A. Berding, A.B. Chen: J. Electron. Matter. **24**, 1119 (1995)
23.10 A.B. Chen, M. van Schilfgaarde, A. Sher: J. Electron. Matter. **22**, 843 (1993)
23.11 M. van Schilfgaarde, A.B. Chen, S. Krishnamurthy, A. Sher: Appl. Phys. Lett. **65**, 2714 (1994)
23.12 H. Asahi, K. Yamamoto, M. Fushida, H. Koh, K. Asami, S. Gonda, K. Oe: Proc. 9th Int. Conf. InP Relat. Mater. (1997) pp. 448–451
23.13 H.J. Lee, A. Mizobata, O. Maeda, K. Konishi, K. Asami, H. Asahi: Jpn. J. Appl. Phys. **41**, 1016 (2002)
23.14 M.A. Berding, M. van Schilfgaarde, A. Sher, M.J. Antonell, C.R. Abernathy: J. Electron. Matter. **26**, 683 (1997)
23.15 W. Mönch: Appl. Phys. Lett. **71**, 1231 (1997)
23.16 M. Kondow, M.C. Larson, T. Kitatani, K. Nakahara, K. Uomi: Proc. Conf. Indium Phosphide Relat. Mater. (1998) p. 233
23.17 C.K. Kim, T. Miyamoto, Y.H. Lee: Jpn. J. Appl. Phys. **37**, 5994 (1998)
23.18 Y.H. Choi, C. Besikci, R. Sudharsanan, M. Razeghi: Appl. Phys. Lett. **63**, 361 (1993)
23.19 K.T. Huang, R.M. Cohen, G.B. Stringfellow: J. Cryst. Growth **156**, 320 (1995)
23.20 N.H. Karam, R. Sudharsanan, T. Parodos, M.A. Dodd: J. Electron. Matter. **25**, 1209 (1996)
23.21 K. Yamamoto, H. Asahi, M. Fushida, K. Iwata, S. Gonda: J. Appl. Phys. **81**, 1704 (1997)
23.22 H. Asahi, M. Fushida, K. Yamamoto, K. Iwata, H. Koh, K. Asami, S. Gonda, K. Oe: J. Cryst. Growth **175/176**, 1195 (1997)
23.23 M. Fushida, H. Asahi, K. Yamamoto, H. Koh, K. Asami, S. Gonda, K. Oe: Jpn. J. Appl. Phys. **36**, L665 (1997)
23.24 H. Koh, H. Asahi, M. Fushida, K. Yamamoto, K. Asami, S. Gonda, K. Oe: J. Cryst. Growth **188**, 107 (1998)
23.25 K. Takenaka, H. Asahi, H. Koh, K. Asami, S. Gonda, K. Oe: Jpn. J. Appl. Phys. **38**, 1078 (1999)
23.26 M.J. Antonell, C.R. Abernathy, A. Sher, M. Berding, M. van Schilfgaarde, A. Sanjuro, K. Wong: J. Cryst. Growth **188**, 13 (1998)

23.27 M.D. Lange, D.F. Storm, T. Cole: J. Electron. Matter. **27**, 36 (1998)
23.28 D.L. Lubyshev, W.Z. Cai, G.L. Catchen, T.S. Mayer, D.L. Miller: Proc. IEEE 24th Intern. Symp. Compound Semicond. (1997) p. 125
23.29 M. Razeghi, J.D. Kim, S.J. Park, Y.H. Choi, D. Wu, E. Michel, J. Xu, E. Bigan: Inst. Phys. Conf. Ser. **145**, 1085 (1996)
23.30 J.D. Kim, E. Michel, S. Park, J. Xu, S. Javadpour, M. Razeghi: Appl. Phys. Lett. **69**, 343 (1996)
23.31 P.T. Staveteig, Y.H. Choi, G. Labeyrie, E. Bigan, M. Razeghi: Appl. Phys. Lett. **64**, 60 (1994)
23.32 M.J. Antonell, C.R. Abernathy, A. Sher, M. Berding, M. van Schilfgaarde: Proc. of 9th Intern. Conf. InP Relat. Mater. (1997) pp. 444–447
23.33 H. Asahi, H. Koh, K. Takenaka, K. Asami, K. Oe, S. Gonda: J. Cryst. Growth **201/202**, 1069 (1999)
23.34 H. Asahi, K. Takenaka, H. Koh, K. Oe, K. Asami, S. Gonda: Inst. Phys. Conf. Ser. **162**, 541 (1999)
23.35 H. Asahi, S. Hasegawa: *IPAP Book 2* (The Institute of Pure and Applied Physics, Tokyo 2005) p. 91
23.36 T. Matsumoto, D. Krishnamurthy, A. Fujiwara, S. Hasegawa, H. Asahi: J. Cryst. Growth **295**, 133 (2006)
23.37 K.M. Kim, S. Emura, D. Krishnamurthy, S. Hasegawa, H. Asahi: J. Appl. Phys. **108**, 053501 (2010)
23.38 A. Ayabe, H. Asahi, H.J. Lee, O. Maeda, K. Konishi, K. Asami, S. Gonda: Appl. Phys. Lett. **77**, 2148 (2000)
23.39 A. Imada, H.-J. Lee, A. Fujiwara, S. Emura, S. Hasegawa, H. Asahi: J. Appl. Phys. **94**, 6976 (2003)
23.40 A. Imada, H.J. Lee, A. Fujiwara, T. Mukai, H. Asahi: Appl. Phys. Lett. **84**, 4212 (2004)
23.41 E. Zielinski, H. Schweizer, K. Streubel, H. Eisele, G. Weimann: J. Appl. Phys. **59**, 2196 (1986)
23.42 A. Fujiwara, H.J. Lee, A. Imada, S. Hasegawa, H. Asahi: Jpn. J. Appl. Phys. **42**, L1359 (2003)
23.43 A. Fujiwara, D. Krishnamurthy, T. Matsumoto, S. Hasegawa, H. Asahi: J. Cryst. Growth **301**, 109 (2007)
23.44 M. Yoshimoto, K. Oe: Molecular beam epitaxy of GaAsBi and related alloys. In: *Molecular Beam Epitaxy*, ed. by M. Henini (Elsevier, Oxford 2013) p. 159
23.45 B. Joukoff, A.M. Jean-Louis: J. Cryst. Growth **12**, 169 (1972)
23.46 J.L. Zilko, J.E. Greene: J. Appl. Phys. **51**, 1549 (1980)
23.47 K. Oe, S. Ando, K. Sugiyama: Jpn. J. Appl. Phys. **20**, L303 (1981)
23.48 A.J. Noreika, W.J. Takei, M.H. Francombe: J. Appl. Phys. **53**, 4932 (1982)
23.49 Y.K. Ma, Z.M. Fang, D.H. Jaw, R.M. Cohen, G.B. Stringfellow: Appl. Phys Lett. **55**, 2420 (1989)
23.50 H. Ofuchi, T. Kubo, M. Tabuchi, Y. Takeda, H. Okamoto, K. Oe: Jpn. J. Appl. Phys. **38**, 545 (1999)
23.51 M. Yoshimoto, T. Fuyuki: Localized states in GaAsBi and GaAs/GaAsB heterostructures. In: *Bismuth-Containing Compounds*, ed. by H. Li, Z.M. Wang (Springer, New York 2013) pp. 201–224
23.52 M. Yoshimoto, S. Murata, A. Chayahara, Y. Horino, J. Saraie, K. Oe: Jpn. J. Appl. Phys. **42**, L1235 (2003)
23.53 K. Oe, H. Okamoto: Jpn. J. Appl. Phys. **37**, L1283 (1998)
23.54 S. Tixier, M. Adamcyk, T. Tiedje, S. Francoeur, A. Mascarenhas, P. Wei, F. Schiettekatte: Appl. Phys. Lett. **82**, 2245 (2003)
23.55 P. Laukkannen, M.P.J. Punkkinen, H.-P. Komsa, M. Ahola-Tuomi, K. Kokko, M. Kuzumin, J. Adell, J. Sadowski, R.E. Peraelae, M. Ropo, T.T. Rantala, I.J. Vaeyrynen, M. Pessa, L. Vitos, J. Kollar, S. Mirbt, B. Johansson: Phys. Rev. Lett. **100**, 086101 (2008)
23.56 M.P.J. Punkkinen, P. Laukkannen, H.-P. Komsa, M. Ahola-Tuomi, M. Raesaenen, K. Kokko, M. Kuzmin, J. Adell, J. Sadowski, R.E. Peraelae, M. Ropo, T.T. Rantala, I.J. Vaeyrynen, M. Pessa, L. Vitos, J. Kollar, S. Mirbt, B. Johansson: Phys. Rev. B **78**, 195304 (2008)
23.57 Y. Tominaga, Y. Kinoshita, K. Oe, M. Yosimoto: Appl. Phys. Lett. **93**, 131915 (2008)
23.58 W. Huang, K. Oe, G. Feng, M. Yoshimoto: J. Appl. Phys. **98**, 053505 (2005)
23.59 W. Huang, M. Yoshimoto, Y. Takehara, J. Saraie, K. Oe: Jpn. J. Appl. Phys. **43**, L1350 (2004)
23.60 G. Feng, M. Yoshimoto, K. Oe, A. Chayahara, Y. Horino: Jpn. J. Appl. Phys. **44**, L1161 (2005)
23.61 G. Pettinari, A. Polimeni, M. Capizzi, J.H. Blokland, P.C.M. Christianen, J.C. Maan, E.C. Young, T. Tiedje: Appl. Phys. Lett. **92**, 262105 (2008)
23.62 Y. Tominaga, K. Oe, M. Yosimoto: Appl. Phys. Express **3**, 062201 (2010)
23.63 J. Yoshida, T. Kita, O. Wada, K. Oe: Jpn. J. Appl. Phys. **42**, 371 (2003)
23.64 T. Fuyuki, K. Yoshida, R. Yoshida, M. Yoshimoto: Appl. Phys. Express **7**, 082101 (2014)
23.65 T. Fuyuki, R. Yoshida, K. Yoshida, M. Yoshimoto: Appl. Phys. Lett. **103**, 202105 (2013)
23.66 P. Ludewig, N. Knaub, N. Hossain, S. Reinhard, L. Nattermann, I.P. Marko, S.R. Jin, H. Hild, S. Chatterjee, W. Stols, S.J. Sweeney, K. Volt: Appl. Phys. Lett. **102**, 242115 (2013)

24. Amorphous Semiconductors: Structure, Optical, and Electrical Properties

Kazuo Morigaki, Chisato Ogihara

This chapter is devoted to a survey of the structural, optical and electrical properties of amorphous semiconductors on the basis of their fundamental understanding. These properties are important for various types of applications using amorphous semiconductors.

First, we review general aspects of the electronic states and defects in amorphous semiconductors, i. e., a-Si:H and related materials, and chalcogenide glasses, and their structural, optical and electrical properties.

Further, we survey the two types of phenomena associated with amorphous structure, i. e., light-induced phenomena, and quantum phenomena associated with nanosized amorphous structure. The former are important from the viewpoint of amorphous-silicon solar cells. The latter phenomena promise novel applications of amorphous semiconductors from the viewpoint of nanotechnology.

24.1 **Electronic States** 557
24.2 **Structural Properties** 560
24.2.1 General Aspects 560
24.2.2 a-Si:H and Related Materials 560
24.2.3 Chalcogenide Glasses 561
24.3 **Optical Properties** 561
24.3.1 General Aspects 561
24.3.2 a-Si:H and Related Materials 563
24.3.3 Chalcogenide Glasses 564
24.4 **Electrical Properties** 565
24.4.1 General Aspects 565
24.4.2 a-Si:H and Related Materials 566
24.4.3 Chalcogenide Glasses 567
24.5 **Light-Induced Phenomena** 567
24.6 **Nanosized Amorphous Structure** 569
References 570

Amorphous semiconductors are promising electronic materials for a wide range of applications such as solar cells, thin-film transistors, light sensors, optical memory devices, vidicons, electrophotographic applications, x-ray image sensors, europium-doped optical-fibre amplifications etc., particularly, hydrogenated amorphous silicon (a-Si:H) for solar cells, thin-film transistors, x-ray image sensors, and chalcogenide glasses for optical memory devices including digital video/versatile disk (DVD). In this chapter, we emphasize the basic concepts and general aspects of the electronic properties of amorphous semiconductors such as their electrical and optical properties as well as their structural properties [24.1–3]. Furthermore, some basic and important results of these properties are described to understand these applications and to consider their further development. Light-induced phenomena in amorphous semiconductors, which have been considered to be associated with amorphous structure, are also described.

Nanosized amorphous structures exhibit quantum effects associated with two-dimensional (quantum well), one-dimensional (quantum wire) and zero-dimensional (quantum dot) structures, so they have received significant attention on both the fundamental and application sides. These topics are briefly described.

24.1 Electronic States

Long-range disorder in amorphous network breaks down the periodic arrangement of constituent atoms, as shown in Fig. 24.1. In the figure, the structures of amorphous and crystalline silicon are shown. The periodic arrangement of atoms makes it easy to treat the electronic states mathematically, i. e., the so-called Bloch

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_24

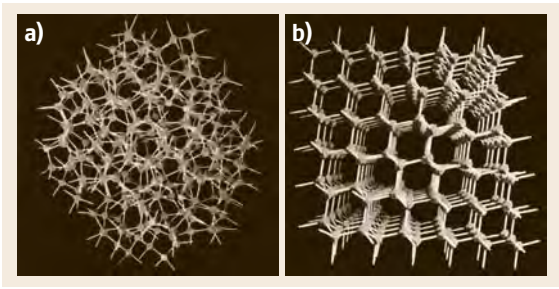


Fig. 24.1a,b Structural models of (a) amorphous silicon and (b) crystalline silicon

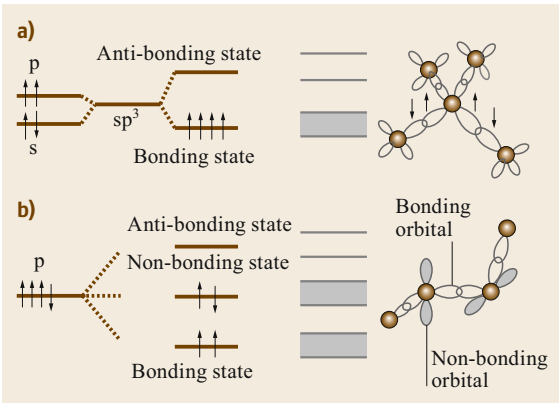


Fig. 24.2a,b Schematic diagram of the energy levels of atomic orbitals, hybridized orbitals and bands for (a) tetrahedrally bonded semiconductors, e.g., a-Si and for (b) selenium

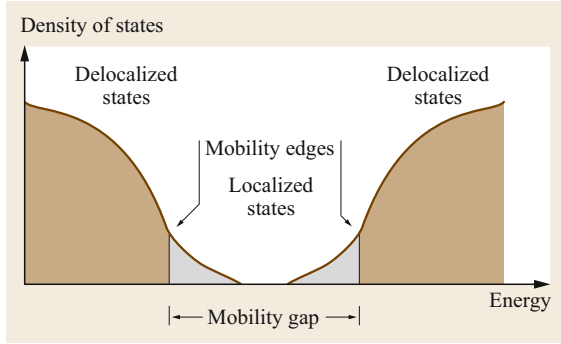


Fig. 24.3 Schematic diagram of the density of states of the conduction band (delocalized states) and the valence band (delocalized states) shown by the *brown-colored*. The mobility edge and mobility gap are shown

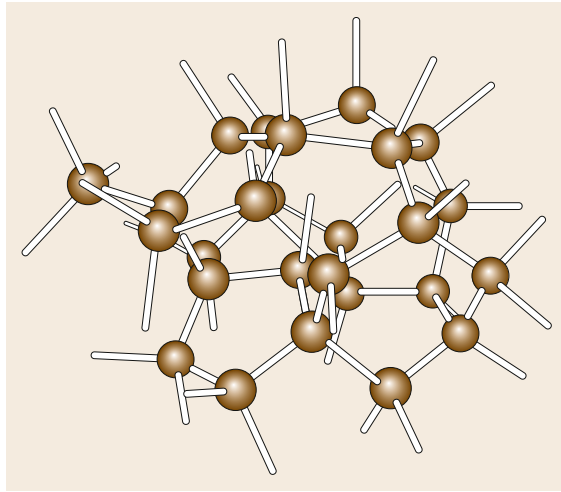


Fig. 24.4 Schematic illustration of a dangling bond in tetrahedrally bonded amorphous semiconductors

theory can be applied to crystalline solids. On the other hand, it becomes difficult to treat the electronic states in amorphous solids mathematically. However, electronic states in amorphous semiconductors are simply described in terms of tight-binding approximations and Hartree–Fock calculations. Using these approaches, spatial fluctuations of bond length, bond angle and dihedral angle lead to broadening of the edges of the conduction and valence bands, constructing the band tail. This is due to spatial fluctuations of bond energy between constituent atoms in tetrahedrally coordinated semiconductors such as Si and Ge, in which the bonding state and the antibonding state constitute the valence band and the conduction band, respectively, as shown in Fig. 24.2a. In chalcogenide glasses, the conduction band arises from the antibonding state, while the valence band arises from the nonbonding state, as shown in Fig. 24.2b. The broadening of those bonds also occurs as a result of potential fluctuations associated with the amorphous network of constituent atoms. The static charge fluctuation associated with bond-length and bond-angle variations in the amorphous network has been discussed theoretically [24.4] and experimentally [24.5, 6] in amorphous silicon.

In the band tails, the electronic states have a localized character and their nature changes from localized to delocalized at a critical boundary called the mobility edge, as shown in Fig. 24.3. The energy separation between the two mobility edges of the conduction and valence bands is called the mobility gap. The nature of the conduction and valence bands has been elucidated by means of photoemission spectroscopy [24.7]. The structure of actual samples of amorphous semiconductors deviates from the ideal random network, namely, coordination of constituent atoms deviates from the normal coordination following the $8-N$ rule [24.1], where N designates the number of valence electrons. For instance, the normal coordinations of Si and Se are four and two, respectively. However, in actual samples Si

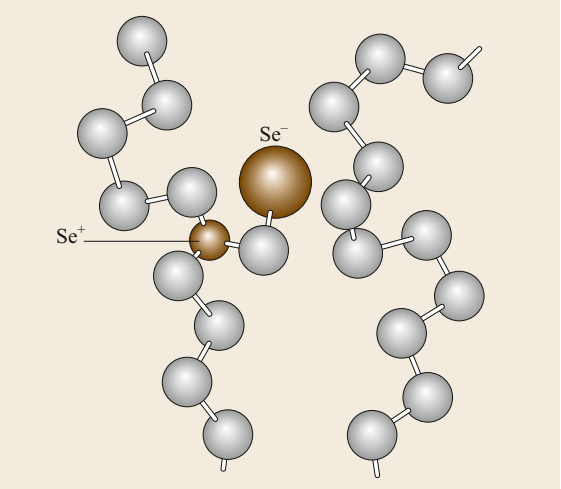


Fig. 24.5 Schematic illustration of two selenium chains with positively charged threefold-coordinated selenium and negatively charged onefold-coordinated selenium

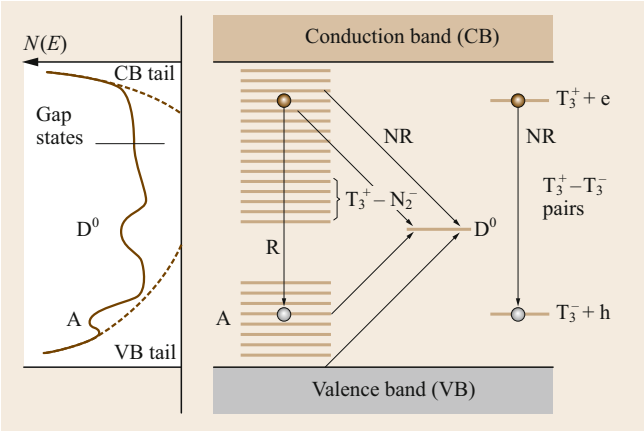


Fig. 24.6 Schematic diagram of the tail and gap states and the recombination processes in a-Si:H. R and NR designate radiative and nonradiative recombination, respectively. A diagram of the density of states spectrum is illustrated on the *left-hand side*. For the definition of other symbols, see the text. (After [24.9])

atoms with a threefold coordination are present in a-Si (Fig. 24.4), and Se atoms with onefold and threefold coordinations are present in a-Se (Fig. 24.5). These atoms are generally called structural defects (gap states) whose electronic energy levels are located within the band-gap region. In the case of chalcogenide glasses, onefold- and threefold-coordinated Se atoms are called valence-alternation pairs (VAP) [24.8]. The detailed properties of structural defects have been elucidated by means of optical spectroscopies such as photoinduced absorption (PA), photothermal deflection spectroscopy (PDS) and the constant-photocurrent method (CPM), capacitance measurements such as deep-level transient spectroscopy (DLTS) and isothermal capacitance transient spectroscopy (ICTS), and resonance methods such as electron spin resonance (ESR) and electron–nuclear double resonance (ENDOR) [24.2].

In the following, we take a-Si:H as an example and describe the nature of its electronic states. Figure 24.6 shows a schematic diagram of the electronic states involved in the band-gap region of a-Si:H proposed by the authors' group [24.9]. The tails of the conduction and valence bands are expressed by an exponential function with widths of 25 meV [24.10] and 48−51 meV [24.11–13], respectively.

In this figure, the processes of radiative recombination and nonradiative recombination are shown, along with the energy levels of the gap states involved in the recombination processes. The radiative recombination occurs between tail electrons and tail holes at higher temperatures and between tail electrons and self-trapped holes (A centres) at lower temperatures, particularly below 60 K. Such recombinations contribute to the principal luminescenceband peaked at 1.3−1.4 eV. These have been elucidated by optically detected magnetic resonance (ODMR) measurements [24.14–16], and are briefly mentioned in Sect. 24.3. The radiative recombination center participating in the low-energy luminescence (defect luminescence) may be attributed to distant T_3^+–N_2^- pair defects [24.17, 18], in which T_3^+ and N_2^- are the positively charged threefold-coordinated silicon center and the negatively charged twofold-coordinated nitrogen center, respectively. The N_2^- centers are created from contaminating nitrogen atoms introduced during sample preparation. On the other hand, close T_3^+–N_2^- pair defects act as nonradiative recombination centers. A typical nonradiative recombination center in a-Si:H is the neutral silicon dangling bond, i. e., T_3^0. The details of the recombination processes in a-Si:H are described in Sect. 24.3.

24.2 Structural Properties

24.2.1 General Aspects

The structural properties of amorphous semiconductors have been investigated by means of x-ray, electron and neutron diffraction, transmission electron microscope (TEM) and scanning electron microscope (SEM), extended x-ray absorption fine structure (EXAFS), small-angle x-ray scattering (SAXS), Raman scattering, infrared absorption (IR) and nuclear magnetic resonance (NMR). The absence of long-range order in amorphous semiconductors is manifested in diffraction techniques, e.g., electron diffraction for a-Si results in a halo pattern as shown in Fig. 24.7a, while Laue spots are seen for c-Si, as shown in Fig. 24.7b. The medium-range order of 0.5–50 nm has been discussed based on SAXS measurements.

24.2.2 a-Si:H and Related Materials

The radial distribution function (RDF) in a-Si has been obtained from an analysis of curves of scattered electron intensity versus scattering angle from electron diffraction measurements, as shown in Fig. 24.8 [24.19]. From a comparison of RDF curves of a-Si and c-Si, it was concluded that the first peak of a-Si coincides with that of c-Si, the second peak of a-Si is broadened compared to that of c-Si, the third peak of c-Si almost disappears and a small peak appears, shifted from the third peak of c-Si, as shown in Fig. 24.8. These results indicate that the bond length of a-Si is elongated by 1% over that of c-Si (= 2.35 Å), and that the bond angle is tetrahedral (= 109.47°) with a variation of 10% from that of c-Si. Using the neutron diffraction technique, the RDF was also deduced from the measured structure factor, $S(Q)$, as a function of Q (the momentum transfer) in a-Si [24.20]. The RDF curve of a-Si exhibiting these features was well simulated by the continuous random network model [24.21]. Since then, computer-generated models have been constructed (see, e.g., *Kugler* et al. [24.20]), using the Monte Carlo (MC) method [24.22], the molecular dynamics (MD) method [24.23] and the reverse MC method [24.24, 25]. Here, the reverse MC method, i. e., a MC simulation repeatedly carried out to reach consistency with the measured RDF curve and curve of structure factor $S(Q)$ versus Q from the neutron diffraction data, is briefly mentioned. First we start with an initial set of Cartesian coordinates (particle configuration) and calculate its RDF and $S(Q)$. Comparing the calculated RDF and $S(Q)$ with the measured ones, the new particle configuration is generated by random motion of a particle, being consistent with the constraint, e.g., the coordination number. From the reverse MC simulation, the bond-angle distribution is derived, i. e., it shows a peak at the tetrahedral angle, 109.5° and small peaks at ≈ 60° and ≈ 90° [24.26].

In a-Si:H, knowledge of the hydrogen configuration is important to understand its electronic properties. This is obtained by NMR [24.27, 28] and IR measurements [24.29]. The bonding modes of Si and H atoms are Si–H, $Si–H_2$, $(Si–H_2)_n$ and $Si–H_3$ bonds. The concentrations of these hydrogen configurations depend on the preparation conditions, particularly on the substrate temperature, i. e., the deposition temperature. a-Si:H films prepared at 250 °C mostly contain Si–H bonds in incorporated hydrogen (hydrogen content of ≈ 10 at.%), while those prepared at lower temperatures such as room temperature contain all these bonds in incorporated hydrogen (hydrogen content of ≈ 30 at.%). The hydrogen configuration has also been investigated by neutron scattering measurements using the isotope-substitution method [24.30, 31]. Partial pair-correlation functions of Si–Si, Si–H and H–H are

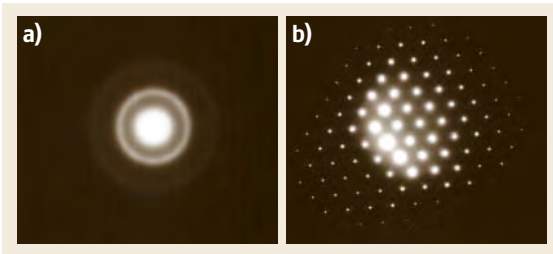


Fig. 24.7a,b Electron diffraction patterns. **(a)** amorphous silicon, **(b)** crystalline silicon

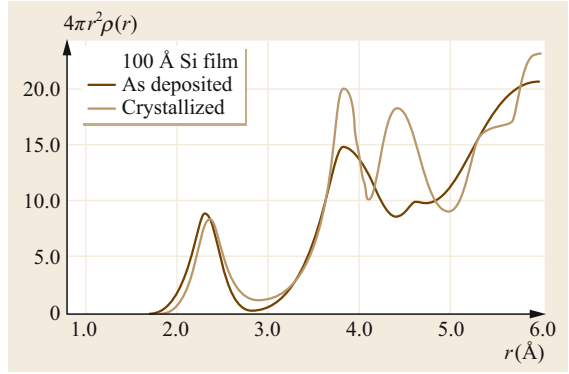


Fig. 24.8 Radial distribution function of amorphous (evaporated) and crystalline silicon obtained from electron diffraction patterns. (After [24.19])

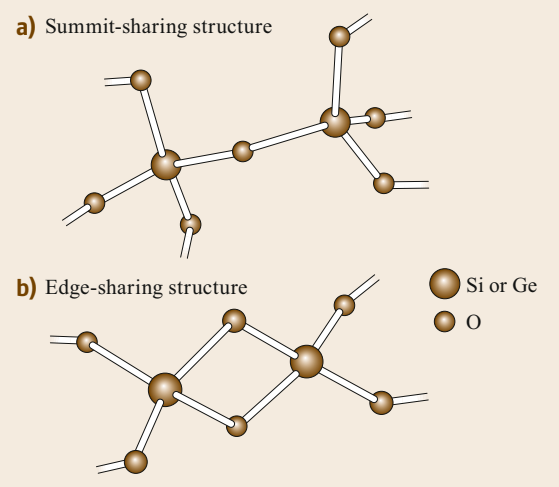


Fig. 24.9a,b Schematic illustration of (a) summit-sharing structure and (b) edge-sharing structure. (After [24.33])

obtained from those measurements. The Si−H bond length is 1.48 Å.

Raman scattering measurements provide useful information about microcrystallinity, i. e., the volume fraction of the amorphous phase in microcrystalline silicon [24.32]. The Raman shift associated with transverse optical (TO) phonons is observed at 520 cm^{-1} for c-Si, while it is observed at 480 cm^{-1} for a-Si.

24.2.3 Chalcogenide Glasses

The chalcogenide glasses are composed of chalcogenide atoms and other constituent atoms such as As, Ge etc. Their crystalline counterpart has a structure consisting of threefold-coordinated As atoms and twofold-coordinated chalcogens, i. e., c-Se and c-As_2S_3 exhibit a chain structure and a layered structure, respectively. The structure of the chalcogenide glasses is fundamentally built from these constituent atoms, keeping their coordination in the crystalline counterpart. However, actual glasses contain wrong bonds, i. e., homopolar bonds such as Se−Se, As−As, Ge−Ge etc. In the following, we take two examples of chalcogenide glasses, As_2X_3 (X = S, Se) and GeX_2 (X = S, Se):

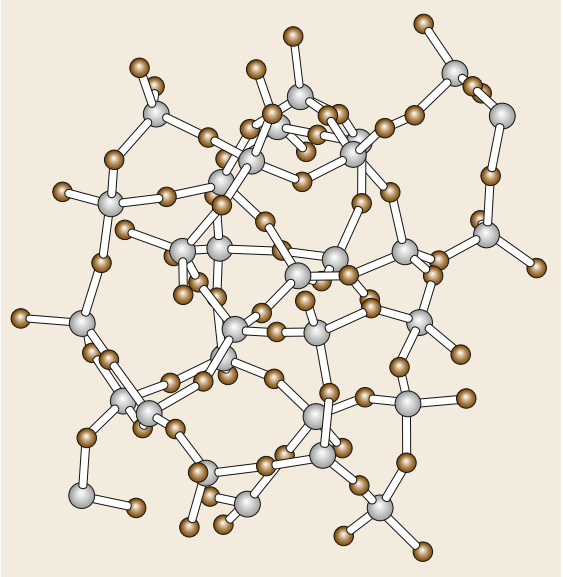


Fig. 24.10 Structural model of germanium disulfide. (After [24.33])

- As_2X_3 (X = S, Se): This type of glass, e.g., As_2S_3, is mainly composed of AsS_3 pyramid units, i. e., a random network of these units sharing the twofold-coordinated S site [24.34]. There is no correlation between As atoms, but interlayer correlation seems to hold even for the amorphous structure.
- GeX_2 (X = S, Se): Silica glass (SiO_2) is a well-known material, which is mainly composed of SiO_4 tetrahedron units sharing the twofold-coordinated O site (the summit-sharing structure. See Fig. 24.9a). On the other hand, SiS_2 and $GeSe_2$ glasses have a different structure from SiO_2 glass: the tetrahedron units are connected through two X sites (the edge-sharing structure) along with one X site (the summit-sharing structure. See Fig. 24.9b). For example, in $GeSe_2$, the edge-sharing structure occurs at a level of more than 30% and the summit-sharing structure with less than 70%. Actually, wrong bonds (homopolar bonds) Ge−Ge, Se−Se exist in $GeSe_2$ glass. A structural model of GeS_2 [24.33] is shown in Fig. 24.10.

24.3 Optical Properties

24.3.1 General Aspects

Optical absorption and luminescence occur by transition of electrons and holes between electronic states such as conduction and valence bands, tail states, and gap states. In some cases, electron–phonon coupling is strong and, as a result, self-trapping occurs. Exciton formation has also been suggested in some amorphous semiconductors.

The absorption of photons due to interband transition, which occurs in crystalline semiconductors, is also observed in amorphous semiconductors. However,

the absorption edge is not clear since interband absorption near the band gap is difficult to distinguish from tail absorption in the absorption spectra. Figure 24.11 schematically illustrates typical absorption spectra of amorphous semiconductors. The absorption coefficient, α, due to interband transition near the band gap is known to be well described by the following equation [24.35]

$$\alpha\hbar\omega = B(\hbar\omega - E_g)^2 \,, \tag{24.1}$$

where $\hbar\omega$ and E_g denote the photon energy and optical gap, respectively. In most amorphous semiconductors, the optical gap E_g is determined by a plot of $(\alpha\hbar\omega)^{1/2}$ versus $\hbar\omega$, which is known as Tauc's plot. The photon energy at which the absorption coefficient is $10^4\,\mathrm{cm}^{-1}$, E_{04}, is also used for the band gap in a-Si:H. The absorption coefficient at the photon energy just below the optical gap (tail absorption) depends exponentially on the photon energy, $E = \hbar\omega$, as expressed by

$$\alpha(E) \propto \exp\left(\frac{E}{E_U}\right) , \tag{24.2}$$

where E_U is called the Urbach energy and also the Urbach tail width. In addition, optical absorption by defects appears at energies lower than the optical gap.

Photoluminescence (PL) occurs as a result of the transition of electrons and holes from excited states to the ground state. After interband excitation, electrons relax to the bottom of the conduction band by emitting phonons much more quickly than the radiative transition. Similarly, the holes also relax to the top of the valence band. In the case of crystalline semiconductors without defects or impurities, there is no localized state in the band gap and PL occurs by transition between the bottom of the conduction band and the top of the valence band. In this case the k-selection rule, $k_{photon} = k_i - k_f$, must be satisfied, where k_{photon}, k_i and k_f denote the wavenumbers of photons, and electrons in the initial and final states, respectively. Since k_{photon} is much smaller than k_i and k_f, we can rewrite the selection rule as $k_i = k_f$. Semiconductors satisfying this condition are called direct-gap semiconductors. Crystalline silicon is one semiconductor in which the direct transition is not allowed by the k-selection rule (indirect-gap semiconductors), but the transition is allowed by either absorption of phonons or their emission. On the other hand, strong PL is observed in a-Si:H. In amorphous semiconductors, the k-selection rule is relaxed and, furthermore, the electrons and holes relax to localized states in the gap before radiative recombination. Thus, PL occurs by transitions between localized states.

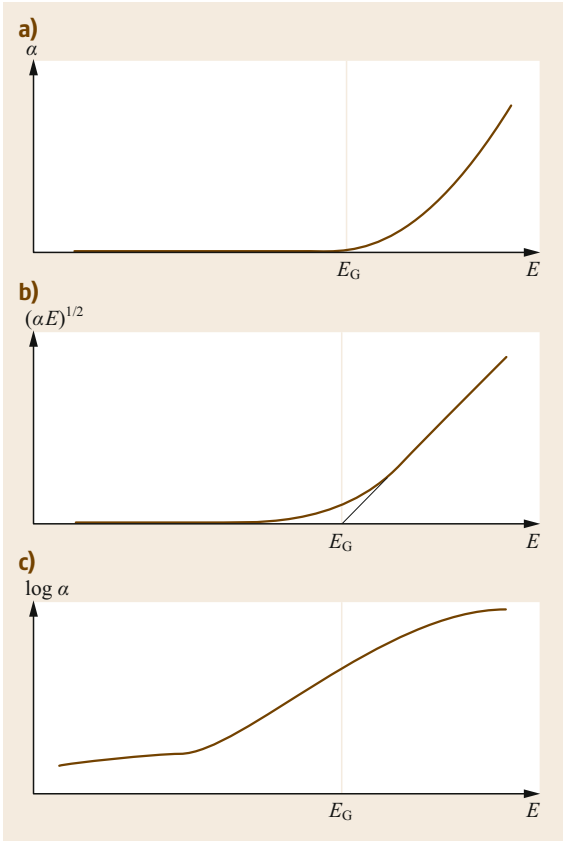


Fig. 24.11a–c Schematic illustration of the absorption spectra of amorphous semiconductors. **(a)** The absorption coefficient α plotted as a function of photon energy E. **(b)** $(\alpha E)^{1/2}$ versus E (Tauc's plot). **(c)** $\log\alpha$ versus E. E_G is the optical gap determined from **(b)**

The lifetime measurement is important for the identification of the origin of the PL since the PL spectra in most amorphous semiconductors are featureless and do not provide much information. The probability of radiative transition between localized electrons and holes depends exponentially on their separation, R. The lifetime is estimated by $\tau = \tau_0 \exp(2R/R_0)$ where R_0 denotes the radius of the most extended wavefunction [24.36]. When the Coulomb interaction is strong enough, excitons are formed. Exciton formation is also suggested in some amorphous semiconductors. The total spin of excitons S is either 0 or 1. Singlet excitons of $S = 0$ have a lifetime in the nanosecond region, while triplet excitons with $S = 1$ have a lifetime much longer than that of singlet excitons, e.g., $\tau = 1\,\mathrm{ms}$, since the transition is forbidden by the spin selection rule.

PL measurements combined with magnetic resonance are a powerful means to study recombination processes [24.14, 15]. This is called optically detected

Table 24.1 Optical gap of a-Si:H prepared at various deposition temperatures

T_s (°C)	[H] (at.%)	E_g (eV)	References
300	7	1.70	[24.37, 38]
200	18	1.75	[24.37, 38]
120	28	1.9	[24.37, 38]
75	33	2.0	[24.38, 39]

T_s: Deposition temperature, [H]: Hydrogen content

magnetic resonance (ODMR) measurements. PL from triplet excitons has been suggested from ODMR measurements, as described in Sects. 24.3.2 and 24.3.3.

The photoconductivity measurements provide us with useful information about the processes of carrier transport and recombination. Their results on a-Si:H and chalcogenide glasses are briefly reviewed in Sects. 24.3.2 and 24.3.3, respectively.

24.3.2 a-Si:H and Related Materials

The band gap of a-Si:H determined from Tauc's plot is 1.7−1.9 eV, depending on the preparation condition, particularly the deposition temperature T_s, i. e., hydrogen content [H] [24.37–40] as shown in Table 24.1. The absorption coefficient of visible light for a-Si:H is of the order 10^5−10^6 cm^{-1}, which is large compared to that of crystalline silicon. A thickness of 1 μm is enough to absorb visible light. Thus a-Si:H is suitable for application to solar cells especially when a thin film is desired.

The optical gap E_g and the Urbach tail width E_U of a-Si:H, a-Ge:H and Si-based alloys are shown in Table 24.2. Some a-Si:H-based alloys such as a-$Si_{1-x}N_x$:H and a-$Si_{1-x}C_x$:H have a band gap wider than that of a-Si:H depending on the composition x. In the preparation of these alloys, x can be varied continuously over a certain range. Thus, the band gap can be varied arbitrarily. It is possible to prepare multilayer films consisting of layers with different band gaps, such as a-Si:H/a-$Si_{1-x}N_x$:H. Quantum size effects in multilayer films will be described in Sect. 24.6.

It is difficult to obtain the absorption spectra below the band gap from transmittance measurements in a-Si:H, which is normally prepared as a thin film. Such low absorption in a-Si:H has been measured by PDS and CPM.

PL from a-Si:H was observed by *Engemann* and *Fischer* [24.49] for the first time. The quantum efficiency has been found to be of the order of unity [24.50], although it decreases with increasing density of dangling bonds [24.51]. PL spectra from a-Si:H films generally consist of two components. The first is observed as a peak at 1.3−1.4 eV with a full-width at half magnitude (FWHM) of 0.3 eV. The other component, which is called low-energy PL (defect PL), is observed at 0.8−0.9 eV.

Table 24.2 Band gap energies, Urbach tail widths and dark conductivities at room temperature for a-Si:H, a-Ge:H and related materials

Material	E_g (eV)	E_U (meV)	σ (300 K) (Ω^{-1} cm^{-1})	References
a-Si:H	1.75	48	10^{-11}	[24.41, 42]
a-Ge:H	1.05	50	10^{-4}	[24.41, 42]
a-$Si_{0.7}C_{0.3}$:H	2.28	183.4		[24.43]
a-$Si_{0.8}C_{0.2}$:H	2.2		10^{-15}	[24.44]
a-$Si_{0.4}N_{0.6}$:H	3.0	≈ 200	≈ 10^{-8}	[24.45–47]
a-$Si_{0.74}N_{0.26}$:H	≈ 2.0	≈ 100	1.4×10^{-8}	[24.45, 46, 48]

The origin of the main peak has not been fully understood. Electrons and holes in the tail states may give rise to this PL. *Dunstan* and *Boulitrop* [24.52] have shown that the PL spectra of a-Si:H can be understood without considering electron–phonon coupling. However, the electron–phonon coupling has still been discussed. *Morigaki* proposed a model on the basis of ODMR measurements [24.14] in which self-trapped holes and tail electrons are the origin of the PL at low temperatures [24.53], while tail holes and tail electrons participate in the PL at high temperatures.

In the case of radiative recombination of localized electron–hole pairs the lifetime depends exponentially on the separation of the electron and the hole, as described in Sect. 24.3.1. When the spatial distribution of electrons and holes is random, a broad lifetime distribution is predicted. This model describes some properties well, e.g., the generation-rate dependence of the lifetime. However, it has been pointed out that we have to consider the PL from specific electronic states such as excitonic PL to understand the lifetime distribution at low generation rate [24.54–56]. PL due to triplet excitons in a-Si:H has been suggested by optically detected magnetic resonance measurements [24.57, 58]. The PL from singlet excitons, which is expected to have lifetime of about 10 ns, has also been reported [24.59–61].

The low-energy PL is emission from deep gap states created by defects. However the origin of low-energy PL is not neutral dangling bonds (T_3^0) of Si since they act as nonradiative centers, as has been suggested from optically detected magnetic resonance [24.62]. *Yamaguchi* et al. [24.17, 18] proposed a model in which the origin of the low-energy PL is $T_3^+ - N_2^-$ pairs.

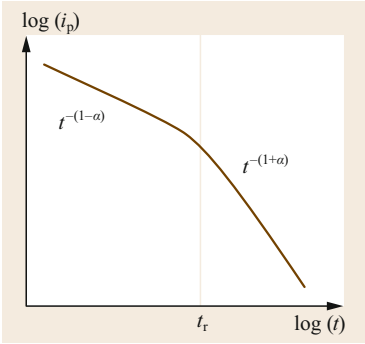


Fig. 24.12 Schematic illustration of the transient photocurrent curve for the dispersive transport

The drift mobility of carriers, i. e., electrons and holes, in a-Si:H has been measured by the method of time of flight (TOF), using blocking electrodes [24.63]. In the measurement, carriers are created by an optical pulse near one of the electrodes and run as a sheet-like shape against another electrode. The transient photocurrent i_p associated with a pulsed optical excitation exhibits a dispersive behavior, as shown in Fig. 24.12 [24.63, 64], which is given by

$$i_p \propto t^{-(1-\alpha)} \quad \text{for} \quad t < t_r \tag{24.3}$$

and

$$i_p \propto t^{-(1+\alpha)} \quad \text{for} \quad t > t_r \,, \tag{24.4}$$

where t_r designates the transit time, as shown in Fig. 24.12, given by

$$t_r = \frac{L}{\mu_d F} \,, \tag{24.5}$$

in which L, μ_d and F are the separation of the two electrodes, the drift mobility and the magnitude of the electric field, respectively. In dispersive conduction, the drift mobility depends on the thickness of the samples. This is due to dispersive carrier processes, which are caused by dispersion of the flight time. Band conduction with multiple trapping of carriers also exhibits dispersive conduction, as was observed in a-Si:H and chalcogenide glasses. For undoped a-Si:H, the drift mobility was on the order of 1 cm^2/V at room temperature. $\alpha = 0.51$ was obtained as 160 K [24.11]. This transient behavior depends on temperature, i. e., it becomes more dispersive at lower temperatures. α is related to the exponential tail width (normally equal to E_U) $E_c \equiv k_B T_c$ as follows

$$\alpha = \frac{T}{T_c} \,. \tag{24.6}$$

From the value of α, the tail width of the conduction band has been obtained to be 25 meV. For hole transport, dispersive behavior has also been obtained and a drift mobility of the order of $10^{-2}\,cm^2/V$ has been obtained at room temperature.

For the steady-state photoconductivity of undoped a-Si:H, it has been generally accepted that photoconduction occurs through electrons in the conduction band at temperatures above ≈ 60 K and through hopping of tail electrons at temperatures below ≈ 60 K. The detailed processes of trapping and recombination of carriers involved in photoconduction have been discussed in many literatures (see, e.g., *Morigaki* [24.2]; *Singh* and *Shimakawa* [24.3]). In relation to this issue, spin-dependent photoconductivity measurements are very useful, and have therefore been extensively performed on a-Si:H [24.68–70].

24.3.3 Chalcogenide Glasses

As mentioned in Sect. 24.3.1, the absorption coefficient for the interband transition in many chalcogenide glasses is well described by Tauc's relationship (24.1). The optical gap E_g is shown in Table 24.3 for various chalcogenide glasses. The effect of light on optical absorption in chalcogenide glasses is known as

Table 24.3 Band gap energies, drift mobilities, and dark conductivities at room temperature for chalcogenide glasses [24.2]

Material	E_g [24.66] (eV)	μ_d ($cm^2/(V\,s)$) [24.1, 65] n-type	μ_d ($cm^2/(V\,s)$) [24.1, 65] p-type	σ (300 K) [24.66] ($\Omega^{-1}\,cm^{-1}$)
Se	2.05	$3-6\times10^{-3}$	$1-2\times10^{-1}$	10^{-16}
As_2S_3	2.32			10^{-17}
As_2Se_3	1.76		10^{-3}	10^{-12}
As_2Te_3	0.83		10^{-2}	10^{-4}
GeS_2	3.07			10^{-14}
$GeSe_2$	2.18	1.4×10^{-1}	4×10^{-2}	10^{-11}
Sb_2Se_3	0.70			
Cd−In−S [24.67]	2.2	30[a]		9×10^{-1}

[a] Hall mobility

photo-darkening. This effect will be briefly reviewed in Sect. 24.5.

PL has been observed in a-Se, arsenic chalcogenide glasses and germanium chalcogenide glasses. PL in chalcogenide glasses is known to have a large Stokes shift. The peak energy in the PL spectra is approximately half of the band gap. This Stokes shift has been attributed to strong electron–phonon coupling [24.71]. The electronic states responsible for the PL have not been well understood. There has been some experimental evidence for PL from self-trapped excitons. The observation of triplet excitons has been reported in ODMR measurements for chalcogenide glasses such as As_2S_3 and As_2Se_3 [24.72–74].

Dispersive transport was observed in the transient photocurrent for the first time for chalcogenide glasses such as a-As_2Se_3, in which the hole transport dominates over the electron transport [24.63]. The drift mobility of holes ranges between 10^{-3}–$10^{-5}\,\mathrm{cm^2/V}$, depending on the temperature and electric field. The zero-field hole mobility at room temperature is about $5\times10^{-5}\,\mathrm{cm^2/V}$. The steady-state photoconduction in a-As_2Se_3 has been considered to be governed by charged structural defects such as VAPs.

24.4 Electrical Properties

24.4.1 General Aspects

Electrical Conductivity

Electrical conduction in amorphous semiconductors consists of band conduction and hopping conduction. Band conduction in undoped amorphous semiconductors is characterized by

$$\sigma = \sigma_0 \exp\left(-\frac{E_\mathrm{a}}{k_\mathrm{B}T}\right), \tag{24.7}$$

where σ and σ_0 are the electrical conductivity and a prefactor, respectively, and E_a, k_B and T are the activation energy, the Boltzmann constant and the temperature, respectively. E_a is given by either $E_\mathrm{c}-E_\mathrm{F}$ or $E_\mathrm{F}-E_\mathrm{V}$, depending on whether electrons or holes are considered, where E_c, E_V and E_F are the mobility edges of the conduction band and the valence band, and the Fermi energy, respectively. Hopping conduction in amorphous semiconductors consists of nearest-neighbor hopping and variable-range hopping [24.1]. Nearest-neighbor hopping is well known in crystalline semiconductors, in which electrons (holes) hop to nearest-neighbor sites by emitting or absorbing phonons. Variable-range hopping is particularly associated with tail states, in which electrons (holes) in tail states hop to the most probable sites. This type of hopping conductivity σ_p is characterized by the following temperature variation

$$\sigma_\mathrm{p} = \sigma_{\mathrm{p}0} \exp\left(-\frac{B}{T^{1/4}}\right). \tag{24.8}$$

The Hall Effect

The Hall effect is used for the determination of carrier density and its sign in crystalline semiconductors. For amorphous semiconductors, however, the sign of the Hall coefficient does not always coincide with that of the carriers [24.75]. Such a sign anomaly has been observed in a-Si:H, as will be shown in Sect. 24.4.2.

Thermoelectric Power

The thermoelectric power S associated with band conduction of electrons is given by

$$S = -\frac{k_\mathrm{B}}{e}\left(\frac{E_\mathrm{c}-E_\mathrm{F}}{k_\mathrm{B}T}+A\right), \tag{24.9}$$

where A is a quantity depending on the energy dependence of the relaxation times associated with electrical conduction. For hole conduction, S is given by (24.9) except that $E_\mathrm{c}-E_\mathrm{F}$ is replaced by $E_\mathrm{F}-E_\mathrm{V}$ and that the sign of S is positive. The sign of S coincides with that of the carriers. The thermoelectric power S (24.9) is related to the electrical conductivity σ (24.7) as follows [24.76]

$$\ln\sigma + \left|\frac{e}{k_\mathrm{B}}S\right| = \ln\sigma_0 + A \equiv Q\,. \tag{24.10}$$

The temperature-dependent quantity Q is defined as above. The activation energies of σ and S, E_σ and E_S, are equal to each other. Then, Q is defined by Q_0 as follows

$$Q = \ln\sigma_0 + A \equiv Q_0\,. \tag{24.11}$$

However, if E_σ is not equal to E_S, Q is generally expressed by

$$Q = Q_0 - \frac{E_Q}{k_\mathrm{B}T} \tag{24.12}$$

$$E_Q = E_\sigma - E_S\,. \tag{24.13}$$

It has been generally observed that E_σ is greater than E_S, i. e., $E_\sigma > E_S$. This has been accounted for in terms of long-range fluctuation of the band edge. The electrical conduction is due to those carriers that are thermally

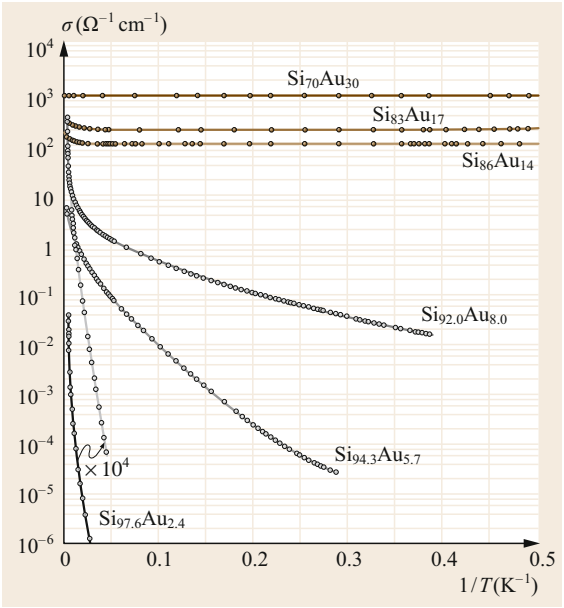


Fig. 24.13 Temperature dependence of the electrical conductivity for an a-Si_xAu_{1-x} film. (After [24.77])

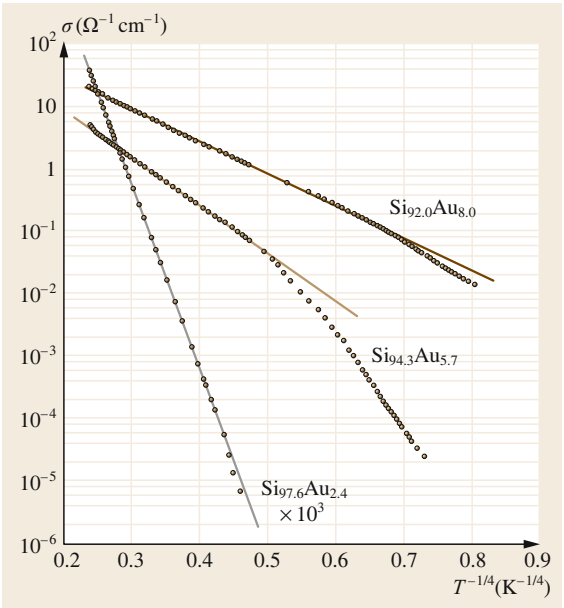


Fig. 24.14 Electrical conductivity versus $T^{-1/4}$ for a-Si_xAu_{1-x} films. (After [24.77])

excited into the valley of the band-edge fluctuation and cross over its barrier, while the thermoelectric power, i. e., transport of phonon energy by carriers, is governed by carriers in the valleys. Thus, a relationship of $E_\sigma > E_S$ can be accounted for in terms of long-range fluctuation of the band edge [24.76].

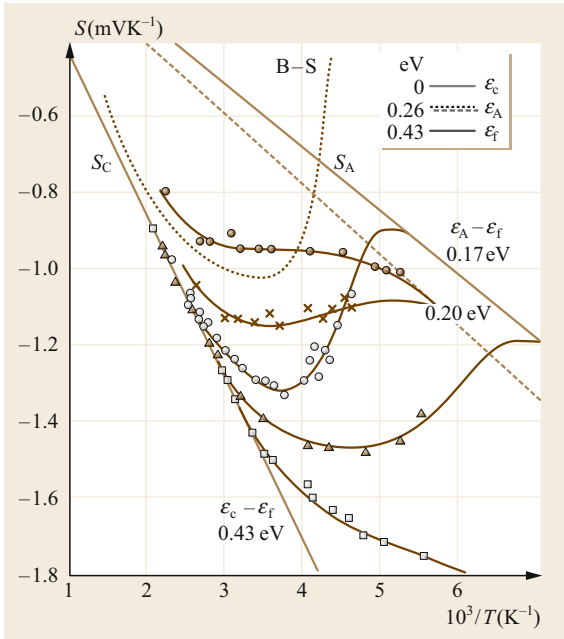


Fig. 24.15 Thermoelectric power versus T^{-1} for various a-Ge:H films. The *solid lines* are the calculated results. The *insert* shows the model on which the calculations are based. Carriers in states just above E_A move in extended states, those at E_A move by hopping. The *dotted curve* marked B–S was obtained by *Beyer* and *Stuke* [24.78] for a slowly evaporated sample annealed at 310 °C [24.79]

Drift Mobility

The drift mobility of carriers is measured by the TOF method, as mentioned in Sect. 24.3.2. The drift mobility μ_d is estimated from (24.5).

In the following, the electrical properties of a-Si:H and related materials and chalcogenide glasses are described.

24.4.2 a-Si:H and Related Materials

A typical example of band conduction and variable-range hopping conduction is shown in Fig. 24.13, which is the temperature dependence of the electrical conductivity for a-Si_xAu_{1-x} films prepared by vapor evaporation [24.77]. Band conduction occurs at high temperatures and variable-range hopping conduction at low temperatures. The $T^{-1/4}$ law of temperature variation of electrical conductivity is clearly seen in Fig. 24.14. The values of the dark conductivities of a-Si:H, a-Ge:H and related materials at room temperature are shown in Table 24.2. Figure 24.15 shows the temperature dependencies of S for a-Ge:H, in which two slopes of the curve of S versus T^{-1} are seen with 0.43 eV in the high-temperature range and 0.17 eV in

the low-temperature range [24.79]. This crossover of two slopes is accounted for in terms of two different mechanisms of electrical conduction, i. e., band conduction at high temperatures and hopping conduction in the band-tail states at low temperatures [24.80]. In amorphous semiconductors, doping control is generally difficult, because the constituent atoms obey the so called $8-N$ rule to satisfy the coordination of covalent bonds with their neighboring atoms. In a-Si:H, however, n-type doping and p-type doping were performed by using phosphorus (group V element) and boron (group III element), respectively [24.81].

The Hall coefficient of a-Si:H has been measured [24.75]. In crystalline semiconductors, when the carriers are electrons, its sign is negative, and when the carriers are holes, its sign is positive. The sign of the Hall coefficient is generally negative irrespective of that of the carriers in amorphous semiconductors. This has been accounted for in terms of the random phase model [24.82]. However, for the Hall coefficient of a-Si:H, this is not the case, i. e., when the carriers are electrons, the sign is positive, and when the carriers are holes, the sign is negative. Such an anomalous Hall coefficient in a-Si:H is called double reversal of the Hall coefficient and has been accounted for, taking into account antibonding orbitals for electrons and bonding orbitals for holes [24.83].

24.4.3 Chalcogenide Glasses

The values of the dark conductivities of chalcogenide glasses at room temperature are shown in Table 24.3 along with those of μ_d at room temperature. In chalcogenide glasses, the Fermi level is pinned near the midgap and this has been considered to be due to VAP defects with negative correlation energy. Thus, doping seems difficult, except for a few cases, e.g., the incorporation of Bi into a-Ge:S and a-Ge:Se increases the direct current (DC) conductivity by 6–7 orders of magnitude [24.84, 85]. This doping effect is due to chemical modification instead of conventional doping. Furthermore, Cd–In–S chalcogenide glasses exhibit $\sigma \approx 1\times10^{-2}\,\Omega^{-1}\,\mathrm{cm}^{-1}$ at room temperature [24.67].

24.5 Light-Induced Phenomena

Light-induced phenomena in amorphous semiconductors were first observed for chalcogenide glasses as photo-darkening, PL fatigue and photostructural changes associated with illumination whose energy equals to or exceeds the band-gap energy (see, e.g., *Shimakawa* et al. [24.86]; *Singh* and *Shimakawa*, [24.3]). Photo-darkening is an effect in which the absorption edge shifts towards lower photon energy under illumination, as shown in Fig. 24.16 [24.87]. PL fatique is the decrease of the PL intensity associated with illumination. Photostructural changes are observed as changes of the volume (generally expansion) of the sample with illumination. PL fatigue may be due to light-induced creation of charged structural defects acting as nonradiative recombination centers. For the photo-darkening and the photostructural change, it has been discussed whether they are independent of each other or not. Thus, the origins of these phenomena are still unclear.

Light-induced phenomena in a-Si:H were observed first in the dark conductivity and photoconductivity, i. e., drops in their values after prolonged illumination of band-gap light (the so-called *Staebler–Wronski* effect [24.88]). Subsequently, such phenomena have been observed in PL, optical absorption and ODMR etc. [24.2]. This effect also gives rise to degradation of the performance of amorphous-silicon solar cells, so that this has received great attention from the viewpoint of application. After prolonged illumination, dangling bonds were found to be created [24.89, 90],

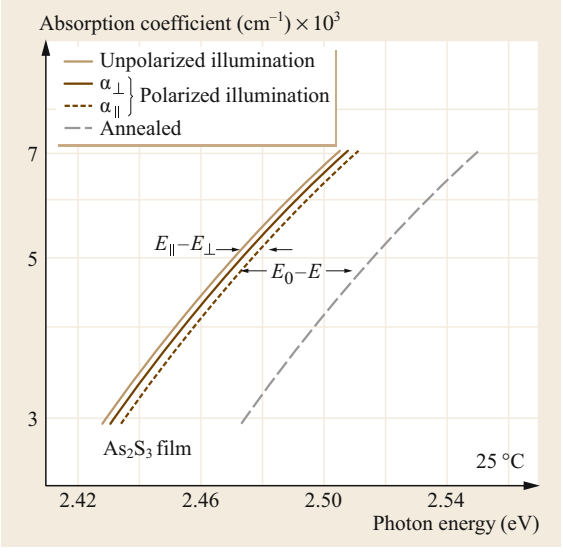


Fig. 24.16 Optical absorption edge of As_2S_3 films annealed or illuminated with unpolarized or linearly polarized light ($\alpha_\perp$: *solid line* and $\alpha_{||}$: *dashed line*). $\alpha_{||}$ and $\alpha_\perp$ refer to polarization along the plane of illumination and perpendicular to it, respectively. (After [24.87])

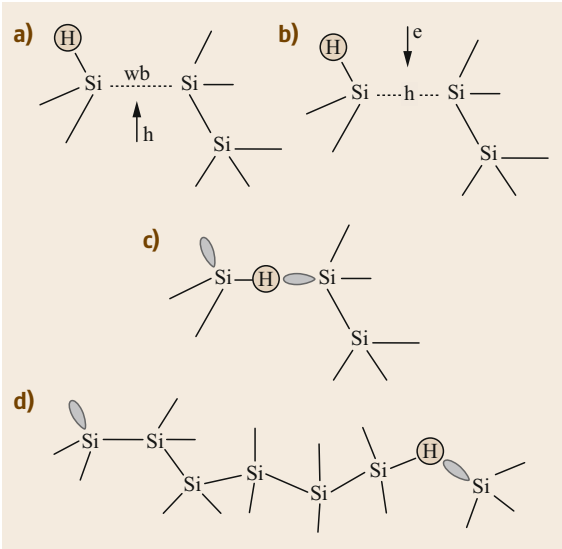


Fig. 24.17 **(a)** Self-trapping of a hole by a weak Si−Si bond adjacent to a Si−H bond. **(b)** Nonradiative recombination of a self-trapped hole with an electron. **(c)** A Si−H bond is switched toward the weak Si−Si bond and a dangling bond is left behind. **(d)** Formation of two separate dangling bonds after hydrogen movements and repeating of processes shown in **(a–c)**

so that light-induced creation of dangling bonds has been considered as the origin of these phenomena. Several models have been proposed for mechanisms for light-induced creation of dangling bonds [24.53, 91–96], but this issue is still controversial (see, e.g., *Morigaki* [24.2]; *Singh* and *Shimakawa* [24.3]). Very recently, the creation of a number of dangling bonds such as 1×10^{19} cm^{-3} after pulsed optical excitation has been observed in high-quality a-Si:H films [24.97, 98]. This result has been accounted for in terms of the authors' model [24.99], based on a combination of the following processes occurring during illumination. Self-trapping of holes in weak Si−Si bonds adjacent to Si−H bonds triggers these weak bonds to break using the phonon energy associated with nonradiative recombination of electrons with those self-trapped holes (Fig. 24.17a,b). After Si−H bond switching and hydrogen movement, two types of dangling bonds are created, i. e., a normal dangling bond and a dangling bond with hydrogen at a nearby site, i. e., so-called hydrogen-related dangling bonds (Fig. 24.17c,d). In the latter, hydrogen is dissociated from the Si−H bond as a result of nonradiative recombination at this dangling bond site. Dissociated hydrogen can terminate two types of dangling bonds (Fig. 24.18c,d) or can be inserted into a nearby weak Si−Si bond (Fig. 24.18a). The dependencies of the density of light-induced dangling bonds on illumination time and generation rate have been calculated using the rate equations governing these processes. The results have been compared with experimental results obtained under continuous illumination and pulsed illumination with good agreement. Light-induced structural changes, i. e., volume change [24.100] and local structural changes around a Si−H bond [24.101] have also been observed in a-Si:H and the origins for these changes have been discussed [24.102–104]. Further references are included in [24.105].

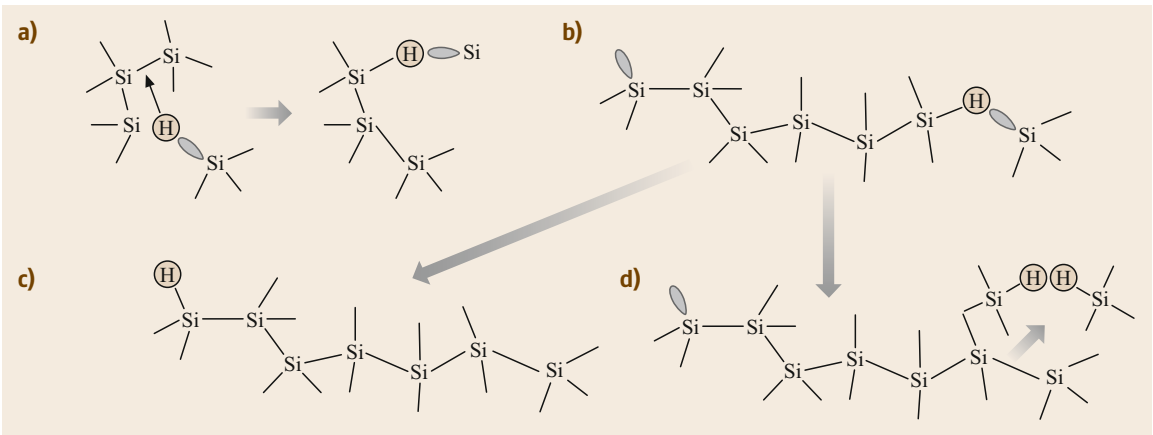


Fig. 24.18 **(a)** Dissociation of hydrogen from a hydrogen-related dangling bond and insertion of hydrogen into a nearby weak Si−Si bond (formation of a new hydrogen-related dangling bond). **(b)** Two separate dangling bonds, i. e., a normal dangling bond and a hydrogen-related dangling bond. **(c)** Dissociation of hydrogen from a hydrogen-related dangling bond and termination of a normal dangling bond by hydrogen. **(d)** Termination of a hydrogen-related dangling bond by hydrogen

24.6 Nanosized Amorphous Structure

Recent technologies have enabled us to control the properties of semiconductors by introducing artificial structures of nanometer size such as quantum wells, quantum wires and quantum dots. The quantum well is formed by preparing multilayers consisting of two semiconductors. Figure 24.19a illustrates the conduction and valence band edges in such multilayers consisting of two semiconducting materials with different band gaps, plotted as functions of the distance from the substrate z. The curves of the conduction and valence bands in Fig. 24.19a are considered to be the potentials for the electrons and holes, respectively. The motion of the electrons and holes is assumed to be described by effective mass equations which are similar to the Schrödinger equation. When the barriers of the potentials are high enough to prevent the carriers from moving to the adjacent well, the quantum levels (for motion along the z-axis) are as illustrated in Fig. 24.19a. The band gap of the multilayer E_g is equal to the separation of the quantum levels of the lowest energy in the conduction and valence bands, as shown in Fig. 24.19a. E_g increases with decreasing thickness of the well layer L_w; this is well known as the quantum-size effect. When the height of the barrier is infinite, $\Delta E = E_g - E_{gw}$ is proportional to L_w^{-2}, where E_{gw} denotes the band gap of the material of the well layer.

Multilayers of various amorphous semiconductors have been prepared. Observation of the quantum size effect has been reported in some amorphous semiconducting multilayers. For example, the shift of the optical gap of an a-Si:H/a-$Si_{1-x}N_x$:H multilayer from that of a-Si:H, ΔE, is proportional to L_w^{-2}, where L_w is the thickness of the a-Si:H layers (e.g., *Morigaki* [24.2]; see *Yamaguchi* and *Morigaki* [24.106]; and references therein). The result, $\Delta E \propto L_w^{-2}$, is consistent with quantum size effect in a square-well potential.

In the case of amorphous semiconductors, it is easy to introduce a potential well of arbitrary shape, because the optical gap of the amorphous semiconducting alloy, which depends on the composition, can be varied continuously within a certain range. The authors have prepared a new type of a-Si:H-based film called band-edge modulated (BM) a-$Si_{1-x}N_x$:H, in which the potentials for electrons and holes are sinusoidal functions of z [24.107, 108]. The band-edges for the BM films are illustrated in Fig. 24.19b. In this case, the quantum size effect is also expected, similarly to the case of multilayer films. However, the quantum levels and the size dependencies of the band gap are different from those in the case of multilayer films. In the case of BM, the potential is approximately parabolic at the bottom of the well. In this case the quantum levels have the same separation as that illustrated in Fig. 24.19b. We expect that the shift of the optical gap in BM films ΔE will be proportional to L^{-1}, where L denotes the modulation period. The band gap of BM films observed in experiments are in agreement with that expected from the above consideration [24.107]. Parabolic potentials in crystalline semiconductors have also been introduced (see *Gossard* et al. [24.109]). However it is difficult to obtain such structures that are small enough to observe significant quantum effects in crystalline semiconductors.

Recently, preparation of amorphous silicon quantum dots in silicon nitride has been reported [24.110]. Theoretical calculations for silicon nanostructures [24.111] have predicted that the oscillator strength in low symmetry is larger than that in high symmetry. A large quantum efficiency for PL is expected in the case of amorphous quantum dots. Thus, amorphous quantum wires and dots have potential applications in light-emitting devices with high efficiency.

Chalcogenide glasses multilayers have been prepared [24.112–114], whose optical properties such as optical absorption and PL, and electrical properties have been measured.

Acknowledgments. We wish to thank M. Ichihara, K. Suzuki and M. Yamaguchi, Institute for Solid State Physics, University of Tokyo, for providing us with their unpublished materials (Fig. 24.7a,b). Stimulating and helpful discussions were held with Prof. S. Kugler during the stay of one of us (K.M.) at the Budapest University of Technology and Economics, for which K.M. is grateful.

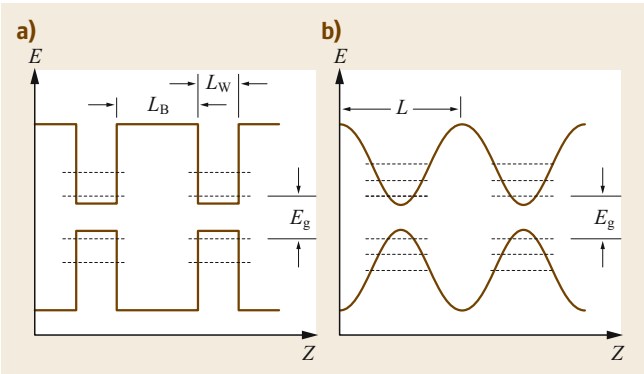


Fig. 24.19 The edges of the conduction and valence bands in **(a)** multilayer films and **(b)** band-edge modulated films, as functions of z, where z denotes distance from the substrates. The *dashed lines* indicate quantum levels. The band gap E_g, which increases due to the quantum size effect, is also shown

References

24.1 N.F. Mott, E.A. Davis: *Electronic Processes in Non-crystalline Materials*, 2nd edn. (Clarendon, Oxford 1979)
24.2 K. Morigaki: *Physics of Amorphous Semiconductors* (World Scientific, Imperial College Press, Singapore, London 1999)
24.3 J. Singh, K. Shimakawa: *Advances in Amorphous Semiconductors* (Taylor Francis, London 2003)
24.4 S. Kugler, P.R. Surján, G. Náray-Szabó: Phys. Rev. B **37**, 9069 (1988)
24.5 L. Ley, J. Reichardt, R.L. Johnson: Phys. Rev. Lett. **49**, 1664 (1982)
24.6 L. Brey, C. Tejedor, J.A. Verges: Phys. Rev. Lett. **52**, 1840 (1984)
24.7 L. Ley, S. Kawalczyk, R. Pollak, D.A. Shirley: Phys. Rev. Lett. **29**, 1088 (1972)
24.8 M. Kastner, D. Adler, H. Fritzsche: Phys. Rev. Lett. **37**, 1504 (1976)
24.9 K. Morigaki, M. Yamaguchi, I. Hirabayashi, R. Hayashi: *Disordered Semiconductors*, ed. by M.A. Kastner, G.A. Thomas, S.R. Ovshinsky (Plenum, New York 1987) p. 415
24.10 G.D. Cody: In: *Semiconductors and Semimetals*, Vol. 21, ed. by J.I. Pankove (Academic, Orlando 1984) p. 1 Part B
24.11 T. Tiedje: In: *Semiconductors and Semimetals*, Vol. 21, ed. by J.I. Pankove (Academic, Orlando 1984) p. 207 Part C
24.12 K. Winer, I. Hirabayashi, L. Ley: Phys. Rev. B **38**, 7680 (1988)
24.13 K. Winer, I. Hirabayashi, L. Ley: Phys. Rev. Lett. **60**, 2697 (1988)
24.14 K. Morigaki: In: *Semiconductors and Semimetals*, Vol. 21, ed. by J.I. Pankove (Academic, Orlando 1984) p. 155 Part C
24.15 K. Morigaki, M. Kondo: Solid State Phenom. **44-46**, 731 (1995)
24.16 K. Morigaki, H. Hikita, M. Kondo: J. Non-Cryst. Solids **190**, 38 (1995)
24.17 M. Yamaguchi, K. Morigaki, S. Nitta: J. Phys. Soc. Jpn. **58**, 3828 (1989)
24.18 M. Yamaguchi, K. Morigaki, S. Nitta: J. Phys. Soc. Jpn. **60**, 1769 (1991)
24.19 S.C. Moss, J.F. Graczyk: Proc. 10th Int. Conf. Phys. Semicond., ed. by J.C. Hensel, F. Stern (US AEC Div. Tech. Inform., Springfield 1970) p. 658
24.20 S. Kugler, G. Molnár, G. Petö, E. Zsoldos, L. Rosta, A. Menelle, R. Bellissent: Phys. Rev. B **40**, 8030 (1989)
24.21 D.E. Polk: J. Non-Cryst. Solids **5**, 365 (1971)
24.22 F. Wooten, K. Winer, D. Weaire: Phys. Rev. Lett. **54**, 1392 (1985)
24.23 R. Car, M. Parrinello: Phys. Rev. Lett. **60**, 204 (1988)
24.24 S. Kugler, L. Pusztai, L. Rosta, P. Chieux, R. Bellissent: Phys. Rev. B **48**, 7685 (1993)
24.25 L. Pusztai: J. Non-Cryst. Solids **227-230**, 88 (1998)
24.26 S. Kugler, K. Kohary, K. Kádas, L. Pusztai: Solid State Commun. **127**, 305 (2003)
24.27 J.A. Reimer: J. Phys. (Paris) **42**, C4–715 (1981)
24.28 K.K. Gleason, M.A. Petrich, J.A. Reimer: Phys. Rev. B **36**, 3259 (1987)
24.29 G. Lucovsky, R.J. Nemanich, J.C. Knights: Phys. Rev. B **19**, 2064 (1979)
24.30 R. Bellissent, A. Menelle, W.S. Howells, A.C. Wright, T.M. Brunier, R.N. Sinclair, F. Jansen: Physica B **156/157**, 217 (1989)
24.31 A. Menelle: Thése Doctorat Ser. (Université Piere et Marie Curie, Paris 1987)
24.32 J.H. Zhou, K. Ikuta, T. Yasuda, T. Umeda, S. Yamasaki, K. Tanaka: J. Non-Cryst. Solids **227–230**, 857 (1998)
24.33 T. Uchino: Kotai Butsuri (Solid State Physics) **37**, 965 (2002)
24.34 G. Lucovsky, F.L. Galeener, R.H. Geils, R.C. Keezer: Proc. Int. Conf. Amorph. Liq. Semicond., ed. by W.E. Spear (University of Edinburgh, Edinburgh 1977) p. 127
24.35 J. Tauc: In: *Amorphous and Liquid Semiconductors*, ed. by J. Tauc (Plenum, New York 1974) p. 159
24.36 C. Tsang, R.A. Street: Phys. Rev. B **19**, 3027 (1979)
24.37 K. Morigaki, Y. Sano, I. Hirabayashi: J. Phys. Soc. Jpn. **51**, 147 (1982)
24.38 K. Morigaki, Y. Sano, I. Hirabayashi: In: *Amorphous Semiconductor Technologies and Devices–1983*, ed. by Y. Hamakawa (Ohomsha, Tokio 1983), Chap. 3.2
24.39 I. Hirabayashi, K. Morigaki, M. Yoshida: Sol. Ener. Mat. **8**, 153 (1982)
24.40 M. Yamaguchi, K. Morigaki: Philos. Mag. B **79**, 387 (1999)
24.41 G.H. Bauer: Solid State Phenom. **44–46**, 365 (1995)
24.42 W. Paul: In: *Amorphous Silicon and Related Materials*, ed. by H. Fritzsche (World Scientific, Singapore 1989) p. 63
24.43 Y. Tawada: In: *Amorphous Semiconductors – Technologies and Devices*, ed. by Y. Hamakawa (Ohmusha, North Holland, Tokyo, Amsterdam 1983), Chap. 4.2
24.44 F. Demichelis, C.F. Pirri: Solid State Phenom. **44-46**, 385 (1995)
24.45 K. Maeda, I. Umezu: J. Appl. Phys. **70**, 2745 (1991)
24.46 M. Yamaguchi, K. Morigaki: Philos. Mag. B **79**, 387 (1999)
24.47 B. Dunnett, D.I. Jones, A.D. Stewart: Philos. Mag. B **53**, 159 (1986)
24.48 M. Hirose: Jpn. J. Appl. Phys. **21**(suppl. 21-1), 297 (1981)
24.49 D. Engemann, R. Fischer: In: *Amorphous and Liquid Semiconductors*, ed. by J. Stuke, W. Brenig (Taylor Francis, London 1974) p. 947
24.50 D. Engemann, R. Fischer: Proc. 12th Int. Conf. Phys. Semicond., ed. by M.H. Pilkuhn (B.G. Teubner, Stuttgart 1974) p. 1042
24.51 R.A. Street, J.C. Knights, D.K. Biegelsen: Phys. Rev. B **18**, 1880 (1978)
24.52 D.J. Dunstan, F. Boulitrop: Phys. Rev. B **30**, 5945 (1984)
24.53 K. Morigaki: J. Non-Cryst. Solids **141**, 166 (1992)

24.54 R. Stachowitz, M. Schubert, W. Fuhs: J. Non-Cryst. Solids **227–230**, 190 (1998)
24.55 C. Ogihara: J. Non-Cryst. Solids **227–230**, 517 (1998)
24.56 T. Aoki, T. Shimizu, S. Komedoori, S. Kobayashi, K. Shimakawa: J. Non-Cryst. Solids **338–340**, 456 (2004)
24.57 M. Yoshida, M. Yamaguchi, K. Morigaki: J. Non-Cryst. Solids **114**, 319 (1989)
24.58 M. Yoshida, K. Morigaki: J. Phys. Soc. Jpn. **58**, 3371 (1989)
24.59 B.A. Wilson, P. Hu, T.M. Jedju, J.P. Harbison: Phys. Rev. B **28**, 5901 (1983)
24.60 C. Ogihara, H. Takemura, H. Yoshida, K. Morigaki: J. Non-Cryst. Solids **266–269**, 574 (2000)
24.61 H. Takemura, C. Ogihara, K. Morigaki: J. Phys. Soc. Jpn. **71**, 625 (2002)
24.62 M. Yoshida, K. Morigaki: J. Non-Cryst. Solids **59/60**, 357 (1983)
24.63 G. Pfister, H. Scher: Adv. Phys. **27**, 747 (1978)
24.64 H. Scher, E.W. Montroll: Phys. Rev. B **12**, 2455 (1975)
24.65 A. Feltz: *Amorphous Inorganic Materials and Glasses* (VCH, Weinheim 1993)
24.66 S.R. Elliott: In: *Material Science and Technology*, Vol. 9, ed. by R.W. Cahn et al. (VCH, Weinheim 1991) p. 376
24.67 H. Hosono, H. Maeda, Y. Kameshima, H. Kawazoe: J. Non-Cryst. Solids **227–230**, 804 (1998)
24.68 I. Solomon: In: *Amorphous Semiconductors*, ed. by M.H. Brodsky (Springer, Berlin, Heidelberg, New York 1979) p. 189
24.69 K. Lips, C. Lerner, W. Fuhs: J. Non-Cryst. Solids **198–200**, 267 (1996)
24.70 M. Stutzmann, M.S. Brandt, M.W. Bayerl: J. Non-Cryst. Solids **266–269**, 1 (2000)
24.71 R.A. Street: Adv. Phys. **25**, 397 (1976)
24.72 B.C. Cavenett: J. Non-Cryst. Solids **59/60**, 125 (1983)
24.73 J. Ristein, P.C. Taylor, W.D. Ohlsen, G. Weiser: Phys. Rev. B **42**, 11845 (1990)
24.74 D. Mao, W.D. Ohlsen, P.C. Taylor: Phys. Rev. B **48**, 4428 (1993)
24.75 P.G. LeComber, D.I. Jones, W.E. Spear: Philos. Mag. **35**, 1173 (1977)
24.76 H. Overhof, W. Beyer: Philos. Mag. B **44**, 317 (1983)
24.77 N. Kishimoto, K. Morigaki: J. Phys. Soc. Jpn. **46**, 846 (1979)
24.78 W. Beyer, J. Stuke: Proc. Int. Conf. Amorph. Liq. Semicond., 1973, ed. by J. Stuke (Taylor Francis, London 1974) p. 251
24.79 D.I. Jones, W.E. Spear, P.G. LeComber: J. Non-Cryst. Solids **20**, 259 (1976)
24.80 N.F. Mott: J. Phys. C **13**, 5433 (1980)
24.81 W.E. Spear, P.G. LeComber: Philos. Mag. **33**, 935 (1976)
24.82 L. Friedman: J. Non-Cryst. Solids **6**, 329 (1971)
24.83 D. Emin: Philos. Mag. **35**, 1189 (1977)
24.84 N. Tohge, T. Minami, Y. Yamamoto, M. Tanaka: J. Appl. Phys. **51**, 1048 (1980)
24.85 L. Tichy, H. Ticha, A. Triska, P. Nagels: Solid State Commun. **53**, 399 (1985)
24.86 K. Shimakawa, A. Kolobov, S.R. Elliott: Adv. Phys. **44**, 475 (1995)
24.87 K. Kimura, K. Murayama, T. Ninomiya: J. Non-Cryst. Solids **77/78**, 1203 (1985)
24.88 D.L. Staebler, C.R. Wronski: Appl. Phys. Lett. **31**, 292 (1977)
24.89 I. Hirabayashi, K. Morigaki, S. Nitta: Jpn. J. Appl. Phys. **19**, L357 (1980)
24.90 H. Dersch, J. Stuke, J. Beichler: Appl. Phys. Lett. **38**, 456 (1981)
24.91 M. Stutzmann, W.B. Jackson, C.C. Tsai: Phys. Rev. B **32**, 23 (1985)
24.92 C. Godet, P. Roca i Cabarrocas: J. Appl. Phys. **80**, 97 (1996)
24.93 H.M. Branz: Phys. Rev. B **59**, 5498 (1999)
24.94 K. Morigaki, H. Hikita: Solid State Commun. **114**, 69 (2000)
24.95 K. Morigaki, H. Hikita: J. Non-Cryst. Solids **266–269**, 410 (2000)
24.96 K. Morigaki, H. Hikita: Proc. Int. Conf. Phys. Semicond., ed. by T. Ando, N. Miura (Springer, Berlin, Heidelberg, New York 2000) p. 1485
24.97 C. Ogihara, H. Takemura, T. Yoshimura, K. Morigaki: J. Non-Cryst. Solids **299–302**, 637 (2002)
24.98 K. Morigaki, H. Hikita, H. Takemura, T. Yoshimura, C. Ogihara: Philos. Mag. Lett. **83**, 341 (2003)
24.99 K. Morigaki, H. Hikita: J. Non-Cryst. Solids **299–302**, 455 (2002)
24.100 T. Gotoh, S. Nonomura, M. Nishio, S. Nitta, M. Kondo, A. Matsuda: Appl. Phys. Lett. **72**, 2978 (1998)
24.101 Y. Zhao, D. Zhang, G. Kong, G. Pan, X. Liao: Phys. Rev. Lett. **74**, 558 (1995)
24.102 H. Fritzsche: Solid State Commun. **94**, 953 (1995)
24.103 R. Biswas, Y.P. Li: Phys. Rev. Lett. **82**, 2512 (1999)
24.104 K. Morigaki: Res. Bull. Hiroshima Inst. Tech. **35**, 47 (2001)
24.105 K. Morigaki, H. Hikita, C. Ogihara: *Light-Induced Defects in Semiconductors* (Pan Stanford, Singapore 2014)
24.106 M. Yamaguchi, K. Morigaki: Phys. Rev. B **55**, 2368 (1997)
24.107 C. Ogihara, H. Ohta, M. Yamaguchi, K. Morigaki: Philos. Mag. B **62**, 261 (1990)
24.108 M. Yamaguchi, C. Ogihara, K. Morigaki: Mat. Sci. Eng. B **97**, 135 (2003)
24.109 A.C. Gossard, M. Sundaram, P.F. Hopkins: In: *Semiconductors and Semimetals*, Vol. 40, ed. by A.C. Gossard (Academic, Boston, Tokio 1994), Chap. 2
24.110 N.-M. Park, C.-J. Choi, T.-Y. Seong, S.-J. Park: Phys. Rev. Lett. **86**, 1355 (2001)
24.111 J. Koga, K. Nishio, T. Yamaguchi, F. Yonezawa: J. Phys. Soc. Jpn. **70**, 3143 (2001)
24.112 R. Ionov, D. Nesheva, D. Arsova: J. Non-Cryst. Solids **137/138**, 1151 (1991)
24.113 R. Ionov: Ph.D. Thesis (Technical Univ., Sofia 1993)
24.114 H. Hamanaka, S. Konagai, K. Murayama, M. Yamaguchi, K. Morigaki: J. Non-Cryst. Solids **198–200**, 808 (1996)

25. Amorphous and Microcrystalline Silicon

Akihisa Matsuda

Processes used to grow hydrogenated amorphous silicon (a-Si:H) and microcrystalline silicon (μc-Si:H) from SiH_4 and H_2/SiH_4 glow discharge plasmas are reviewed. Differences and similarities between growth reactions of a-Si:H and μc-Si:H in a plasma and on a film-growing surface are discussed, and the process of nucleus formation followed by epitaxial-like crystal growth is explained as being unique to μc-Si:H. The application of a reaction used to determine the dangling-bond defect density in the resulting a-Si:H and μc-Si:H films is emphasized, since it can provide clues about how to improve the optoelectronic properties of those materials for device applications, especially thin-film silicon-based solar cells. Material issues related to the realization of low-cost and high-efficiency solar cells are described, and finally recent progress in this area is reviewed.

25.1 **Reactions in SiH_4 and SiH_4/H_2 Plasmas** 574
25.2 **Film Growth on a Surface** 575
25.2.1 Growth of a-Si:H 575
25.2.2 Growth of μc-Si:H 576
25.3 **Defect Density Determination for a-Si:H and μc-Si:H** 580
25.3.1 Dangling Bond Defects 580
25.3.2 Dangling Bond Defect Density in μc-Si:H 582
25.4 **Device Applications** 582
25.5 **Recent Progress in Material Issues Related to Thin-Film Silicon Solar Cells** 583
25.5.1 Controlling Photoinduced Degradation in a-Si:H 583
25.5.2 High Growth Rates of Device-Grade μc-Si:H 584
25.6 **Summary** 586
References 586

Hydrogenated amorphous silicon (a-Si:H) and microcrystalline silicon (μc-Si:H) are recognized as being useful materials for constructing devices related to optoelectronics, such as solar cells, thin-film transistors, etc. [25.1, 2]. Several methods have been proposed for the preparation of device-grade a-Si:H and μc-Si:H. These include: reactive sputtering of a crystalline silicon target with $Ar+H_2$ plasma [25.3]; mercury-sensitized photochemical vapor deposition (CVD) utilizing a decomposition reaction of silane (SiH_4) molecules with photoexcited Hg (Hg*) [25.4]; a direct photo CVD method where high-energy photons from a Xe-resonance lamp or a low-pressure Hg lamp are used for the direct excitation of SiH_4 molecules to excited electronic states [25.5, 6]; a hot-wire CVD method for decomposing SiH_4 by means of catalytic reactions on a heated metal surface [25.7]; and a plasma-enhanced CVD method (PECVD). The PECVD method is the most widely used of these due to its ability to consistently prepare uniform, high-quality materials on a large-area substrate.

In this chapter, the PECVD method is highlighted, details regarding the processes used to grow a-Si:H and μc-Si:H from reactive plasmas are explained, and the determination reaction (which is used to obtain the dangling-bond defect density in the films: one of the most important structural properties that influences device performance) is interpreted in order to obtain clues about how to control the optoelectronic properties of those materials for device applications.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_25

25.1 Reactions in SiH_4 and SiH_4/H_2 Plasmas

The initial event required for the growth of a-Si:H and μc-Si:H is the decomposition of the source gas material in SiH_4 or SiH_4/H_2 glow discharge plasma. Figure 25.1 shows a schematic of the dissociation pathway of SiH_4 and H_2, during which the molecules are excited to higher electronic states due to inelastic collisions with high-energy electrons in the plasma [25.8]. As the electrons in the plasma usually have a wide variety of energies, from zero to several tens of electron volts (eV), ground-state electrons of source gas molecules are excited into their electronic excited states almost simultaneously due to inelastic collisions with energetic electrons. Excited electronic states of complicated molecules like SiH_4 are usually dissociating states, from which dissociation occurs spontaneously to SiH_3, SiH_2, SiH, Si, H_2 and H, as shown in Fig. 25.1, depending on the stereochemical structure of the excited state. Hydrogen molecules are also decomposed to atomic hydrogen. Excitation of ground-state electron to vacuum-state gives rise to ionization events, generating new electrons and ions, which maintains the plasma.

Reactive species produced in the plasma also experience secondary reactions, mostly with parent SiH_4 and H_2 molecules, as shown in Fig. 25.2, resulting in a steady state. Reaction-rate constants for each reaction are summarized in the literature [25.9]. Steady-state densities of reactive species are basically determined by the balance between their generation and annihilation rate. Therefore, highly reactive species such as SiH_2, SiH and Si (short-lifetime species) have much smaller densities than SiH_3 in the steady-state plasma, although the generation rates of those species are not very different from that of SiH_3 showing low reactivity along with SiH_4 and H_2 (long-lifetime species).

Steady-state densities of reactive species have been measured using various gas-phase diagnostic techniques [25.10–16], such as optical emission spectroscopy (OES) [25.10], laser-induced fluorescence (LIF) [25.12], infrared laser absorption spectroscopy (IRLAS) [25.13], and ultraviolet light absorption spectroscopy (UVLAS) [25.16]. Figure 25.3 shows the steady state number densities of the chemical species, including both emissive and ionic species, in the SiH_4 and SiH_4/H_2 plasmas used to prepare device-grade a-Si:H and μc-Si:H. It is clear from Fig. 25.3 that the SiH_3 radical is the dominant chemical species in the growth of both a-Si:H and μc-Si:H, although the density ratio of short-lifetime species to SiH_3 changes depending on the conditions used to generate the plasma. For instance, when high electric power is supplied to the plasma (high electron density) under low SiH_4 flow rate conditions, SiH_4 is rapidly dissociated and then depleted, giving rise to reduced probabilities of gas-phase reactions of short lifetime species with SiH_4 molecules. This leads to an increased contribution of short-lifetime species to the film growth, which causes a deterioration of the structural properties of the resulting films.

The steady-state density of atomic hydrogen (H) varies widely in the plasma, as shown in Fig. 25.3. This is mainly due to the change in the hydrogen dilution

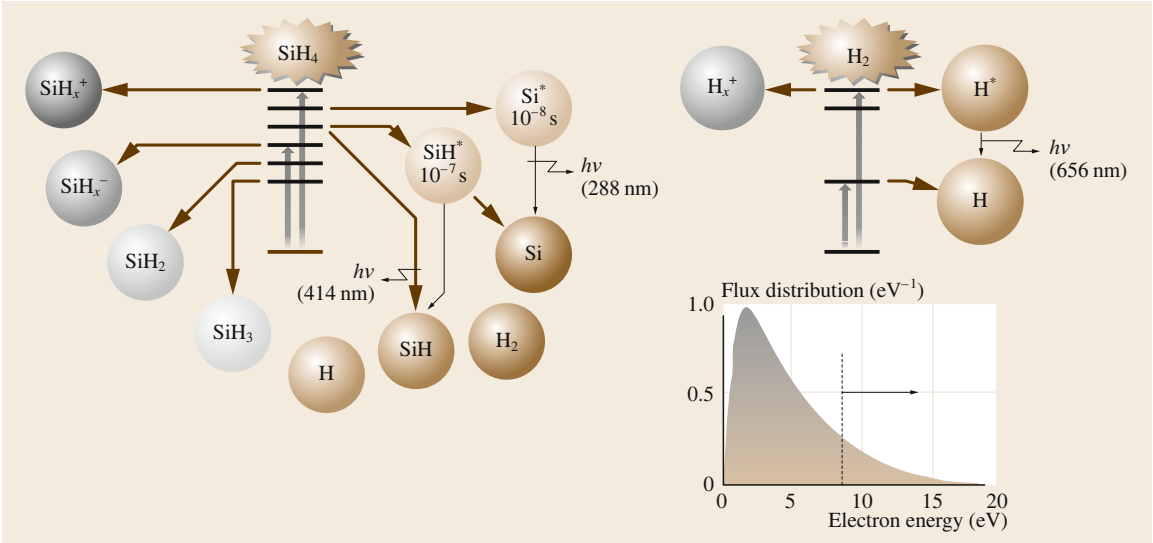


Fig. 25.1 Schematic showing the dissociation of SiH_4 and H_2 molecules to a variety of chemical species in the plasma via excited electronic states. The electron energy distribution function in the plasma is also shown

Ion exchanging	$SiH_x^+ + SiH_4$	$SiH_x + SiH_4^+$
Ion–Molecule	$SiH_x + SiH_4$	$SiH_3 + SiH_3$
Neutral–Molecule	$SiH + SiH_4$	Si_2H_5
Disproportionation	$Si + SiH_4$	$SiH_3 + SiH$
Insertion	$SiH_2 + SiH_4$	Si_2H_6
Recombination	$SiH_2 + H_2$	SiH_4
Abstraction	$SiH_3 + SiH_4$	$SiH_4 + SiH_3$
	$H + SiH_4$	$H_2 + SiH_3$
Ion–Radical, Radical–Radical	less probable	

Fig. 25.2 Representative secondary reactions of the chemical species produced in the plasma with SiH_4 and H_2 molecules. Their reaction rate constants are available in the literature

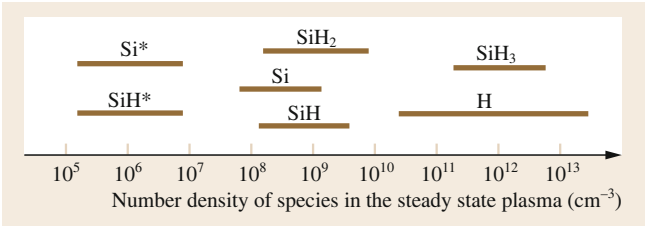


Fig. 25.3 Number densities of chemical species in realistic steady-state plasmas measured or predicted by various diagnostic techniques

ratio R (H_2/SiH_4) in the source gas material, i. e., the density of atomic hydrogen increases with increasing R. Noting the fact that more μc-Si:H is formed with increasing R at constant electron density in the plasma and constant substrate (surface) temperature, it is clear that atomic hydrogen plays an important role in μc-Si:H growth [25.17] although SiH_3 is the dominant film precursor for both a-Si:H and μc-Si:H growth [25.18].

25.2 Film Growth on a Surface

25.2.1 Growth of a-Si:H

Upon reaching a film-growing surface, some of the SiH_3 is reflected at the surface, while the remaining SiH_3 is adsorbed on the surface. Some adsorbed SiH_3 molecules interacts with the surface hydrogen at its landing site, forming SiH_4 and leaving dangling bonds on the surface. This is known as growth-site formation. Other adsorbed SiH_3 radicals begin to diffuse across the surface to find the site containing the dangling bonds, whereupon Si–Si bond formation (film growth) occurs, as shown schematically in Fig. 25.4. This surface reaction scheme for film growth has been proposed on the basis of two experimental results [25.18].

Figure 25.5 shows the general concept of the surface-reaction process. Some the flux of SiH_3 is reflected off the surface (the proportion of molecules reflected is given by the reflection probability). The remaining SiH_3 is adsorbed onto the surface and it changes its form as follows:

1. SiH_3 abstracts bonded H from the surface, forming SiH_4, or two of the SiH_3 radicals interact on the surface, forming Si_2H_6 (with a recombination probability γ);
2. Surface-diffusing SiH_3 sticks to the site containing the dangling bond, forming Si–Si bond (with a sticking probability s).

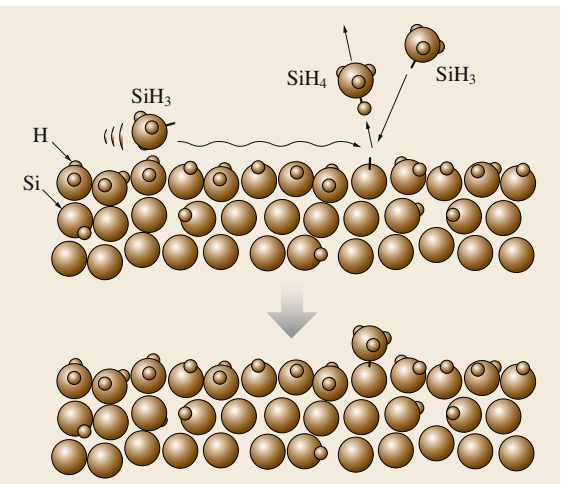


Fig. 25.4 Schematic of the surface growth of a-Si:H. The *large spheres* and *small spheres* represent Si and H, respectively

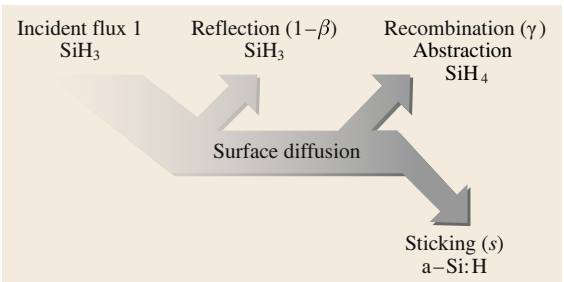


Fig. 25.5 General concepts behind the surface reactions of incoming SiH_3 radicals

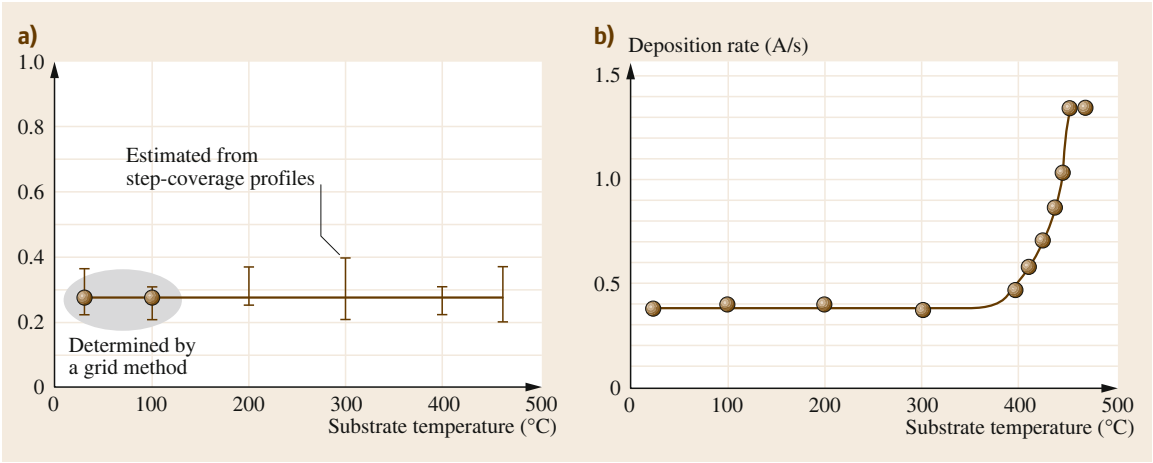


Fig. 25.6 Loss probability β **(a)** of SiH_3 radicals reaching the film-growing surface, as measured using the grid method and the step-coverage method, and the deposition rate (corresponding to the sticking probability s for SiH_3) **(b)** as a function of substrate temperature during the growth of a-Si:H

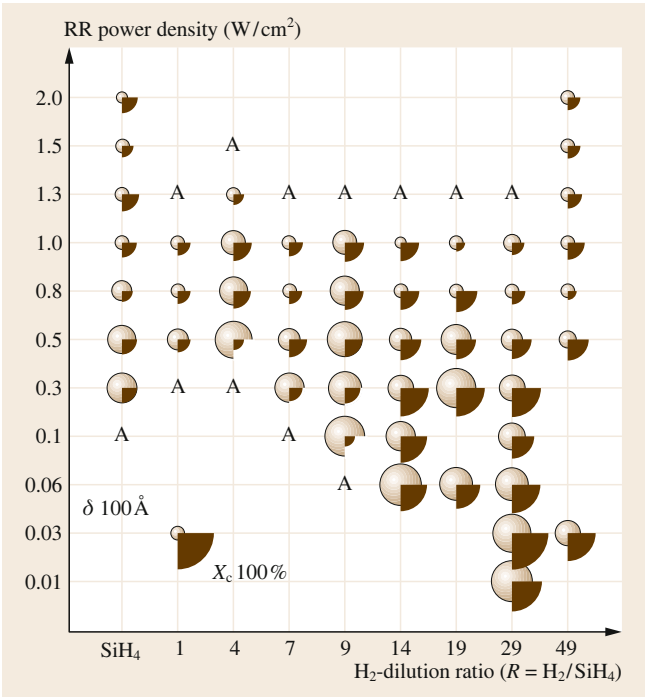


Fig. 25.7 Crystal size (*light*: δ) and volume (*dark*: X_c) of microcrystallites in the resulting films, mapped out on the RF power density/hydrogen dilution ratio plane

The total loss probability (β) is given by the sum of recombination probability and the sticking probability ($\gamma + s$), and the reflection probability is therefore $1-\beta$. Among the reaction probabilities mentioned above, the loss probability has been measured using a grid method (GM) as well as a step-coverage method (SCM). Figures 25.6a and 25.6b show the loss probabilities obtained by GM and SCM, as well as the sticking probability predicted from the film deposition rate as a function of the substrate temperature, respectively.

The loss probability β, related to the reflectivity $(1-\beta)$, is simply dependent on the nature of the site where SiH_3 from the plasma lands. Therefore, the temperature-independent $1-\beta$ seen in Fig. 25.6 suggests that almost all of the surface sites are covered with bonded H over the whole temperature range used in the experiment, although a few dangling bonds are created thermally above 350 °C, as shown by in situ infrared reflection absorption spectroscopy (IRRAS) [25.19]. On the other hand, SiH_3 diffusing across the surface is easily captured by even just a few dangling bonds, because a mobile species on the surface can find a specific site at a distant location from the landing site.

The temperature-independent behavior of $1-\beta$ and the temperature-dependent behavior of s observed in Figs. 25.6a and 25.6b have been perfectly reproduced by theoretical simulations based on the surface-reaction scheme shown in Fig. 25.4, assuming reasonable activation energies for the surface diffusion of SiH_3, the abstraction reaction of H with SiH_3, the saturation of the site containing the dangling bond with SiH_3, and thermal H-removal processes [25.20].

25.2.2 Growth of μc-Si:H

Atomic hydrogen reaching the film-growing surface plays an important role in the growth of μc-Si:H. This has been confirmed in the μc-Si:H-formation map drawing on the radio frequency (RF) power den-

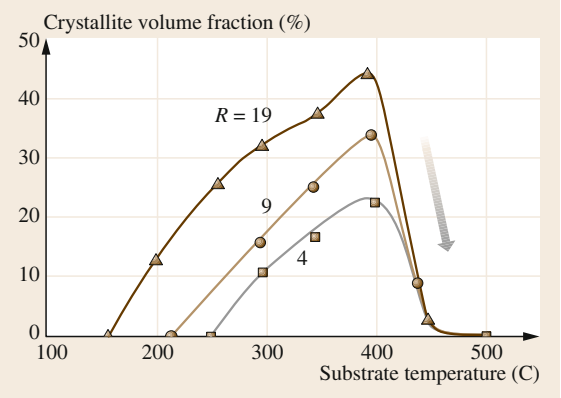


Fig. 25.8 Volume (X_c) of microcrystallites in the resulting films plotted against the substrate temperature during film growth

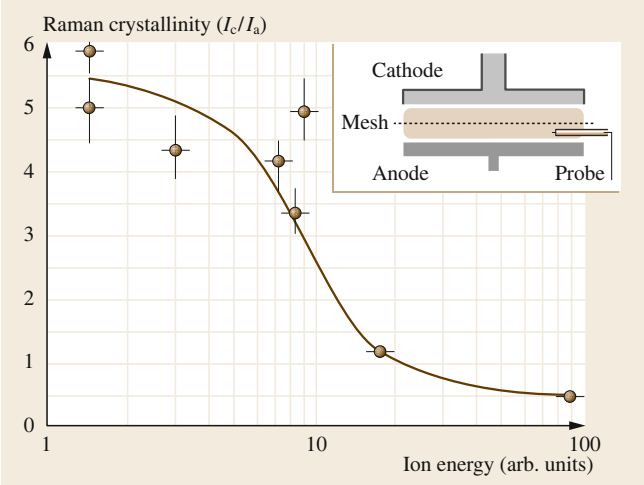


Fig. 25.9 Raman crystallinity in the resulting μc-Si:H as a function of the energy of the ions impinging on the film-growing surface (which is controlled by the application of the positive bias voltage to the mesh electrode during film growth)

sity/hydrogen dilution ratio plane as shown in Fig. 25.7. As seen in the figure, large-crystallite μc-Si:H is prepared under high hydrogen dilution conditions, indicating the importance of atomic H in the growth of μc-Si:H. Figure 25.8 shows the crystalline volume fraction (determined by the x-ray diffraction peak area) in the resulting μc-Si:H as a function of the substrate temperature during film growth for three different hydrogen dilution ratios [25.17]. The crystalline volume fraction increases with increasing substrate temperature, reaching a maximum at around 350 °C, and then suddenly drops to zero above 500 °C. The lack of μc-Si:H formation above 500 °C suggests that surface hydrogen coverage is a requirement for crystallite formation in the resulting film [25.17].

Figure 25.9 shows the negative effect of ionic species impinging on the film-growing surface on the formation of μc-Si:H, studied using a triode reactor, as shown in the inset of the figure [25.21]. The Raman crystallinity I_c/I_a, defined as the ratio of the peak intensity from the crystalline phase at around 520 cm^{-1} to that from the amorphous phase at 480 cm^{-1} in the Raman scattering spectrum, deteriorates as the energy of the ionic species impinging on the film-growing surface (controlled by the bias voltage applied to the mesh electrode) is increased.

The surface reaction behavior of SiH_3 reaching the film-growing surface has also been investigated for μc-Si:H growth using GM and SCM, in contrast to a-Si:H growth [25.18]. Figures 25.10a and 25.10b show the loss probability (β) and the deposition rate (corresponding to the sticking probability s) of SiH_3 reaching the film-growing surface as a function of the substrate temperature for the growth of both μc-Si:H and a-Si:H. Unlike in the case of a-Si:H growth, the loss probability β shows a significant dependence on substrate temperature and the sticking probability s shows no temperature dependence in the case of μc-Si:H growth.

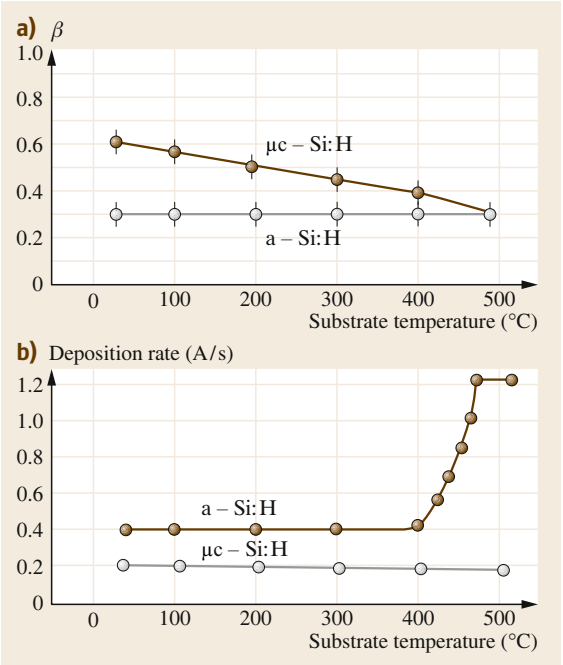


Fig. 25.10 Loss probability β and deposition rate as a function of substrate temperature for μc-Si:H growth in comparison to those for a-Si:H growth

Based on these experimental results, the following properties of the formation process of μc-Si:H can be specified:

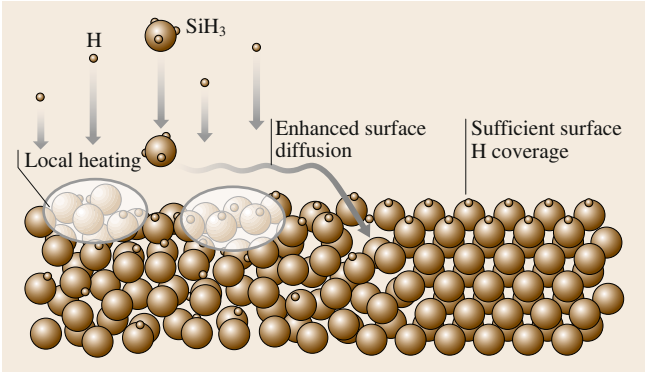


Fig. 25.11 Surface diffusion model for μc-Si:H formation. The *large spheres* and *small spheres* represent Si and H, respectively

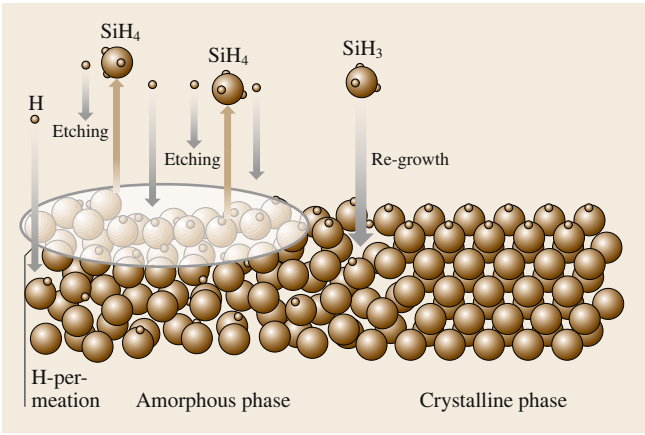


Fig. 25.12 Etching model for μc-Si:H formation

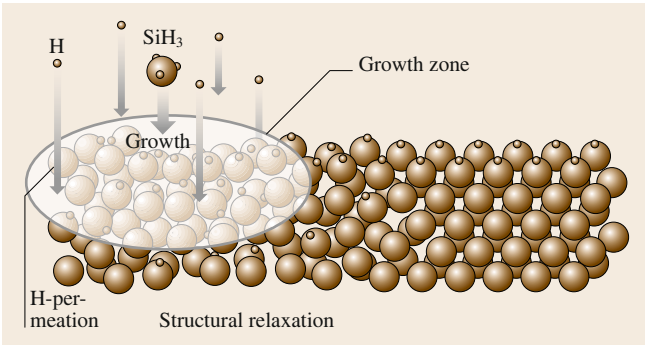


Fig. 25.13 Chemical annealing model for μc-Si:H formation

1. The film precursor is SiH_3, the same as in the case of a-Si:H growth.
2. Atomic hydrogen reaching the film-growing surface plays an important role in the formation of μc-Si:H.
3. The film becomes amorphous when the substrate temperature is higher than 500 °C.
4. High-energy ions impinging on the surface result in crystallinity deterioration.
5. The surface loss probability shows a temperature dependence whereas the sticking probability does not show a temperature dependence in the case of μc-Si:H growth.

Growth Models for μc-Si:H

In an attempt to explain the specific phenomena observed during the formation of μc-Si:H, three models have been proposed:

1. The surface diffusion model [25.17]
2. The etching model [25.22]
3. The chemical annealing model [25.23].

The surface diffusion model is depicted schematically in Fig. 25.11. Here, a high atomic H flux from the plasma results in full bonded hydrogen surface coverage and also local heating through hydrogen exchange reactions on the film-growing surface. These two actions enhance the surface diffusion of film precursors (SiH_3). As a consequence, the SiH_3 adsorbed on the surface can find energetically favorable sites, leading to the formation of an atomically ordered structure (nucleus formation). After the formation of the nucleus, epitaxial-like crystal growth takes place with enhanced surface diffusion of SiH_3 [25.17, 24].

The etching model has been proposed due to the observation that the rate of film growth decreases with increasing hydrogen dilution ratio R. The concept behind the etching model is shown schematically in Fig. 25.12. Atomic H reaching the film-growing surface breaks Si−Si bonds, preferentially weak bonds involved in the amorphous network structure, leading to the removal of Si atoms weakly bonded to other Si atoms. This site is replaced with a new film precursor SiH_3, creating a rigid, strong Si−Si bond, which gives rise to an ordered structure [25.22, 24].

The chemical annealing model has been proposed in order to explain the observation that crystal formation is observed during hydrogen plasma treatment; growth occurs layer-by-layer via an alternating sequence of thin amorphous film growth and hydrogen plasma treatment. Several monolayers of amorphous silicon are deposited, and these layers are exposed to hydrogen atoms produced in the hydrogen plasma. This procedure is repeated several tens of times in order to fabricate the proper thickness to be able to evaluate the film structure. The absence of any significant reduction in film thickness during the hydrogen plasma treatment is difficult to explain using the etching model, and so the chemical annealing model was proposed, as schematically shown in Fig. 25.13. During the hydrogen plasma treatment, many atomic hydrogens permeate through the subsur-

face (the growth zone), giving rise to the crystallization of an amorphous network through the formation of a flexible network without any significant removal of Si atoms [25.23, 24].

These three models have been carefully examined, and the merits and drawbacks of each model have been discussed [25.24, 25].

More microscopic observations have recently been reported, based on the use of in situ diagnostic techniques, and a detailed mechanism for the formation process of μc-Si:H has been proposed.

Formation of the Nucleus

Figure 25.14 shows the evolution in surface roughness during film growth, as obtained by spectroscopic ellipsometry (SE), for three hydrogen dilution ratios R of 0, 10 and 20 [25.26]. As is seen in the figure, after the formation of an island, the enforced coalescence of islands takes place, which results in a smooth surface under μc-Si:H-growth conditions ($R = 20$). After the smooth surface is obtained (formation of the nucleus is confirmed at this point in time), surface roughness is then enhanced due to an orientation-dependent crystal growth rate [25.27]. As soon as the smooth surface appears, a particular surface absorption band is observed in the infrared absorption spectrum, as measured using the in situ attenuated total reflection technique (ATR) during film growth [25.27, 28]. Figure 25.15 shows the surface infrared absorption spectrum, showing the presence of specific bands at 1897 cm^{-1} and 1937 cm^{-1} together with the usual Si−H_x surface and bulk absorption bands (which occur between 2000 cm^{-1} and 2150 cm^{-1}). This new absorption band is assigned to the SiH_2(Sid) complex, which is also sketched in Fig. 25.15. Note that the number density (absorption intensity) of the SiH_2(Sid) complex is found to be proportional to the magnitude of the internal stress in the film just before these complexes appear. A nucleation model has been proposed based on the experimental facts mentioned above. The enforced island coalescence due to the enhanced surface diffusion of SiH_3 gives rise to an internal stress involving many strained Si−Si bonds in the amorphous incubation layer [25.27]. Atomic hydrogen attacks the strained Si−Si bonds, forming SiH_2(Sid) complexes on the film-growing surface. These complexes provide structural flexibility, which enables structural order to be obtained via successive Si−SiH_3 bond formation at this site; in other words it acts as a prenucleation site on the film-growing surface.

Epitaxial-Like Crystal Growth

Figure 25.16 shows a cross-sectional transmission electron microscope (TEM) image of typical μc-Si:H films

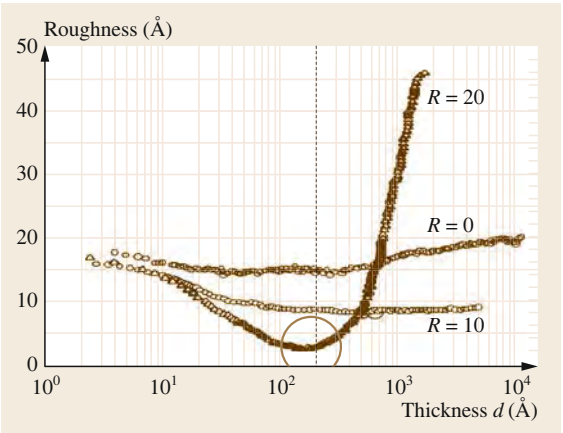


Fig. 25.14 Evolution in surface roughness, as measured using spectroscopic ellipsometry, during film growth for three different R_s

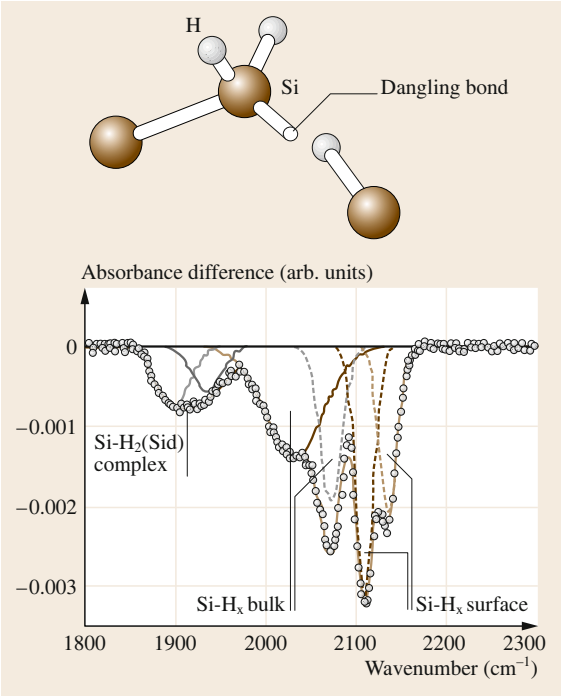


Fig. 25.15 Surface infrared absorption spectrum from the film just before nucleus formation, showing the appearance of the Si−H_2(Sid) complex whose structure is also shown

deposited on a glass substrate [25.29]. In the figure, epitaxial-like crystal growth is clearly observed to occur from the nucleus. It is a well-known fact that epitaxial crystal growth occurs only when the surface diffusion length of the film precursor is sufficiently long. In order to investigate the origin of the enhanced surface diffusion of SiH_3 during the formation of μc-

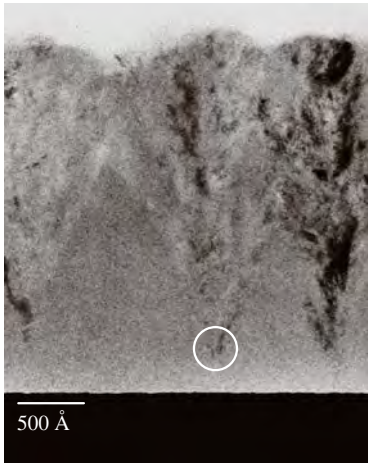


Fig. 25.16 Cross-sectional transmission electron microscope image of typical μc-Si:H deposited on a glass substrate

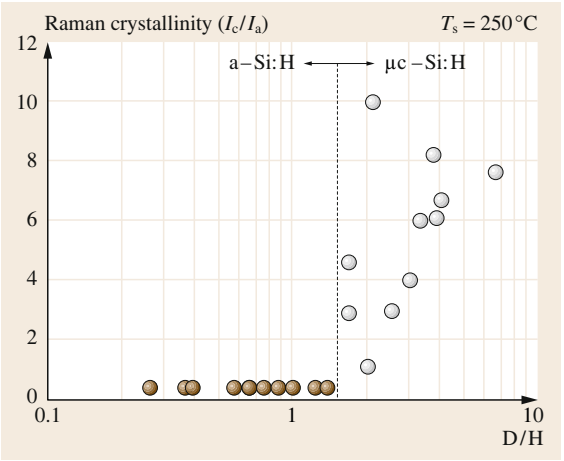


Fig. 25.17 Raman crystallinity plotted against the number density ratio of D/H incorporated into the resulting films prepared from D_2/SiH_4 plasmas

Si:H, an isotope labeling experiment has been carried out [25.30]. D_2 was used as the source gas material (D_2/SiH_4) instead of H_2/SiH_4 during film growth under constant substrate temperature conditions, and the number densities of both D and H incorporated into the resulting film are measured by infrared absorption spectroscopy in order to estimate the degree of H-to-D exchange reactions, which mostly occur on the film-growing surface (and the first layers below).

Figure 25.17 shows the Raman crystallinity I_c/I_a plotted against the number density ratio of the D/H incorporated in the resulting films. As is clearly from the figure, μc-Si:H is obtained only when D/H exceeds a critical value, indicating that the D/H exchange reaction is required to some extent for the formation of μc-Si:H. The D/H exchange reaction involves two steps:

1. Atomic D reaching the film-growing surface abstracts bonded H on the surface (the Eley-Rideal reaction), forming HD and providing a Si dangling bond site;
2. A recombination reaction occurs between the dangling bond and another atomic D adsorbed onto the surface (including the subsurface), forming a Si−D bond.

These abstraction and recombination reactions are known to be strong exothermal reactions, with 1.4 eV and 3.1 eV of energy released, respectively. Considering the highly exothermal nature of the D/H exchange reaction, and the existence of a threshold value for the degree of D/H exchange reactions required for the formation of μc-Si:H, as shown in Fig. 25.17, *local heating* on the film-growing surface is believed to play a major role in the enhanced surface diffusion of the film precursor (SiH_3) during both nucleus formation (strained bond formation and SiH_2(Sid)-complex formation) and the epitaxial-like crystal growth associated with μc-Si:H film growth [25.17, 24, 28, 31].

25.3 Defect Density Determination for a-Si:H and μc-Si:H

One of the most important structural properties of a-Si:H and μc-Si:H for device applications is their dangling bond defect densities, because each dangling bond creates a localized deep state in the band gap of the material, which acts as a recombination center for photoexcited electrons and holes, although the free carrier mobility is another important property in semiconductors.

25.3.1 Dangling Bond Defects

Figure 25.18 shows the dangling bond defect density in the resulting a-Si:H as a function of substrate temperature. This dependency of the dangling bond density on the substrate temperature has been explained by taking into account the steady-state dangling-bond density on the film-growing surface. The structural properties

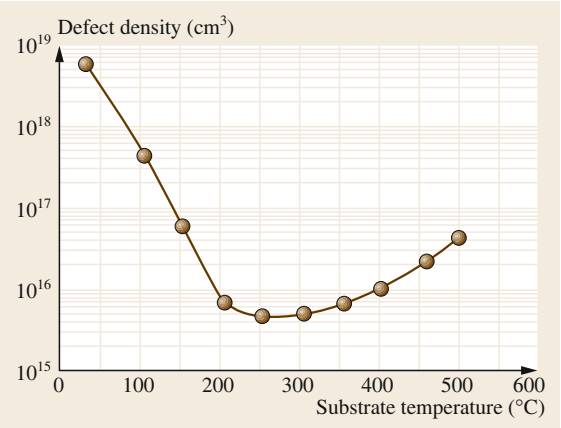


Fig. 25.18 Dangling bond defect density in the resulting a-Si:H films as a function of the substrate temperature during film growth

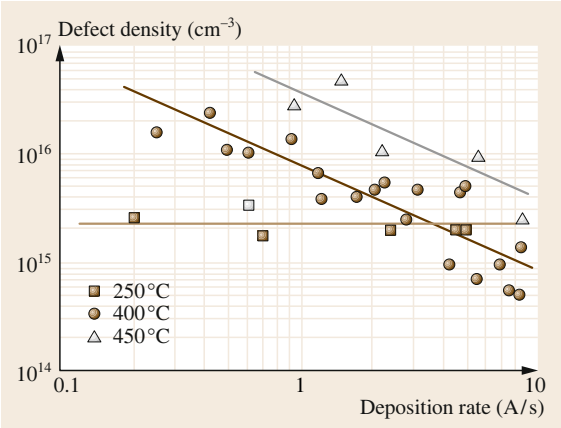


Fig. 25.19 Dangling bond defect density in a-Si:H films plotted against their deposition rates for three different substrate temperatures

of the resulting film are generally largely determined by the steady-state surface properties during the thin film growth process, because the surface formed at any given instant is incorporated into the bulk in the next instant due to the successive layering nature of the film growth [25.32, 33]. The steady state number density of surface dangling bonds is determined by the balance between the rate of generation of dangling bonds and their annihilation rate.

At low substrate temperatures, surface dangling bonds are produced by abstraction reactions of surface bonded H atoms with SiH_3; the reaction rate for this is almost independent of substrate temperature, since the reaction rate involves two rates, i. e., the adsorption rate of SiH_3 on the surface (negative temperature dependence) and dissociation rate of bonded H with SiH_3 (positive temperature dependent). The surface dangling bond is saturated by SiH_3 diffusing across the surface (this is dangling bond annihilation), which exhibits a slower rate at low temperatures due to the slower surface diffusion of SiH_3, which results in high dangling bond density on the steady-state film-growing surface. As a consequence, the dangling-bond density in a-Si:H grown at low substrate temperatures can be as high as $10^{19}\,cm^{-3}$. When the substrate temperature is increased during film growth, the steady state number density of surface dangling bonds is reduced drastically due to the combination of the temperature-independent H-abstraction reaction and the thermally enhanced surface diffusion of SiH_3, which gives rise to a minimum defect density of $10^{15}\,cm^{-3}$ in the resulting a-Si:H at a substrate temperature of $\approx 250\,°C$. The increased dangling bond density of the a-Si:H prepared at substrate temperatures higher than 350 °C is explained by the increased rate of generation of steady-state dangling bonds due to the addition of a new term in the generation rate of these bonds associated with the thermal removal of surface bonded H, although its annihilation rate strictly increases with increasing substrate temperature [25.33, 34].

We note here that a remarkable increase in the contribution from short-lifetime species such as SiH_2, SiH and Si, which show high reactivity and no diffusion on the film-growing surface, is observed upon the depletion of SiH_4-parent molecules, which causes an increase in the number density of dangling bond defects in the resulting films through an enhancement of the dangling bond generation rate and a reduction of the dangling bond annihilation rate on the film-growing surface [25.20, 32].

Based on our understanding of the defect density determination reaction during film growth, several trials have attempted to control the defect density in a-Si:H. The steady state defect density on the film-growing surface could be reduced when the growth rate was much faster than the thermal H removal rate in the substrate temperature range above 350 °C, where the steady state defect density is mainly determined by the thermal H removal process. Figure 25.19 shows the number density of dangling bond defects in a-Si:H films plotted against their growth rate. As expected, the defect density in a-Si:H shows no growth- rate dependence when the film is prepared at substrate temperatures lower than 300 °C, whereas the defect density monotonically decreases with increasing growth rate when the film is deposited at 400 °C and 450 °C, and a defect density of $10^{14}\,cm^{-3}$ has been demonstrated [25.35].

25.3.2 Dangling Bond Defect Density in μc-Si:H

Figure 25.20 shows typical dangling bond defect densities in μc-Si:H as a function of substrate temperature together with those in its a-Si:H counterpart [25.33, 35, 36]. As seen from the figure, the substrate temperature-dependent dangling bond defect density in μc-Si:H in the temperature range above 300 °C shows a similar trend to that for a-Si:H, indicating that the defect determination reaction on the film-growing surface is identical for both μc-Si:H and a-Si:H growth in this temperature range; in other words, the increase in defect density with increasing substrate temperature is controlled by the thermal H removal process on the film-growing surface. On the other hand, the number density of defects in μc-Si:H prepared at low substrate temperatures shows much lower values than those in a-Si:H. Considering that the defect density in the resulting film is largely determined by the defect density on the steady-state film-growing surface, the much lower defect density in μc-Si:H is caused by an increase in the defect annihilation rate (through an enhanced surface diffusion of SiH_3 film precursors) on the film-growing surface due to the local heating from H-exchange reactions that occur during the course of μc-Si:H film growth.

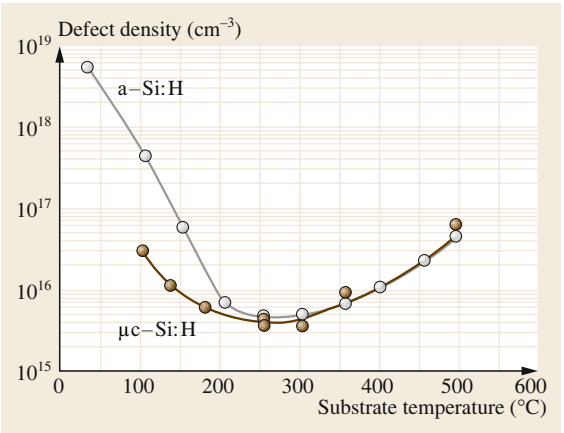


Fig. 25.20 Dangling bond defect density in μc-Si:H films as a function of substrate temperature during film growth in comparison to that in a-Si:H

25.4 Device Applications

a-Si:H and μc-Si:H are highly promising materials applicable to electronic or optoelectronic thin-film devices such as thin-film transistors (TFTs), position sensors, color sensors, solar cells, etc. [25.1, 2].

A thin-film transistor array with a-Si:H active layer has been developed for switching devices in liquid crystal displays (LCD), and large-area LCDs more than 50 in across have already been made commercially available based on this technology. A laser crystallization technique has been developed in order to increase the carrier mobility of a-Si:H-based TFTs and thus reduce the active area of the transistors in the LCD. LCDs with a-Si:H or laser-crystallized thin-film Si-based transistor arrays are widely used as flat panel displays in televisions and monitors, where they are in competition with plasma display panels (PDPs) and other flat panel display systems. If the carrier mobility in as-deposited μc-Si:H is drastically improved through the enhanced control of the film growth process, μc-Si:H will be widely used not only for thin-film transistors but also for signal-scanning devices such as charge-coupled devices monolithically arranged in the periphery of the LCD.

Thin-film Si-based solar cells has also been widely expected to provide low-cost photovoltaics. Actually, a-Si:H-based solar cells have already been widely used in pocket calculators, and now large-area solar cells are being developed for electricity generation. One big advantage of using a-Si:H for solar cell applications is its large optical absorption coefficient compared to single-crystalline or polycrystalline silicon counterparts, showing indirect optical transition properties, and so a thickness of less than 1 μm is enough to absorb sufficient sunlight for electricity generation when using a-Si:H-based solar cells. Low-temperature processes using PECVD are also advantageous in terms of reducing the cost of producing a-Si:H-based solar cells. However, there is a well-known phenomenon that occurs in a-Si:H called photo-induced degradation [25.37], where the conversion efficiency of a-Si:H-based solar cells, usually ≈ 10%, is degraded to less than 8% after prolonged exposure to light. Recombinations photogenerated electrons and holes are believed to trigger this photoinduced degradation in a-Si:H; therefore, increasing the field in the a-Si:H-based solar cell has proved an effective way to reduce the degradation. In fact, 0.1 μm-thick a-Si:H-based solar cells do not show any photoinduced degradation, although the conversion efficiency naturally deteriorates due to the reduced absorption of sunlight. The fabrication of a tandem-type stacked solar cell structure consisting of a top cell with a thin a-Si:H layer and a bottom cell

containing narrow-gap materials such as a-SiGe:H, μc-Si:H and μc-SiGe:H has been proposed as a promising way to overcome photoinduced degradation and so to achieve high conversion efficiency in thin-film silicon-based solar cells. Initially, a-SiGe:H was adopted for the bottom cell material [25.38]; however, this material also shows severe photoinduced degradation. Recently, μc-Si:H or μc-SiGe:H have been proposed as promising candidates for bottom cell materials, because these materials do not exhibit photoinduced degradation [25.39, 40]. In this proposal, high rates of growth of those materials are crucial to the low-cost fabrication of tandem-type solar cells, since μc-Si:H and μc-SiGe:H undergo largely indirect optical transitions. Therefore, urgent material issues for the realization of low-cost/high-efficiency thin-film silicon-based solar cells include the need to improve the photoinduced stability of high-quality a-Si:H and the need to achieve a high rate of growth of device-grade μc-Si:H or μc-SiGe:H.

25.5 Recent Progress in Material Issues Related to Thin-Film Silicon Solar Cells

25.5.1 Controlling Photoinduced Degradation in a-Si:H

A relationship between the degree of photoinduced degradation and the dihydride bonding ($Si{-}H_2$) density has been reported in a-Si:H prepared under a variety of deposition conditions, where the substrate temperature, plasma-excitation frequency, gas-flow rate, hydrogen dilution ratio, working pressure, power density, etc. have all been varied [25.41]. Figure 25.21 shows the degree of photoinduced degradation, defined as the difference in the fill factors of photo-I–V characteristics of Ni-a-Si:H Schottky diode before and after light soaking plotted against the $Si{-}H_2$ density in a-Si:H film, as measured by infrared absorption spectroscopy. Furthermore, it has been suggested from mass spectrometric results that the $Si{-}H_2$ density in the resulting a-Si:H is strongly increased by the contributions from higher silane-related species (HSRS) such as Si_4H_9 when the substrate temperature is kept constant. The contribution of these HSRS to the film precursor (SiH_3) has been theoretically analyzed using a couple of gas-phase reaction-rate equations [25.42]. This analysis predicted that the contribution ratio is a complex function of the electron temperature in the plasma, the electron density in the plasma, the gas temperature, the hydrogen dilution ratio R, and the gas residence time during film growth. Based on our understanding of the network structure ($Si{-}H_2$ bonding configuration) responsible for the photoinduced degradation in a-Si:H and the chemical species (HSRS) responsible during film growth (shown schematically in Fig. 25.22), a guiding principle for obtaining highly stabilized a-Si:H has been proposed [25.42, 43].

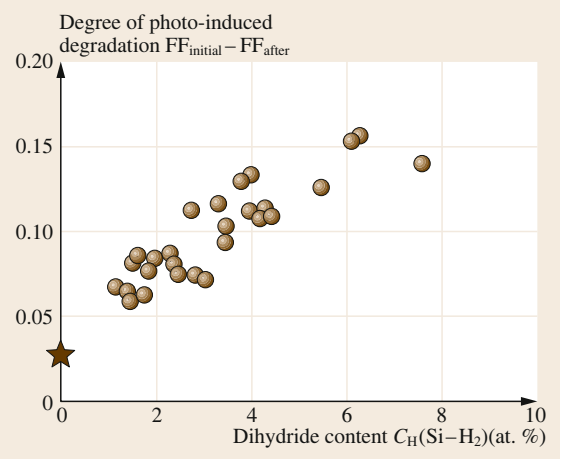


Fig. 25.21 Relationship between the degree of photoinduced degradation and the dihydride ($Si{-}H_2$) content in a-Si:H prepared under various deposition conditions. The *star* symbol represents for a-Si:H prepared under conditions of a reduced ratio of contributions from higher silane-related species versus the SiH_3 contribution at a substrate temperature of 250 °C

By following the guiding principle, a-Si:H with minimized $Si{-}H_2$ density in the network has been prepared by adjusting the plasma parameters during film growth under high growth rate conditions, and an a-Si:H-based solar cell showing a stable conversion ef-

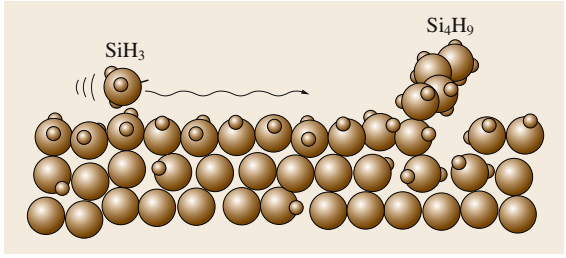


Fig. 25.22 Surface reaction image of the incorporation of $Si{-}H_2$ bonds into the resulting a-Si:H due to the contributions from higher silane-related species such as Si_4H_9

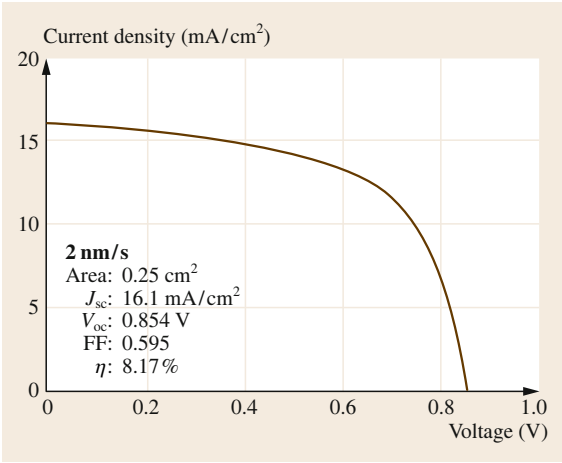


Fig. 25.23 Photo-I–V characteristics of a stable n–i–p a-Si:H-based solar cell fabricated at a high growth rate of 2 nm/s

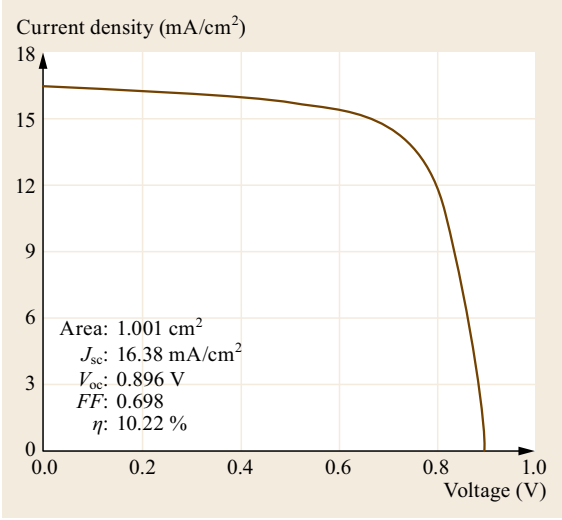


Fig. 25.24 Photo-I-V characteristics of a stable p–i–n a-Si:H-based solar cell fabricated using a triode reactor

ficiency of 8.2% (Fig. 25.23) has been fabricated at a high growth rate of 2 nm/s [25.41].

Moreover, a-Si:H containing a Si−H_2 density of 0% has been successfully prepared using a triode reactor at substrate temperatures as low as 250 °C by making use of the difference in the gas phase diffusion coefficients of SiH_3 (light) and HSRS (heavy) during film growth. High efficiency stable p–i–n a-Si:H based solar cells (efficiency after light soaking of 10.2% as shown in Fig. 25.24) have been fabricated using the triode reactor, although the growth rate is as low as 0.1 nm/s [25.44].

25.5.2 High Growth Rates of Device-Grade μc-Si:H

In the past, conventional high hydrogen dilution methods performed at relatively low working gas pressures (several tens of mTorr) have been used to obtain device-grade μc-Si:H (with low dangling bond defect densities of $\approx 10^{16}/cm^3$) [25.17]. Recently, a simple concept for preparing device-grade μc-Si:H at a high growth rate has been proposed, known as the high-pressure depletion (HPD) method [25.21, 45]. The production rate of the film precursor SiH_3 in the plasma, which is proportional to the growth rate of μc-Si:H, is determined by the product of the electron density and the number density of SiH_4. To increase the production rate of SiH_3 in the plasma, a high power density (radio frequency, RF, or very high frequency, VHF) and high partial pressures are needed, because the electron density is basically a function of the power density applied to the plasma and the number density of SiH_4 is proportional to the partial pressure. However, hydrogen atoms (the chemical species responsible during the formation of μc-Si:H) are strongly scavenged by SiH_4 molecules during their transportation from their production area to the film-growing surface. In order to enhance the survival of the hydrogen atoms, it has been suggested that the SiH_4 molecules should be depleted by applying high power density to the plasma. As conditions of high total pressure are also useful for decreasing the effects of ion bombardment during film growth through the reduction of the electron temperature in the plasma, high working pressures along with SiH_4 depletion conditions (HPD) have recently been popularly adopted for the high-rate growth of device-grade μc-Si:H. It has also been suggested that short lifetime species such as SiH_2, SiH, Si are scavenged by a large number of H_2 molecules present under HPD conditions, although the reaction rate of short lifetime species with H_2 is lower than that with SiH_4 [25.9], by two order of magnitudes.

Figure 25.25 shows the number density of dangling bond defects, as measured by electron spin resonance (ESR), in the resulting μc-Si:H as a function of growth rate [25.46]. When the conventional low pressure regime is used for the growth of μc-Si:H, the defect density increases exponentially with increasing growth rate, as seen in the figure (see A). This is caused by both an increase in the ion bombardment (due to the high power density) and an increase in the contributions from short-lifetime chemical species such as SiH_2, SiH and Si due to the reduced ability of those species to react with SiH_4 and H_2 (due to the low pressure and SiH_4 depletion) during film growth. On the other hand, the slope in Fig. 25.25 becomes shallower when HPD con-

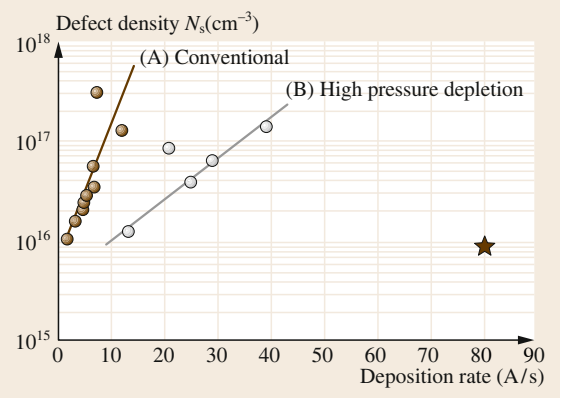


Fig. 25.25 Relationships between the dangling bond defect density in μc-Si:H films and their deposition rates for (A) μc-Si:H prepared using the conventional low-pressure regime and for (B) μc-Si:H prepared using the high-pressure depletion (HPD) method. The *star* symbol represents μc-Si:H prepared under HPD conditions with the novel cathode design

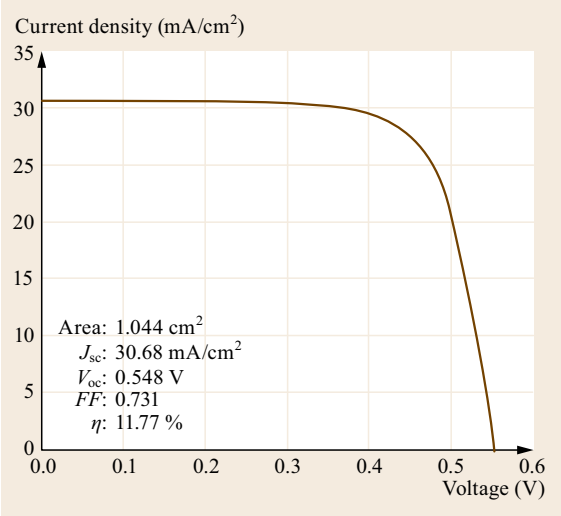


Fig. 25.26 Photo-I–V characteristics of a p–i–n μc-Si:H-based solar cell fabricated at reasonably high growth rate of 2 nm/s under HPD conditions

ditions are used during the growth of μc-Si:H, as shown in the figure (see B), which illustrates the usefulness of the HPD method for obtaining high-quality μc-Si:H at high growth rates.

The validity of the HPD method has also been demonstrated during the fabrication of μc-Si:H-based solar cells. Figure 25.26 shows the photo-I–V characteristics of a μc-Si:H-based p–i–n single-junction solar cell prepared at a high growth rate of 2 nm/s using the HPD method, which exhibits a reasonably high conversion efficiency of 11.8% [25.47].

However, HPD conditions require that the spacing between the cathode and anode is reduced in the conventional capacitively coupled plasma reactor, giving rise to nonuniform plasma production and nonuniform film growth. To overcome this problem encountered when using HPD conditions, the structure of the cathode surface has been designed to produce uniform plasma production even in a large-area parallel plate electrode configuration. Figure 25.27 shows the structure of this novel design of cathode. A multitude of holes (hollows) with interconnecting slots are arranged on the cathode surface, which cause strong coupling between the high-density plasmas produced in each hole where source gas injection is performed. Using this new type of cathode, quite high growth rates (more than 8 nm/s) have been obtained along with reasonably low defect densities in the resulting μc-Si:H, as shown by star symbol in Fig. 25.25 [25.48].

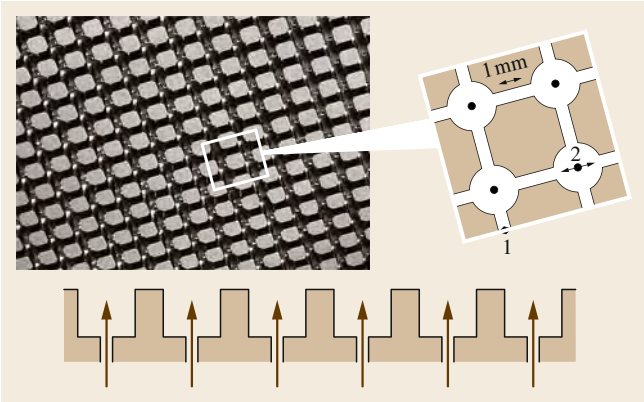


Fig. 25.27 Photograph of the surface structure on the cathode designed for the production of high-density/uniform plasmas

25.6 Summary

In this chapter, the processes involved in the growth of a-Si:H and μc-Si:H from SiH_4 and SiH_4/H_2 plasma have been interpreted in detail. The defect density determination reaction that takes place on the film-growing surface was discussed in order to obtain clues that may lead to enhanced optoelectronic properties in those materials. The recent status of work done in the fields of thin-film transistors and solar cells was reviewed, as these are the main device applications of those materials. Recent progress in resolving material issues related to solar cell applications were also described.

Finally, we note here that the concepts used in and our understanding of the film growth process mentioned here are widely applicable to other processes, especially processes where thin films are grown from reactive plasmas.

References

25.1 W.E. Spear, P.G. LeComber: Solid State Commun. **17**, 1193 (1975)
25.2 C.R. Wronski, D.E. Carlson, R.E. Daniel: Appl. Phys. Lett. **29**, 602 (1976)
25.3 T. Moustakas: *Semiconductors and Semimetals*, Vol. 21A (Academic, New York 1984) p. 55
25.4 T. Saito, S. Muramatsu, T. Shimada, M. Migitaka: Appl. Phys. Lett. **42**, 678 (1983)
25.5 Y. Mishima, M. Hirose, Y. Osaka, K. Nagamine, Y. Ashida, K. Isogaya: Jpn. J. Appl. Phys. **22**, L46 (1983)
25.6 T. Fuyuki, K.Y. Du, S. Okamoto, S. Yasuda, T. Kimoto, M. Yoshimoto, H. Matsunami: J. Appl. Phys. **64**, 2380 (1988)
25.7 A.H. Mahan, B.P. Nelson, S. Salamon, R.S. Crandall: Mater. Res. Soc. Proc. **219**, 673 (1991)
25.8 M. Tsuda, S. Oikawa, K. Saito: J. Chem. Phys. **91**, 6822 (1989)
25.9 J. Perrin, O. Leroy, M.C. Bordage: Contrib. Plasma Phys. **36**, 3 (1996)
25.10 A. Matsuda, K. Nakagawa, K. Tanaka, M. Matsumura, S. Yamasaki, H. Okushi, S. Iizima: J. Non-Cryst. Solids **35–36**, 183 (1980)
25.11 A. Matsuda, K. Tanaka: Thin Solid Films **92**, 171 (1982)
25.12 Y. Matsumi, T. Hayashi, H. Yoshikawa, S. Komiya: J. Vac. Sci. Technol. A **4**, 1786 (1986)
25.13 N. Itabashi, N. Nishiwaki, M. Magane, T. Goto, A. Matsuda, C. Yamada, E. Hirota: Jpn. J. Appl. Phys. **29**, 585 (1990)
25.14 N. Itabashi, N. Nishiwaki, M. Magane, S. Saito, T. Goto, A. Matsuda, C. Yamada, E. Hirota: Jpn. J. Appl. Phys. **29**, L505 (1990)
25.15 K. Tachibana, T. Mukai, H. Harima: Jpn. J. Appl. Phys. **30**, L1208 (1991)
25.16 A. Kono, N. Koike, H. Nomura, T. Goto: Jpn. J. Appl. Phys. **34**, 307 (1995)
25.17 A. Matsuda: J. Non-Cryst. Solids **59–60**, 767 (1983)
25.18 A. Matsuda, T. Goto: Mater. Res. Soc. Proc. **164**, 3 (1990)
25.19 Y. Toyoshima, K. Arai, A. Matsuda, K. Tanaka: J. Non-Cryst. Solids **137–138**, 765 (1991)
25.20 J.L. Guizot, K. Nomoto, A. Matsuda: Surf. Sci. **244**, 22 (1991)
25.21 M. Kondo, M. Fukawa, L. Guo, A. Matsuda: J. Non-Cryst. Solids **266–269**, 84 (2000)
25.22 C.C. Tsai, G.B. Anderson, R. Thompson, B. Wacker: J. Non-Cryst. Solids **114**, 151 (1989)
25.23 K. Nakamura, K. Yoshida, S. Takeoka, I. Shimizu: Jpn. J. Appl. Phys. **34**, 442 (1995)
25.24 A. Matsuda: Thin Solid Films **337**, 1 (1999)
25.25 K. Saito, M. Kondo, M. Fukawa, T. Nishimiya, W. Futako, I. Shimizu, A. Matsuda: Res. Soc. Proc. Mater. **507**, 843 (1998)
25.26 J. Koh, Y. Lee, H. Fujiwara, C.R. Wronski, R.W. Collins: Appl. Phys. Lett. **73**, 1526 (1998)
25.27 H. Fujiwara, M. Kondo, A. Matsuda: Surf. Sci. **497**, 333 (2002)
25.28 H. Fujiwara, Y. Toyoshima, M. Kondo, A. Matsuda: J. Non-Cryst. Solids **266–269**, 38 (2000)
25.29 H. Fujiwara, M. Kondo, A. Matsuda: Phys. Rev. B **63**, 115306 (2001)
25.30 S. Suzuki, M. Kondo, A. Matsuda: J. Non-Cryst. Solids **299–302**, 93 (2002)
25.31 H. Fujiwara, M. Kondo, A. Matsuda: Jpn. J. Appl. Phys. **41**, 2821 (2002)
25.32 A. Matsuda, K. Nomoto, Y. Takeuchi, A. Suzuki, A. Yuuki, J. Perrin: Surf. Sci. **227**, 50 (1990)
25.33 G. Ganguly, A. Matsuda: Phys. Rev. B **47**, 3361 (1993)
25.34 G. Ganguly, A. Matsuda: J. Non-Cryst. Solids **164–166**, 31 (1993)
25.35 G. Ganguly, A. Matsuda: Jpn. J. Appl. Phys. **31**, L1269 (1992)
25.36 Y. Nasuno, M. Kondo, A. Matsuda: Tech. Digest of PVSEC-12. Jeju (2001) 791
25.37 D.L. Staebler, C.R. Wronski: Appl. Phys. Lett. **28**, 671 (1977)
25.38 K. Saito, M. Sano, K. Matsuda, T. Kondo, T. Nishimoto, K. Ogawa, I. Kajita: Proc. WCPEC-2 Vienna (1998) p. 351
25.39 J. Meier, P. Torres, R. Platz, S. Dubail, U. Kroll, J.A. Anna Selvan, N. Pellaton Vaucher, Ch. Hof, D. Fischer, H. Keppner, A. Shah, K.-D. Ufort, P. Giannoules: Mater. Res. Soc. Proc. **420**, 3 (1996)
25.40 K. Yamammoto, M. Yoshimi, T. Suzuki, Y. Tawada, Y. Okamoto, A. Nakajima: Mater. Res. Soc. Proc. **507**, 131 (1998)
25.41 T. Nishimoto, M. Takai, H. Miyahara, M. Kondo, A. Matsuda: J. Non-Cryst. Solids **299–302**, 1116 (2002)
25.42 M. Takai, T. Nishimoto, T. Takagi, M. Kondo, A. Matsuda: J. Non-Cryst. Solids **266–269**, 90 (2000)

25.43 M. Takai, T. Nishimoto, M. Kondo, A. Matsuda: Appl. Phys. Lett. **77**, 2828 (2000)
25.44 T. Matsui, K. Maejima, A. Bidiville, H. Sai, T. Koida, T. Suezaki, M. Matsumoto, K. Saito, I. Yoshida, M. Kondo: Proc. WCPEC-6 Kyoto (2014), 2Tu0.6.2
25.45 L. Guo, M. Kondo, M. Fukawa, K. Saito, A. Matsuda: Jpn. J. Appl. Phys. **37**, L1116 (1998)
25.46 M. Kondo, T. Nishimoto, M. Takai, S. Suzuki, Y. Nasuno, A. Matsuda: Tech. Digest of PVSEC-12 Jeju (2001) 41
25.47 H. Sai, K. Maejima, T. Matsui, T. Koida, M. Kondo, S. Nakao, Y. Takeuchi, H. Katayama, I. Yoshida: Proc. WCPEC-6 Kyoto (2014), 2We0.14.1
25.48 C. Niikura, M. Kondo, A. Matsuda: Proc. WCPEC-3 Osaka (2003), 5P-D4-03

26. Ferroelectric Materials

Roger Whatmore

Ferroelectric materials offer a wide range of useful properties. These include ferroelectric hysteresis (used in nonvolatile memories), high permittivities (used in capacitors), high piezoelectric effects (used in sensors, actuators and resonant wave devices such as radio-frequency filters), high pyroelectric coefficients (used in infra-red detectors), strong electro-optic effects (used in optical switches) and anomalous temperature coefficients of resistivity (used in electric-motor overload-protection circuits). In addition, ferroelectrics can be made in a wide variety of forms, including ceramics, single crystals, polymers and thin films – increasing their exploitability. This chapter gives an account of the basic theories behind the ferroelectric effect and the main ferroelectric material classes, discussing how their properties are related to their composition and the different ways they are made. Finally, it reviews the major applications for this class of materials, relating the ways in which their key functional properties affect those of the devices in which they are exploited.

26.1 **Definitions and Background** 589
26.1.1 Basic Ferroelectric Characteristics and Models 591
26.2 **Ferroelectric Materials** 593
26.2.1 Ferroelectric Oxides 593
26.2.2 Triglycine Sulphate (TGS) 598
26.2.3 Polymeric Ferroelectrics 599
26.3 **Ferroelectric Materials Fabrication Technology** 600
26.3.1 Single Crystals 600
26.3.2 Ceramics 601
26.3.3 Thick Films 604
26.3.4 Thin Films 605
26.4 **Ferroelectric Applications** 607
26.4.1 Dielectrics 607
26.4.2 Computer Memories 607
26.4.3 Piezoelectrics 608
26.4.4 Pyroelectrics 611
References 613

26.1 Definitions and Background

Ferroelectric materials offer a very wide range of useful properties for the electronic engineer to exploit. As we will see, they are also a class of materials that is hard to define accurately in a single sentence. It is useful to start from the class of insulating materials that form dielectrics; in other words materials that will sustain a dielectric polarisation under the application of an electric field. There exists a set of these materials for which the crystal structure lacks a centre of symmetry. (If a crystal structure has a centre of symmetry, it means that for every atom in the structure there is a point in the unit cell through which inversion will bring one to the same type of atom.) A list of the non-centrosymmetric, or *acentric*, point groups is given in Table 26.1. All of the crystalline materials whose structures possess these point groups (with the exception of group 432) exhibit the phenomenon of piezoelectricity, which means that stress will generate a charge separation on the faces of the crystal (the direct piezoelectric effect) and will undergo mechanical strain when subjected to an electric field (the converse piezoelectric effect). Both effects are widely exploited in electronic devices. A well-known example of a non-centrosymmetric material is the mineral α-quartz, which is used for the piezoelectric resonators employed for frequency filtering and electronic clocks.

A sub-set of the non-centrosymmetric crystals also possess a unique axis of symmetry. These crystals are said to be *polar*. The polar point groups are also listed in Table 26.1. Polar crystals are piezoelectric (as they are acentric) and also exhibit pyroelectricity, which means that a charge separation will appear on their surfaces when their temperature is changed. Polar structures effectively have a dielectric polarisation *built in* to the unit

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_26

Table 26.1 Polar and acentric (non-centrosymmetric) point groups

Crystal system	Polar (acentric)	Nonpolar (acentric)
Triclinic	1	
Monoclinic	2, *m*	
Orthorhombic	*mm*2	222
Trigonal	3, 3*m*	32
Hexagonal	6, 6*mm*	$\bar{6}$, $\bar{6}m2$
Tetragonal	4, 4*mm*	$\bar{4}$, 422, $\bar{4}2m$
Cubic	None	23, $\bar{4}3m$, 432

cell of the crystal structure. This is sometimes called a *spontaneous polarisation*. The application of stress or a change in temperature causes a change in this dipole moment and it is this change that causes the separation of charge on the surfaces of the crystal. The direction and magnitude of the spontaneous polarisation in a polar dielectric can be changed by the application of an electric field, but on removal of the field it will return to its zero-field value. A well-known example of a polar dielectric is ZnO, which possesses the wurtzite crystal structure in which Zn^{2+} ions sit in the tetrahedrally coordinated sites between hexagonal close-packed layers of oxygen ions. Thin films of ZnO are widely used for piezoelectric applications.

A sub-set of the set of polar dielectrics exists for which the application of a field of sufficient magnitude will cause the spontaneous polarisation to switch to a different, stable direction. Upon removal of the field the polarisation will not spontaneously return to its original direction and magnitude. These crystals are called *ferroelectric*. We can view the relationships between the sets of ferroelectric, polar, acentric and centrosymmetric dielectrics as shown in the Venn diagram in Fig. 26.1.

The history of ferroelectrics is long, as has been described in the excellent review by *Busch* [26.1]. Many very distinguished scientists were involved in its early development, including Brewster (who was one of those who studied pyroelectricity), J. & P. Curie (who discovered piezoelectricity), Boltzmann, Pockels and Debye, to name but a few. Indeed, the term ferroelectricity was first coined by E. Schrodinger. However, credit for the discovery of the effect goes to Joesph Valasek, who found in 1920 that the polarisation of sodium potassium tartrate (Rochelle salt) could be switched by the application of an electric field, thus providing the first demonstration of the process that is the hallmark of ferroelectricity.

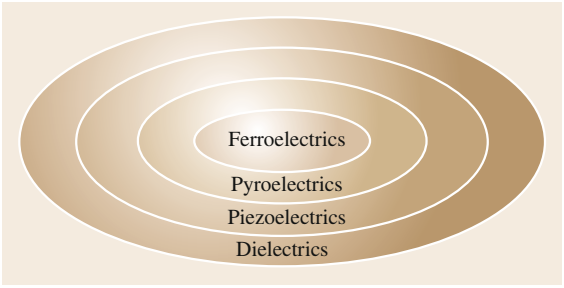


Fig. 26.1 Venn diagram showing how ferroelectrics fit into the different classes of dielectric materials

A reasonable working definition of ferroelectricity is *a polar dielectric in which the polarisation can be switched between two or more stable states by the application of an electric field*. However, as we will see in the discussion which follows, there are exceptions to this definition: some ferroelectrics are semiconducting (and thus are not dielectrics because they cannot sustain an electrical polarisation); the spontaneous polarisation in some ferroelectrics cannot be switched because they cannot sustain an electric field of sufficient magnitude to effect the switching, either because they reach electrical breakdown first, or because they are too conducting.

Since the initial discovery of ferroelectricity in Rochelle salt, the effect has been demonstrated in a wide range of materials, from water-soluble crystals through oxides to polymers, ceramics and even liquid crystals. Many of these will be discussed in this chapter. The range of useful properties exhibited by ferroelectrics covers:

- Ferroelectric hysteresis is used in nonvolatile computer information storage.
- Ferroelectrics can exhibit very high relative permittivities (several thousand) which means that they are widely used in capacitors.
- The direct piezoelectric effect (the generation of charge in response to an applied stress) is widely used in sensors such as accelerometers, microphones, hydrophones etc.
- The converse piezoelectric effect (the generation of strain in response to an applied electric field) is widely used in actuators, ultrasonic generators, resonators, filters etc.
- The pyroelectric effect (the generation of charge in response to a change in material temperature) is widely used in uncooled infra-red detectors.
- The electro-optic effect (a change in birefringence in response to an applied electric field) is used in laser Q-switches, optical shutters and integrated optical (photonic) devices.
- Ferroelectrics exhibit strong nonlinear optical effects that can be used for laser frequency doubling and optical mixing.
- Illumination of transparent ferroelectrics with light of sufficient energy causes excitation of carriers into

the conduction band. Their movement under the internal bias field caused by the spontaneous polarisation causes a refractive-index modulation that can be used for a variety of optical applications, including four-wave mixing and holographic information storage.

- Ferroelectrics exhibit strong coupling between stress and birefringence, which can be used to couple acoustic waves to optical signals with applications in, for example, radar signal processing.
- Doping certain ferroelectric ceramics with electron donors (e.g., $BaTiO_3$ with La^{3+}) can render them semiconducting. Heating these ceramics through their Curie temperature causes a very large, reversible increase in resistivity (by several orders of magnitude in some cases) over a narrow range of temperature (ca. 10 °C). This large positive temperature coefficient of resistance (PTCR) is widely exploited in electric-motor overload-protection devices and self-stabilising ceramic heating elements.

26.1.1 Basic Ferroelectric Characteristics and Models

The switchable spontaneous polarisation in ferroelectrics gives rise to the first characteristic property of the materials: ferroelectric hysteresis. Figure 26.2 shows a schematic plot of polarisation versus electric field as would be observed on a typical ferroelectric. As the field is increased from zero, the overall polarisation in the crystal increases as the polarisations in different dipolar regions are aligned. Eventually, it reaches a saturation point where the only further increase in P is that due to the relative permittivity of the material. (The gradient of the P/E curve for a linear dielectric is equal to its permittivity). Extrapolation of this line back to the abscissa gives the saturation value of the spontaneous polarisation (P_s). Reduction of the field to zero leaves a *remanent polarisation* (P_r), which is usually slightly less than P_s. A negative field will cause the polarisation to reduce, until it reaches zero at the *coercive field* ($-E_c$). A further negative increase in the field will eventually cause a reverse saturation polarisation ($-P_s$) to develop. When the field returns to zero the crystal is left with a negative remanent polarisation ($-P_r$). Increasing the field once more increases the polarisation from $-P_r$ to zero at E_c, and then to $+P_s$, completing the ferroelectric hysteresis loop. The ability to switch the polarisation between two states gives rise to the first application of ferroelectrics – as a nonvolatile memory storage medium. It also permits polycrystalline ferroelectrics (especially ceramics) to be polarised. The fact that the polarisation can possess different directions within different regions of the same crystal gives rise to the existence of *ferroelectric domains*. These are given names according to the angle between the polarisation between adjacent regions. Hence, two adjacent regions in which the polarisations are orientated at 180° to each other are, naturally, called 180° domains. In many ferroelectrics, the polarisation can choose between many different possible directions so that adjacent regions can have polarisations at, for example, 90° to one another.

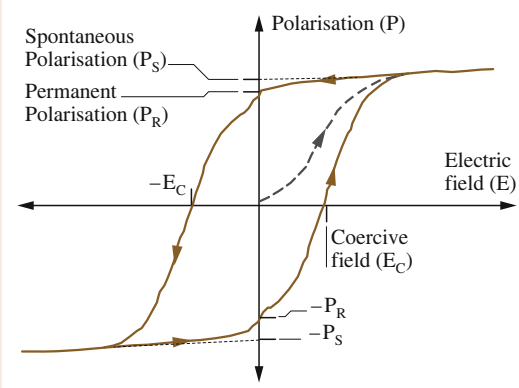


Fig. 26.2 Ferroelectric hysteresis loop

For most ferroelectrics, the polar state only exists over a limited range of temperatures. As the temperature is raised, a point is reached at which there is a transition from the polar, ferroelectric phase to a nonpolar, non-ferroelectric phase (called the *paraelectric* phase). In all cases, the paraelectric phase possesses a higher crystal symmetry than the ferroelectric phase into which it transforms. The temperature at which this occurs is called the *Curie temperature* (T_C). This ferroelectric-to-paraelectric phase transition is a characteristic of most ferroelectrics, but again there are exceptions. Some ferroelectric materials melt or decompose before T_C is reached; the polymeric ferroelectric polyvinylidene fluoride (PVDF) is one such example. For most ferroelectrics (called *proper ferroelectrics*), as the Curie temperature is approached from above, the relative permittivity is observed to increase, reaching a peak at T_C and decreasing below T_C. (There is a class of ferroelectrics, known as *improper* or *extrinsic ferroelectrics*, for which there is no peak in permittivity, just an anomaly. These are described further below.) Very high values of relative permittivity (many thousands) can be reached at T_C, leading to another use of ferroelectric materials – as dielectrics in high-value capacitors. The phase transition at T_C can either be continuous (known as second-order) or discontinuous (known as first-order). In either case, the dependence of permittivity (ε) on temperature above

T_C can be described by $\varepsilon = C/(T - T_0)$. T_0 is called the *Curie Weiss temperature*, and is only equal to T_C for a second-order phase transition. C is called the *Curie Weiss constant*. Figure 26.3 shows typical plots of P_s and ε as functions of temperature for ferroelectrics around T_C. Figs. 26.3a and 26.3b show the behaviour for a first-order phase transition. As can be seen, P_s drops discontinuously to zero at the phase-transition temperature. The permittivity rises as the temperature decreases, peaking at the transition, where there is a discontinuous step down. First-order phase transitions tend to show thermal hysteresis in the transitions, so that the transition occurs at a higher temperature when approached from the low-temperature side than from the high-temperature side. Frequently in practical observations there is actually a region of coexistence where both phases exist together in the same sample at the same temperature. In this case, the Curie temperature is defined as the temperature at which the ferroelectric and paraelectric phases have the same Gibbs free energies. In the case of a second-order phase transition (Figs. 26.3c and 26.3d), P_s drops continuously to zero at T_C and the permittivity rises to a sharp peak. There is no hysteresis in the transition.

There are two types of theoretical model that describe the ferroelectric phenomenon. The first is the phenomenological or thermodynamic theory first proposed by *Devonshire* [26.2]. Here, the Gibbs free energy of the ferroelectric system is described in terms of a power series in the spontaneous polarisation. In this case, P_s is termed the order parameter for the phase transition and the theory is similar in form to the generalised phenomenological theory of phase transitions known as Landau theory. It is not appropriate to go into the details of the theory here, which has been described very well by *Lines* and *Glass* [26.3]. The theory successfully predicts many of the behavioural characteristics of ferroelectric materials in terms of the measurable macroscopic properties, including the ferroelectric hysteresis and the behaviour of the permittivity near T_C. However, it tells us nothing about the microscopic origins of ferroelectricity. This is the province of the two types of microscopic models of ferroelectric behaviour. The first of these is the order–disorder model of ferroelectric behaviour. The second is the displacive model.

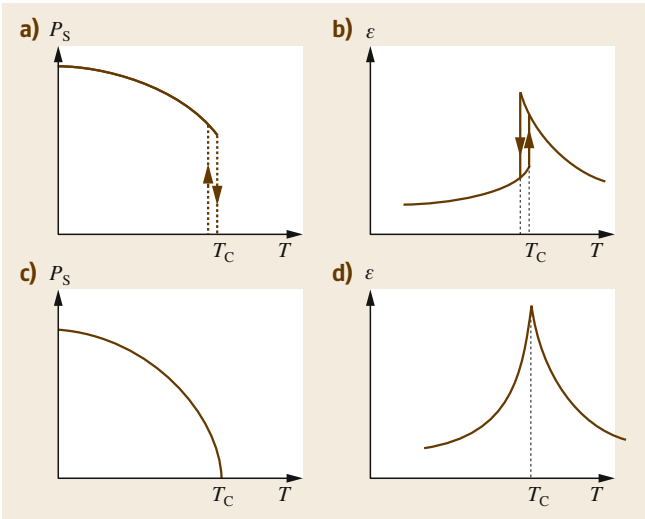


Fig. 26.3a–d The variations of spontaneous polarisation ((**a**) and (**c**)) and permittivity ((**b**) and (**d**)) with temperature for typical ferroelectric materials. (**a**) and (**b**) are for a first-order (discontinuous) phase transition while (**c**) and (**d**) are for a second-order (continuous) phase transition

According to the order–disorder model, the electric dipoles exist within the structure in the paraelectric phase above T_C, but are thermally disordered between two or more states so that the average polarisation is zero. There are several materials for which that is the case, an example being potassium dihydrogen phosphate (KDP), which is paraelectric, tetragonal ($\bar{4}2m$) above 123 K and ferroelectric, orthorhombic ($mm2$) below. The polarisation appears along the tetragonal c-axis. In this crystal structure, the PO_4 groups form tetrahedra that are linked by hydrogen bonds at their corners. The protons in these sit in double potential wells. Above T_C they are delocalised between the the two minima in the wells. Below T_C, they are localised and the PO_4 tetrahedra become distorted. The phosphorous and potassium ions become displaced relative to the oxygen framework, forming the polarisation in the lattice. Replacement of the hydrogen in the structure by deuterium raises T_C to about 220 K because the heavier deuteron delocalises between the two potential minima at a much higher temperature. Another example of an order–disorder ferroelectric is sodium nitrite ($NaNO_2$) in which the polar groups are NO_2^- ions, which become ordered below T_C (163 °C) with their dipoles all pointing along the b-axis of the crystal structure.

In displacive ferroelectrics, there are considered to be no dipoles in the structure above T_C, but the dipoles appear in the structure due to the cooperative motion of ions. Typically this is seen as the *softening* of a zone-centre optical phonon. What this means is that the frequency of an optical phonon mode with zero wavevector goes to zero at T_C. Such a phonon mode involves the displacement in opposite directions of cations and anions in the structure. When the frequency of this mode goes to zero, a dipolar displacement results. Examples of such ferroelectrics include the perovskite-structured oxides such as $BaTiO_3$, which will be discussed further below. One of the successes of soft-mode theory

is that it provides a convincing mechanism for the peak in the dielectric permittivity at T_C. According to the Lyddane–Sachs–Teller relationship (see, for example, [26.3] p. 216), the softening of a zone-centre phonon will make the permittivity diverge at T_C.

In practice, the distinction between order–disorder and displacive ferroelectrics becomes rather blurred, as some ferroelectrics which may be considered to be purely displacive may exhibit significant disordered cation displacement well above T_C, while some order–disorder ferroelectrics show soft-mode behaviour. For a more detailed discussion of these theories, the reader is referred to *Lines* and *Glass* [26.3].

26.2 Ferroelectric Materials

26.2.1 Ferroelectric Oxides

This is by far the most technologically important class of ferroelectric materials. They will be discussed according to the most important crystal classes.

Perovskite Ferroelectrics

These materials possess crystal structures isomorphous with the mineral perovskite ($CaTiO_3$). They all have the general chemical formula ABO_3, where A and B are cations. Typically the A cation will be around 1.2–1.6 Å in radius (similar to the oxygen ions) while the B cations will be around 0.6–0.7 Å in radius. The crystal structure is illustrated in Fig. 26.4. It consists of a network of corner-linked BO_6 octahedra, within which is enclosed the large A cation. Another way to look at the structure is as cubic-close-packed AO_3 layers, with the small B cations sitting in the octahedral sites between these close-packed layers. A pair of layers are shown in Fig. 26.5. The structure is a very tolerant one and will accommodate many different ions. Because of this, it is exhibited by a large number of oxides. The basic criteria for the structure to be stable is that the valencies of the ions should balance and that the ionic radii meet the *Goldschmidt criteria* [26.4]. A tolerance factor t is defined as

$$t = \frac{r_A + r_O}{\sqrt{2}(r_B + r_O)} ,$$

where r_X is the ionic radius of the X cation. The ideal cubic perovskite structure, where the ionstable are just touching each other, will possess $t = 1$. However, the structure will be stable with $0.85 < t < 1.05$. A list of ions that commonly form perovskites, together with their ionic radii, is given in Table 26.2. The closer t is to unity, the more likely the structure will be to be cubic. Conversely, perovskites which have values of $t < 1$ show distorted structures that are frequently ferroelectric. Examples of some perovskites and their tolerance factors are listed in Table 26.3. This table also lists the structures formed by these compounds at

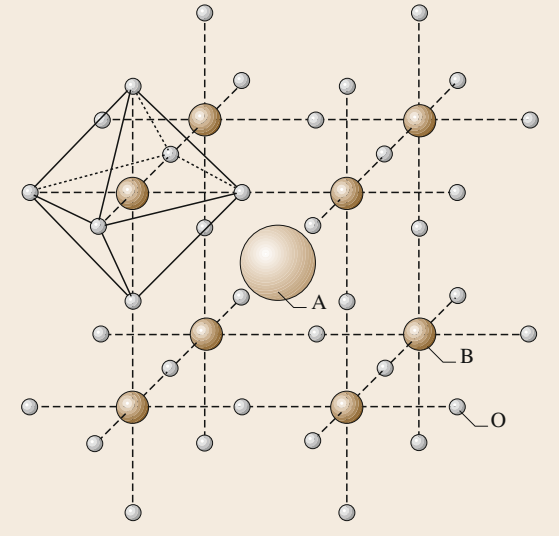


Fig. 26.4 The perovskite crystal structure

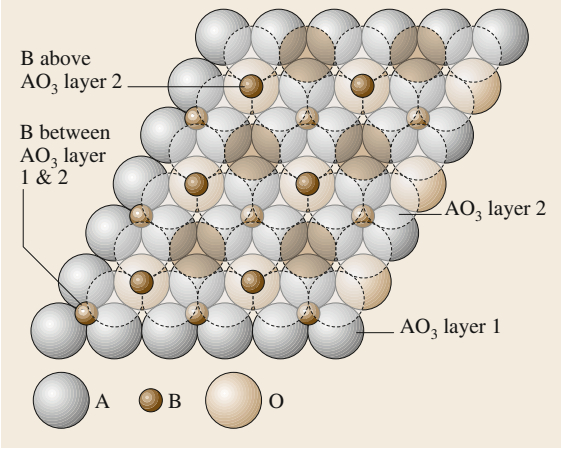


Fig. 26.5 Showing two AO_3 close-packed layers of the perovskite structure, with the B cations sitting in the six-fold coordinated sites between the layers. The second layer has been made semitransparent for clarity

Table 26.2 List of cations frequently found to form perovskite-structured oxides and their ionic radii. (Ionic radii taken from [26.5])

A-site cation	Ionic radius when [12][a] by O^{2-} (Å)	B-site cation	Ionic radius when [6][b] by O^{2-} (Å)
Na^{+}	1.32	Nb^{5+}	0.64
K^{+}	1.6	Ta^{5+}	0.68
Ba^{2+}	1.6	Zr^{4+}	0.72
Sr^{2+}	1.44	Ti^{4+}	0.605
Pb^{2+}	1.49	Pb^{4+}	0.775
Bi^{3+}	1.11	Sc^{3+}	0.73
Ca^{2+}	1.35	Fe^{3+}	0.645

[a] *12-fold coordinated*; [b] *6-fold coordinated*

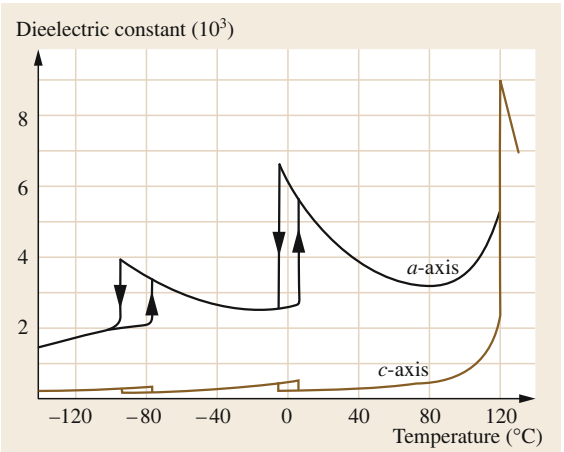


Fig. 26.6 Temperature dependence of the relative permittivity of $BaTiO_3$ measured along [001] and [100] (After [26.7])

room temperature, and whether or not they are ferroelectric. Some of the most interesting perovskites from the point of view of applications are $BaTiO_3$, $PbTiO_3$ and $KNbO_3$. $BaTiO_3$ is cubic above 135 °C, but transforms to a tetragonal ferroelectric structure below this temperature. In this case, the Ba and Ti ions are displaced relative to the anion framework along one of the cubic ⟨001⟩ directions. This means that the polar axis has six choices for direction in the tetragonal phase. At 5 °C there is a second-phase transition from the tetragonal to an orthorhombic phase, where the polarisation now appears due to cation displacements along one of the cubic ⟨110⟩ directions, for which there are 12 choices. Finally, at −90 °C there is a transition to a rhombohedral phase with the cations being displaced along one of the cubic ⟨111⟩ directions, for which there are eight choices. In the case of $PbTiO_3$ there is a single transition to a tetragonal phase at 490 °C, again by cationic displacements along ⟨100⟩. Some perovskites show phase transitions that are not ferroelectric. For example, $SrTiO_3$ shows a transition to a tetragonal phase at 110 K which involves linked rotations, or tilts, of the TiO_6 octahedra about the cubic [100] direction. Tilting of the octahedra is a common feature of the phase transitions that occur in perovskites and can lead to very complex series of phase transitions, as has been observed for $NaNbO_3$. This type of structural modification and a commonly-used notation for it has been described in detail by *Glazer* [26.6].

It is very easy to make solid solutions of the end-member perovskites, such as those listed in Table 26.3, and this has been used to great effect to provide materials with a wide range of properties. For example, Fig. 26.6 shows the temperature dependence of the relative permittivity of a $BaTiO_3$ single crystal [26.7]. There is a peak at each transition where the value perpendicular to the polar axis reaches several thousand, making the material interesting for use in capacitor dielectrics. However, the temperature variation in such a material would make it useless. The formation of ceramic solid solutions of $BaTiO_3$ with $SrTiO_3$, $CaTiO_3$ or $PbTiO_3$ allows the temperature dependence of permittivity to be controlled so that useful specifications can be met with average permittivities of 2000 or more over a wide range of temperatures. *Herbert* [26.8] has discussed the details of these modifications and their

Table 26.3 Some end-member perovskites and their properties

Perovskite oxide	Tolerance factor	Structure at 20 °C	Type	T_C (°C)
$BaTiO_3$	1.06	Tetragonal	Ferroelectric	135
$SrTiO_3$	1.00	Cubic	Paraelectric	
$CaTiO_3$	0.97	Tetragonal	Paraelectric	
$PbTiO_3$	1.02	Tetragonal	Ferroelectric	490
$PbZrO_3$	0.96	Orthorhombic	Antiferroelectric	235
$NaNbO_3$	0.94	Monoclinic	Ferroelectric	−200
$KNbO_3$	1.04	Tetragonal	Ferroelectric	412
$KTaO_3$	1.02	Cubic	Ferroelectric	−260
$BiScO_3$	0.83	Rhombohedral	Ferroelectric	370
$BiFeO_3$	0.87	Tetragonal	Ferroelectric	850

effects on the $BaTiO_3$ transition temperatures. $BaTiO_3$-based ceramics are widely used in ceramic capacitors and form the basis of an industry worth billions of dollars annually. They have also been used for piezoelectric ceramics, but are no longer so important in that field, as they have largely been displaced by ceramics in the $PbZrO_3$–$PbTiO_3$ system. Nevertheless, there is a renewed interest in $BaTiO_3$-based ceramics for lead-free piezoelectrics in response to the legislative drive to reduce lead in the environment.

The perovskite solid solution system between $PbZrO_3$ and $PbTiO_3$ is of great technological importance and is thus worth discussing in some detail. The phase diagram given in Fig. 26.7 shows several phases. Starting at $PbTiO_3$, the ferroelectric tetragonal (F_T) phase persists well across the diagram, until the composition $Pb(Zr_{0.53}Ti_{0.47})O_3$ is reached, where there is a transition to a ferroelectric, rhombohedral phase. The composition where this occurs is called the morphotropic phase boundary (MPB). The rhombohedral phase region splits into two. There is a high-temperature phase ($F_{R(HT)}$) in which the cations are displaced along the cubic [111] direction. There is also a low-temperature ferroelectric rhombohedral ($F_{R(LT)}$) phase in which the $(Zr, Ti)O_6$ octahedra are rotated about the [111] axis. This doubles the unit cell. Close to $PbZrO_3$ (Fig. 26.7b) the room-temperature structure transforms to the antiferroelectric, orthorhombic (A_O) structure. This is a complex structure in which the cations are displaced in double antiparallel rows along the cubic ⟨110⟩ directions, coupled with octahedral tilts. A higher-temperature A_T phase, which is frequently shown when this phase diagram is cited in the literature, has been shown not to exist in pure solid solutions [26.9]. There is another, more modern, modification to this phase diagram. It has been shown by *Noheda* et al. [26.10] that a monoclinic phase exists at the MPB (Fig. 26.7c). The enormous technological and commercial importance of this system derives from the very high piezoelectric, pyroelectric and electro-optic coefficients that can be obtained from ceramic compositions in different parts of the phase diagram, particularly when the base composition is doped with selected ions. These are discussed in more detail below. This solid solution system is frequently referred to generically as *PZT* although, strictly, PZT was the brand name for a set of piezoelectric ceramic compositions manufactured by the Clevite Corporation.

There is another very important class of oxides, termed complex perovskites, where the A or, more commonly B, sites in the structure are occupied by ions of different valency in a fixed molar ratio. These are not solid solutions as they possess a fixed composition. Examples are $PbMg_{1/3}Nb_{2/3}O_3$ (PMN), $PbSc_{1/2}Ta_{1/2}O_3$

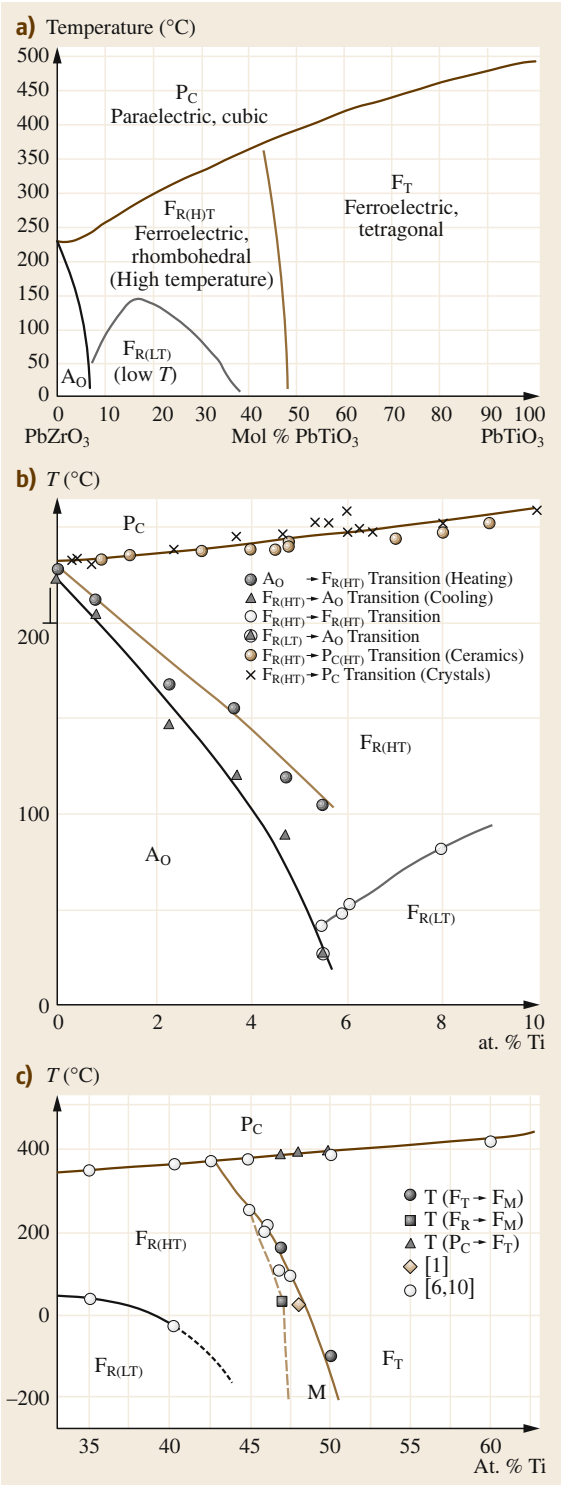


Fig. 26.7 **(a)** $PbZrO_3$–$PbTiO_3$ phase diagram. (After [26.11]) **(b)** Region close to $PbZrO_3$. (After [26.9, 12]) **(c)** Region close to the MPB. (After [26.10])

(PST) and $Bi_{1/2}Na_{1/2}TiO_3$. These are also technologically important and can form end members to solid solutions in their own right, $PbMg_{1/3}Nb_{2/3}O_3$–$PbTiO_3$ (PMN-PT), for example. Many of the complex ferroelectrics (PMN-PT is a classic example) exhibit *ferroelectric relaxor* behaviour. In these materials, the Curie point is no longer a sharp transition but is actually observed over a very wide range of temperatures. A broad permittivity peak is observed and the temperature and height of the peak are strongly dependent on the frequency of measurement. Figure 26.8 shows a typical variation of permittivity with temperature and frequency in PMN [26.13].

The broad tolerance of the perovskite structure to different cations has led to a wide exploration of the inclusion of different iso-valent dopants to obtain different electronic properties. It is also possible to include alio-valent dopants over a wide range, which the structure tolerates through the introduction of cation or anion vacancies, or free charge carriers. For example, it is possible to incorporate Nb^{5+} into the B site of the $BaTiO_3$ system [26.14], and charge balance is maintained through the presence of free electrons. Similarly, Fe^{3+} can be introduced into the B site of the system and this substitution creates O vacancies [26.15].

The vast majority of perovskites are employed in ceramic form, but thin- and thick-film materials are becoming of increasing importance and single-crystal materials are starting to emerge which have exceptionally high piezoelectric coefficients.

Illmenite Ferroelectrics

The illmenite structure is related to the perovskite structure in that it is exhibited by materials with the general formula ABO_3, where the A cation is too small to fill the [12] coordinated site of the perovskite structure. The structure is made up of hexagonal close-packed layers of oxygen ions, with the A and B ions occupying the octahedrally coordinated sites between the layers. Hence, this structure can also be considered to be related to the perovskite structure in that both are based on oxygen octahedra. The two best known illmenite ferroelectrics are $LiNbO_3$ and $LiTaO_3$, whose structure is shown in Fig. 26.9. The materials have high T_C values (ca. 1200 and 620 °C, respectively). In the case of $LiNbO_3$, the T_C is only 50 °C below the melting point. Referring to Fig. 26.9, it can be seen that the cations occupy octahedral sites progressing along the *c*-axis in the order Nb, vacancy, Li, Nb, vacancy, Li etc. The polarisation occurs through displacement of the cations along the three-fold axis, so the structure transforms from $\bar{3}m$ to $3m$ symmetry at T_C. These are *uniaxial* ferroelectrics, in that the polarisation can only be up or down, and only 180° domains exist. The materials are mainly used in

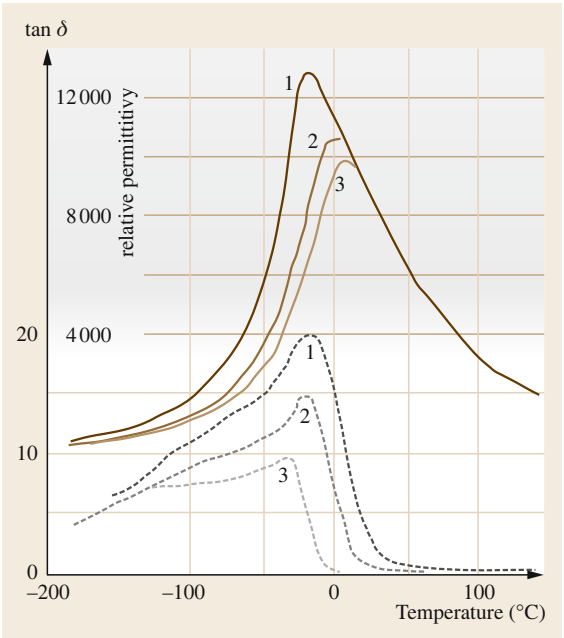


Fig. 26.8 Permittivity (*solid lines*) and dielectric loss (*dashed lines*) versus temperature in PMN. The measurements were made at: (1) 0.4, (2) 45 and (3) 4500 kHz. (Adapted from [26.13])

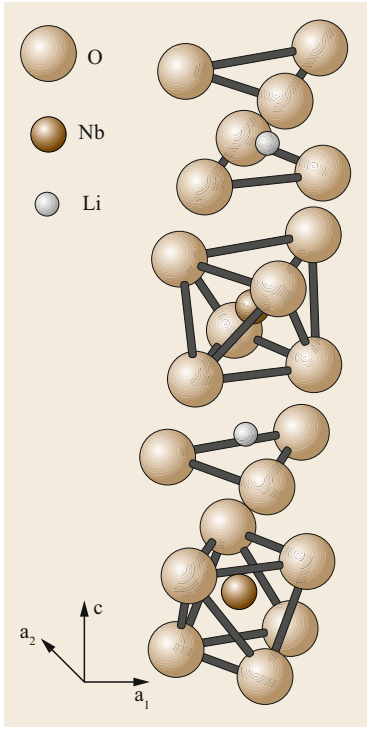


Fig. 26.9 Crystal structure of $LiNbO_3$. (After [26.16])

single-crystal form for piezoelectric and electro-optic devices.

Tungsten Bronze Ferroelectrics

This is another very large family of oxygen octahedral ferroelectrics possessing the general formula $[A1_2A2_4C_4][B1_2B2_8]O_{30}$. The crystal structure is complex, shown schematically in Fig. 26.10. The B1 and B2 sites are octahedrally coordinated by oxygens and have similar sizes and valencies to the B sites in the perovskites. The A1 and A2 sites are surrounded by four and five columns of BO_6 octahedra respectively. The three-fold coordinated C sites in the structure are frequently empty, but can be occupied by small uni- or divalent cations (e.g., Li^+ or Mg^{2+}). There are a wide range of ferroelectric tungsten bronzes, which frequently show nonstoichiometry. All are tetragonal in their paraelectric phase, and can transform to either a tetragonal ferroelectric phase, in which the polar axis appears along the tetrad axis of the paraelectric phase, or (more commonly) into an orthorhombic phase in which the polar axis appears perpendicular to the original tetrad axis. Examples of tungsten bronzes are $PbNb_2O_6$ (lead metaniobate) in which five out of the available six A sites are occupied by Pb^{2+} and the B sites by Nb^{5+}. This crystal and its Ta analogue are metastable below about 1200 °C. They can be stabilised by rapid cooling, or more commonly by doping. Lead metaniobate ceramics are difficult to make, but are commercially available and are occasionally used for piezoelectric devices by virtue of their higher Curie temperatures than members of the PZT system, combined with reduced lateral piezoelectric coupling factors. Substitution of Pb^{2+} by Ba^{2+} can also be used to produce a useful piezoelectric ceramic material $Pb_{1/2}Ba_{1/2}Nb_2O_6$. Other tungsten bronze ferroelectrics such as $Sr_xBa_{1-x}Nb_2O_6$, $Ba_2NaNb_5O_{15}$ and $K_3LiNb_5O_{15}$ have been investigated as single crystals for use in electro-optic devices. However, the growth of highly homogeneous single crystals is difficult because of the formation of optical striations caused by compositional fluctuations.

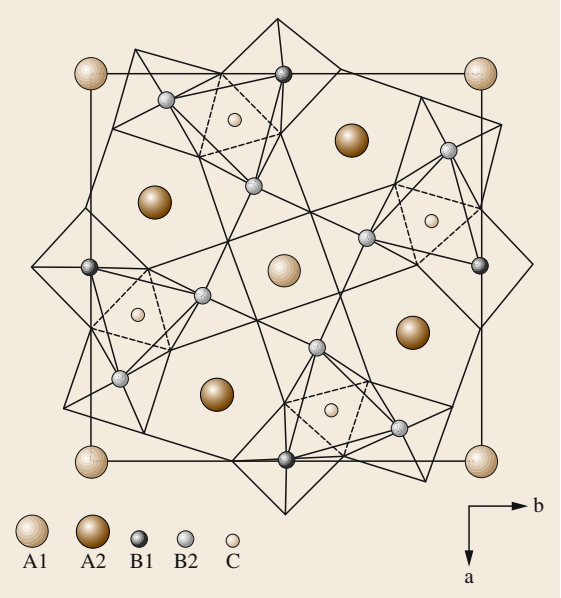


Fig. 26.10 Tungsten bronze crystal structure. (After [26.17])

Aurivillius Compounds

The Aurivillius compounds form another important class of ferroelectrics based on oxygen octahedra. They have a good deal in common with the perovskites in that they consist of layers or slabs of perovskite blocks with the general formula $(A_{m-1}B_mO_{3m+1})^{2-}$ separated by $(M_2O_2)^{2+}$ layers in which the M cation is in pyramidal coordination with four oxygens, the M being at the apex of the pyramid. The structure possesses the general formula $M_2(A_{m-1}B_mO_{3m+3})$. In the ferroelectric phases, M is usually Bi and m is usually between one and five. The A and B cations follow the usual ionic radius and valency criteria of the perovskites, and we see ferroelectric compounds in which A = Bi, La, Sr, Ba, Na, K etc. and B = Fe, Ti, Nb, Ta, etc. Part of the crystal structure is shown schematically in Fig. 26.11, illustrating in this case a structure with $m = 3$. This is just over half of the unit cell, showing one slab of perovskite units and two of the $(Bi_2O_2)^{2-}$ layers. The structures are tetragonal in their paraelectric phases, with the tetrad c-axis being perpendicular to the $(Bi_2O_2)^{2+}$ plane. In the majority of cases they become orthorhombic in their ferroelectric phase, with the spontaneous polarisation appearing in the $a-b$ plane, at 45° to the tetragonal a-axis. Bismuth titanate ($Bi_4Ti_3O_{12}$) is an example of an Aurivillius ferroelectric in which M = A = Bi, B = Ti and $m = 3$, so that there are three blocks in the perovskite slab. In this case, there is a small component of the spontaneous polarisation (ca. 4 $\mu C/cm^2$) out of the a–b plane, with the majority of the polarisation (ca. 50 $\mu C/cm^2$) being in the plane, so that the ferroelectric structure is monoclinic. The T_C is high (675 °C) and ceramics have been explored for use as high-temperature piezoelectrics. The problem with the compounds when used as ceramics is that the piezoelectric coefficients are rather small. Single crystals have been explored for optical light-valve applications but have not received widespread use because of the problems of growth. Thin films of $SrBi_2Ta_2O_9$ (SBT) are receiving considerable attention for applications in ferroelectric nonvolatile memories. Recently, thin films of

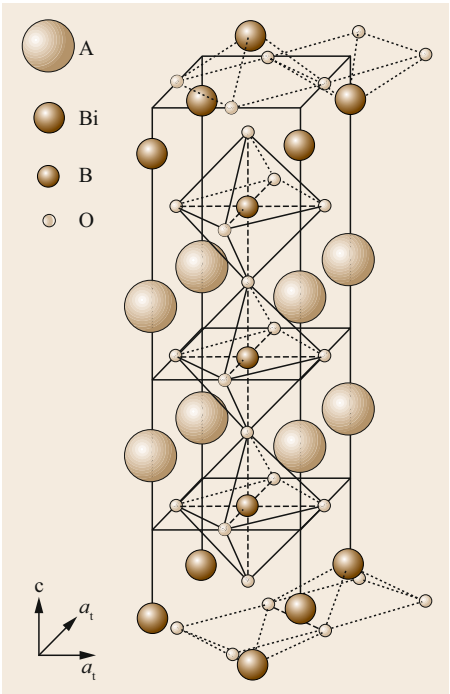


Fig. 26.11 Schematic diagram of half of the unit cell of the crystal structure of an Aurivillius compound with $m = 3$. The cell axes shown are those corresponding to the tetragonal structure

$Bi_6Ti_{2.8}Fe_{1.52}M_{10.68}O_{18}$ with $m = 5$ have been shown to be single phase magnetoelectric multiferroics [26.18].

Other Oxide Ferroelectrics

There is a wide range of other oxide ferroelectric materials, although none have anywhere near the breadth of application of the oxygen octahedral ferroelectrics discussed above. There are two groups of phosphates (in which the phosphorous ions are tetrahedrally coordinated by oxygens) which have found applications in optical systems. The potassium dihydrogen phosphate (KDP) family was referred to above as an example of an order–disorder ferroelectric. KDP, and its deuterated analogue KD*P, can be grown from aqueous solution as large, high-quality single crystals. These can be lapped and polished into plates which are used for longitudinal electro-optic modulators, in which the modulating electric field is applied parallel to the direction of light propagation in the crystal. In this application, the crystals are used at room temperature in the paraelectric phase. One spectacular application for these crystals is as the Q-switches in the large ultra-high-power lasers used for nuclear fusion research, where extremely large (ca. 1 m) crystal slices are needed. Another class of phosphate ferroelectrics is the potassium titanyl phosphate (KTP) system. These crystals can be grown relatively easily from the melt and are used for optical frequency doubling applications, particularly using periodically poled crystals [26.19, 20]. This is a large family in which the potassium can be replaced by caesium or rubidium and the phosphorous by arsenic. Lasers using KTP are finding significant uses in surgical procedures [26.21].

There are other examples of oxygen tetrahedral ferroelectrics. Lead germanate ($Pb_5Ge_3O_{11}$) is an interesting material which has been prepared as a single crystal. It is a hexagonal uniaxial ferroelectric which is also optically active. Ge can be substituted up to 50% by Si, which reduces the Curie temperature from 177 to 60 °C. The sense of the optical rotation switches with the sign of the spontaneous polarisation. This has been considered for optical devices and pyroelectric infra-red detectors, but has not reached any commercial applications. It has been prepared in ceramic and thin-film form and is occasionally used as a sintering aid in lead-containing ferroelectric ceramics. Gadolinium molybdate, $Gd_2(MoO_4)_3$ (GMO), possesses a structure consisting of corner-linked MoO_4 tetrahedra. It is an example of an improper ferroelectric material in which the phase transition from the paraelectric phase is controlled by the softening of a zone boundary mode, which would lead to a nonpolar low-temperature phase, which is in fact *ferroelastic*. The high-temperature phase is acentric, and the ferroelastic lattice distortion couples via the piezoelectric coefficients to give a spontaneous polarisation. The lattice distortion can be reversed by electrically switching the spontaneous polarisation and vice versa. GMO has no commercial applications, although it was at one time considered for optical switches.

26.2.2 Triglycine Sulphate (TGS)

The TGS family is a salt of the simplest amino acid, glycine (NH_2CH_2COOH) with sulphuric acid. In the crystal structure, the glycine groups are almost planar and pairs of them are connected by hydrogen bonds. A ferroelectric transition at 49.4 °C between two monoclinic phases is driven by an ordering of the protons in these hydrogen bonds. The replacement of these protons by deuterons (which can be accomplished by repeated crystallisation of TGS from heavy water) to give DTGS results in a 10 °C increase in T_C. There has been a great deal of academic interest in TGS because it possesses a well-behaved second-order order–disorder ferroelectric phase transition. In addition it has been of considerable technological interest for uncooled pyroelectric infra-red detectors and thermal imaging devices [26.22]. As a consequence, there has been considerable work on the replacement of various components with the objective of either improving performance or increasing T_C, such as can be achieved by deuteration. The replacement of glycine by l-alanine has been demonstrated to result in an internal bias field

in the crystal, which means that the crystals will retain their single-domain structure, even if heated well above T_C.

26.2.3 Polymeric Ferroelectrics

The first polymeric ferroelectric to be discovered was polyvinylidene fluoride (PVDF). The monomer for this possesses the formula CH_2CF_2. It was first shown to be piezoelectric by *Kawai* [26.23] and later *Bergman* et al. [26.24] speculated about the possibility of it being ferroelectric, although no Curie temperature could be shown to exist, as the polymer melts first. The structure of the polymer is complex, as the polymer backbone can adopt a number of configurations depending upon whether the neighbouring carbon-carbon linkages adopt trans or gauche configurations. In a trans (T) configuration, the groups bonded to the carbon atoms sit on opposite sides of the carbon-carbon bond. In a gauche (G) configuration, they sit on the same side. In an all-trans configuration, the carbon backbone of the polymer forms a simple zigzag. In the case of PVDF, there are four different phases with different T and G sequences, and with different ways the polymers stack together in crystal structures. The polymer crystallises from the melt into *form II*, also called the α-phase, in which the bonds arrange themselves into the sequence TG$\overline{\text{T}}\overline{\text{G}}$, see Fig. 26.12. The fluorine atoms are strongly electronegative, which makes the C−F bond polar so that the molecule has a net dipole moment perpendicular to its length. However, the molecules of polymer arrange themselves in the unit cell so that the dipoles cancel each other out. Form II is neither ferroelectric nor piezoelectric, but application of an electric field will convert it into form II_P, also called the δ-phase, in which the polymer molecules are arranged so that the unit cell has a net dipole moment. High-temperature annealing of either of these forms will produce form III (the γ-phase), which has a new TGTTT$\overline{\text{G}}$TT configuration which also has a net dipole moment perpendicular to the long axis of the molecule and these arrange in a crystal structure which is also polar. Subjecting forms II or III to stretching or drawing will produce form I (the β-phase), which is an all-trans configuration (Fig. 26.12). It can be seen from Fig. 26.12 that of the three molecular configurations, the polar bonds all point most-nearly in the same direction and this is retained in the unit cell. Electrical poling makes this the most-strongly piezoelectric phase of PVDF. The stretching

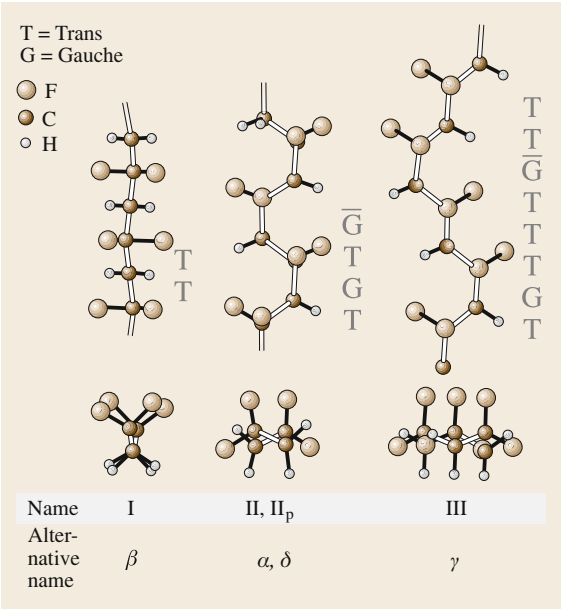


Fig. 26.12 PVDF bond structures

can be in a single direction, called uniaxial and usually achieved by drawing through rollers, or in two perpendicular directions, called biaxial and usually achieved by inflating a tube of polymer. Strong electric fields are needed for poling and these are frequently applied by placing the polymer under a corona discharge. The formation of a copolymer of vinylidene fluoride with between 10 and 46% trifluoroethylene (TrFE) leads to a polymer which will crystallise directly into form I (the β-phase) from either melt or solution. This can be poled to produce a material which is as active as pure PVDF. The copolymers also show clear signs of a ferroelectric-to-paraelectric transition, such as exhibiting peaks in permittivity at T_C, with a T_C that depends on the amount of TrFE in the copolymer. Extrapolation to 0% TrFE implies a T_C in PVDF of 196 °C. PVDF is readily available and has achieved moderately widespread use as a piezoelectric sensor material. It is particularly useful where light weight and flexibility are important, or where very large areas or long lengths are needed. The P(VDF-TrFE) copolymers are not as readily available and are not widely used. Other polymers that have been shown to exhibit ferroelectric behaviour include the odd-numbered nylons, but these are only weakly piezoelectric and have not achieved any technological uses.

26.3 Ferroelectric Materials Fabrication Technology

Ferroelectrics are used commercially in a very wide range of forms, from single crystals through polycrystalline ceramics and thin films to polymers. Hence, only a summary of the fabrication techniques can be presented here.

26.3.1 Single Crystals

As has been shown above, ferroelectrics often have a wide range of complex compositions. This renders the problems of single-crystal growth much more difficult than for single-element crystals such as silicon. In many cases, the most technologically useful compositions are themselves solid solutions of complex ferroelectric end members. Frequently these do not melt congruently and sometimes one or more components of the melt will be volatile. Some of the growth techniques which are used are:

Czochralski Growth

Certain ferroelectric crystals melt congruently and can be pulled from the melt using the Czochralski technique. Examples include $LiNbO_3$ and $LiTaO_3$. Both materials are widely used technologically. $LiNbO_3$ is used in surface acoustic wave (SAW) and electro-optic devices and $LiTaO_3$ in pyroelectric infrared detectors, piezoelectric resonators, SAW and electro-optic devices. $LiNbO_3$ melts congruently at 1240 °C, but the congruently melting composition is not at the stoichiometric composition (Li : Ta = 1), but lies at around 49% Li_2O. Crystals pulled from such melts in platinum crucibles are perfectly adequate for piezoelectric and most optical applications. However, it is known that the stoichiometric composition possesses rather better optical properties than the congruent composition and there is now a premium-grade material available commercially. $LiTaO_3$ melts congruently at 1650 °C. Again, the composition is not at Li : Ta = 1, but is slightly Ta deficient. The growth temperature is above the melting point for platinum. The solution is to use crucibles made from iridium, which is very expensive. Also, IrO_2 is volatile so the growth atmosphere must be N_2, which means that the crystals come out of the melt oxygen-deficient and must be embedded in $LiTaO_3$ powder and annealed under O_2 after growth to make them clear. Pt/Rh crucibles have been used to grow $LiTaO_3$ crystals, but the Rh absorbed into the crystals makes them brown. Such crystals are acceptable for piezoelectric and pyroelectric, but not for optical, applications. Virtually all ferroelectric materials need to be electrically poled before use. In the case of $LiTaO_3$, silver electrodes are applied to faces perpendicular to the polar (c) axis and a pulsed field of about 10 V/cm (1 ms with a 50/50 duty cycle) applied while the crystal is cooled slowly through T_C (620 °C). A similar process can be applied to pole $LiNbO_3$ crystals or, as the T_C is very close to the growth temperature, the field can be applied between the growing crystal and the crucible. This process has also been used to grow periodically poled crystals in which the field is reversed periodically during the growth of the crystal. Such crystals have some advantages in optical frequency-doubling in a process known as quasi-phase matching [26.25].

Solution and Flux Growth

The growth of crystals from aqueous solution is a technology very readily applied to water-soluble materials such as KDP and TGS. There are two processes that have been used. The first consists of rotating crystal seeds in a bath of saturated solution, which is slowly cooled so that material that comes out of solution is deposited on the seeds. The second consists of recirculating the bath solution over feed material, which is held at a slightly higher temperature and then back over the slightly cooler seeds, where the feed material is deposited. In both techniques, very careful control of temperature is essential (better than 0.002 °C) and the growth process can take many days.

The process of growing oxide single crystals from solution or flux is closely related to the above process, in that the crystals are grown from a saturated solution of the required oxide in a molten flux as it is slowly cooled. In the simplest version of this process, the oxides are sealed in a platinum crucible, heated to a *soak* temperature and slowly cooled to room temperature. Small crystals (usually up to a few mm in size) can then be recovered from the solidified melt. This has been applied to growth of a wide variety of small single crystals, such as PZT and PMN-PT solid solutions. PZT, for example, can be grown from $PbO-B_2O_3$ flux mixtures, or even pure PbO [26.26]. $BaTiO_3$ crystals can be grown from KF or TiO_2 fluxes. A major issue with the growth of Pb-containing ferroelectrics from PbO rich fluxes is the evaporation of PbO from the melt at the crystallisation temperature over the long time periods required to grow large crystals. This can be avoided to some degree by the use of sealed crucibles. A variation on the flux growth process is top-seeded solution growth, whereby a seed crystal is immersed from above into a supersaturated solution of the material to be grown. The crystal then grows as the solution is cooled further. The process is difficult to control and it has the major problem, especially when growing from a PbO flux, that the melt is open and therefore very prone to

flux evaporation. It has been successfully applied to the growth of crystals of PMN-PT and PZN-PT solid solutions. The Bridgman process seeks to get over this problem by placing a dense compact of the flux/solute mixture inside a sealed crucible with a conical end, as shown in Fig. 26.13. A seed crystal is placed at the base of the cone which is held cooler than the rest of the crucible by a cooled rod. A molten zone is then allowed to travel up the crucible and the crystal solidifies inside it. Good-quality crystals have been grown this way and it is the preferred technique for the manufacture of PMN-PT and PZN-PT solid solutions. One problem with all flux-growth techniques is that the flux composition is constantly changing during the growth process and therefore the crystals tend to be nonuniform in composition. This is a particular problem for larger crystals, especially for solid solutions.

26.3.2 Ceramics

The vast majority of ferroelectric materials that are used commercially are used in the form of polycrystalline ceramics. These are used in many of the applications listed above, including as dielectrics and in piezoelectric, pyroelectric, PTCR and electro-optic devices. Many different ferroelectric ceramic fabrication technologies have been developed over the years, but they are all based around a similar fabrication sequence. The key points are discussed below. Note that as the vast majority of commercially exploited ferroelectrics are oxides, the discussion will centre around these materials:

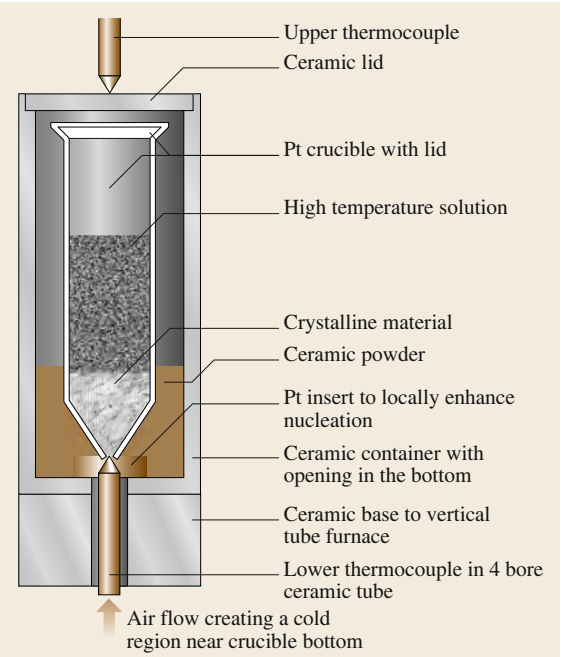


Fig. 26.13 A modified Bridgman apparatus used for growing piezoelectric single crystals. (After [26.27])

- Raw-material selection: The usual starting point for a ceramic fabrication sequence is the selection of the raw materials. For oxide ferroelectrics, the raw materials are usually oxides or carbonates, but occasionally other compounds (e.g nitrates, citrates etc.) are used which will decompose when heated to form the oxides required. High purity (usually $> 99.9\%$ with respect to unwanted cations) is important for good reproducibility. Small quantities of dopants can have a major effect on the final electrical properties, as would be expected for any electronic material, but can also seriously affect the sintering characteristics and grain size of the final ceramic body. Note that high purity does not usually imply the kind of purity that would be needed for a semiconductor material. Indeed, ultra-high purity ($> 99.99\%$) is often obtained by raw-material manufacturers by applying processes which negatively affect the reactivity of the oxide powders. A high degree of solid-state reactivity in the powders is important for the processes which follow. This is usually determined by a combination of raw-material particle size and specific surface area, although selecting the right crystallographic phase can be important. When using TiO_2, for example, it is usually found that the anatase phase is more reactive than rutile. These factors can be analysed for quality control purposes by using laser particle size analysis (taking care to break up loosely bound agglomerates by ultrasonic dispersion in water, with a dispersing agent prior to measurement), BET (Brunauer, Emmett and Teller)-specific gas absorption and x-ray powder diffraction respectively. The majority of ceramics are made by mixing together powdered raw materials, but occasionally solution mixing techniques are used and these will be discussed further below.
- Raw-material mixing: This is usually done in a ball mill consisting of a cylinder containing a mixture of small balls or cylinders made of steel, steatite, zircon or yttria-stabilised zirconia (YSZ). The oxide raw materials are accurately weighed into the ball mill, together with a predetermined amount of a milling fluid (usually deionised water, but occasionally an organic fluid such as acetone). It is usual to add a dispersant to aid the breakup of agglomerates. The ball mill is then sealed and rotated to mix the ingredients. Precise milling conditions are determined by the materials being used, but

the mixing time is usually of the order of a few hours. It is advantageous to use milling balls that are made of a material whose wear products will be reasonably innocuous in the final ceramic. For this reason, YSZ balls are preferred over steel or steatite, as small amounts of iron and silica contaminants can have unwanted or deleterious effects on the properties of many ferroelectric ceramics. For similar reasons, the ball mills are frequently rubber lined. Many variations on this process have been explored. High-energy ball milling is receiving considerable attention. In this process, the milling balls are given very high energy by either aggressively vibrating the ball mill or by stirring them at high speed with a paddle. Raw materials trapped between the balls are both comminuted and, if the energy is sufficiently high, can be forced to react together. Crystalline raw materials can be made amorphous in this process. A significant amount of energy can be stored in the powders after this process so that subsequent sintering can be undertaken at lower temperatures. In an adaptation of this process, it can be done in a continuous flow-through mill, whereby the slurry is pumped through the high-energy mill, making it less of a batch process.

- Drying: The slurry that results from the mixing process is dried after the milling balls are removed. On a small scale, this can be done in an oven or by freeze drying, but on a commercial scale, this is usually achieved by spray drying.
- Calcination: The purpose of this process is to react the raw materials into the required crystallographic phase. The dried powders are placed into a sealed crucible (usually high-purity alumina or zirconia), which is baked in a furnace at a temperature high enough to decompose any non-oxide precursors and cause a solid-state reaction between the raw materials, but not so high as to sinter the particles and form hard agglomerates that will be difficult to break up in subsequent processing. A simple example is the reaction between $BaCO_3$ and TiO_2 to form $BaTiO_3$

$$BaCO_3 + TiO_2 \rightarrow BaTiO_3 + CO_2 \uparrow \ .$$

This reaction will go to completion at about 600 °C. Note that a much higher temperature (> 1000 °C) is required for the decomposition of $BaCO_3$ to BaO in the absence of TiO_2.

- Milling: This is usually done in a similar manner to the mixing process described above, but is done for longer. The objective is to break up hard agglomerates and reduce the powder to its primary particle size. Again, high-energy milling techniques can be used. Dispersants are usually added to aid the process, and pH is controlled so that the powder does not flocculate. Organic binders can be added at this stage. The slurry from the milling process is usually spray-dried or freeze-dried as a free-flowing aggregated powder is required. The objective is to form soft aggregates that will break up on subsequent die pressing. Alternatively, the slurry (or *slip*) can be taken straight to a slip- or tape-casting process.
- Shape forming: After drying, the powders from the above process can be uniaxially pressed into the *green* shape required, typically using a steel punch and die-set. It is important in this process to make sure that the force is applied equally to top and bottom punches. Usually, the die is designed with a slight taper to ease removal of the workpiece after pressing. Faults that can occur at this stage of the process include *capping* or radial cracks in the work piece, which usually result from an uneven distribution of pressure. One way to avoid this is to use cold isostatic pressing. In this process the powder, or more frequently a lightly uniaxially cold-pressed green block, is vacuum-sealed into a rubber mould, which is then immersed in a bath of oil. The oil is taken up to a high pressure, which isostatically compresses the block. The advantage of using this process is that it avoids the nonuniform distribution of pressure that is frequently a problem with uniaxially cold pressing. The high pressure allows the agglomerates to be broken up and the process usually results in a much higher green density, which facilitates the sintering process. Anther process that can be used to form the green item is slip-casting, which involves pouring the ceramic slip into a plaster mould, which is usually gently rotated. The plaster absorbs moisture from the slip and the ceramic particles are deposited as a layer. Surplus slip is poured off and the deposited layer can be separated from the mould after it is dry. This process has been used to make large cylinders of piezoelectric ceramics. A third process that is frequently used to make thin sheets of ceramic is tape-casting, by which the slip is passed under a set of *doctor blades* in the apparatus shown schematically in Fig. 26.14. In this case, the slip is cast onto a continuous roll of plastic tape and a plastic binder is added to the slip which, when dry, holds the powder together and makes a flexible green tape. Sometimes the slip is cast directly onto a glass sheet. The green tapes can be screen-printed with patterns of metallic electrode inks and then laminated by warm-pressing to make multilayer structures with interleaved electrodes. This type of fabrication is widely used for multilayer ceramic (MLC) capacitors and multilayer ceramic piezoelectric actuators.

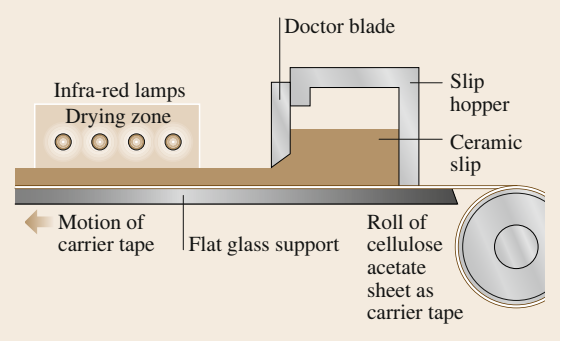


Fig. 26.14 Tape-casting apparatus

- Sintering: The green ceramic bodies are sintered in furnaces at temperatures of 1150–1300 °C, depending on the ceramic being manufactured. There is a good deal of process know-how in the sintering of the ceramic bodies. Key issues are:
  - The heating profile, which must be carefully controlled, especially over the lower range of temperatures up to about 500 °C where the organic matter (dispersants and binder) is burned off. Too rapid heating will cause large pores, or even cracks, to form.
  - For lead-containing ceramics, it is important that air be allowed free access to the vaporising organic material, or it will be carbonised and this can lead to reduction of the PbO in the ceramic to free Pb.
  - Loss of PbO by volatilisation at higher temperatures ($>\approx 800$ °C) means that the ceramic body needs to be sintered in a PbO-rich environment, which is often produced by packing the ceramic body in a *spacer powder* of $PbZrO_3$, or PZT ceramic chips within well-sealed ceramic crucibles or saggers. This tends to restrict the access of air to the sintering ceramic body, which conflicts with the previous requirement.
  - The ceramic body must be free to move during the sintering process, as there is significant linear shrinkage (about 15 to 18%). If the body is unable to slide over the surface on which it sits, this can lead to cracking, especially for large components. One solution to this problem is to sit the body on zirconia sand.

  The sintering process has been very well described elsewhere. It is possible to obtain very high densities ($> 98\%$) by careful control of the sintering conditions. For Pb-containing ferroelectric ceramics, a small excess of PbO is usually added to compensate for PbO loss by evaporation. This also tends to act as a liquid-phase-sintering aid, lubricating the grains of ceramic as they slide over each other in the sintering process, and providing a surface-tension force that pulls the ceramic grains together. The PbO also acts as a solvent, aiding the movement of the ceramic components during sintering and further densifying the ceramic. It is possible to sinter such materials to transparency, which means that there is virtually zero porosity. Very high densities can also be obtained by hot-pressing, in which the green body is placed inside a ceramic punch and die-set and raised to the sintering temperature under a pressure of about 35 MPa. Alternatively, hot isostatic pressing (HIP) can be used. In this process the pressure transmitting medium is a high-pressure gas (usually argon, but if the ceramic is Pb-containing it is important to include a few percent of oxygen to prevent reduction of the ceramic). In this case pressures of 100 MPa or more can be used. If the ceramic is pre-sintered so that there is no open porosity, then this can be achieved without the need for any container, but if there is open porosity, the body must be encapsulated in a suitable metal container, which will usually need to be a noble metal such as Pt.
- Electroding: Good-quality electroding is essential for all ferroelectric devices for a variety of reasons. Any low-permittivity layer between the electrode and the ferroelectric material will manifest itself in a fall-off in capacitance and an increase in loss as the frequency of measurement is increased. Poor-quality electrodes, or even the wrong type of conductive material can lead to problems with device ageing, or even an inability for the device to function as intended. Most piezoelectric materials are supplied with a *fired-on* silver electrode, which is a mixture of silver powder, a finely divided glass frit and a fluxing agent. Obtaining the correct balance between metal and glass contents is important, as too high a proportion of metal will lead to poor electrode adhesion while too high a proportion of glass will lead to poor electrode conductivity. Such electrodes are solderable with the use of appropriate fluxes. Sputtered or evaporated metal electrodes such as Ni or Cr/Au can also be used, as can metal electrodes deposited by electroless processes (e.g., Ni). The ohmic nature of the electrode contact to the ceramic is not usually important for highly insulating materials, such as piezoelectrics, but is important for semiconducting ceramics such as PTCR $BaTiO_3$. In this case it is usual to use Ni electrodes, which need to be annealed after deposition to develop the ohmic contact. In some cases it is necessary to fire on the electrode at the same time as ceramic sintering (cofired electrodes). This is particularly important

where the electrodes are buried in the structure, as with MLC capacitors and actuators. Clearly, there are problems to solve here in terms of the potential oxidation or melting of the electrode material. The conventional materials to use in such electrodes have been noble metals such as palladium (sometimes alloyed with silver) or platinum, which are very expensive. There have been serious efforts to develop ferroelectric compositions that can be fired with base metal electrodes such as Ni. In the case of $BaTiO_3$-based dielectrics, this has entailed the development of heavily acceptor-doped compositions that can stand being fired in a neutral or slightly reducing atmosphere. In the case of MLC actuators, there has recently been some success in developing piezoelectric PZT compositions that can stand being fired in such atmospheres, using Cu as the electrode material.

- Poling: Many devices made from ferroelectric ceramics (all piezoelectric and pyroelectric devices) require poling before they will develop useful properties. This entails applying a field that is significantly in excess of the coercive field (typically 3×), usually at an elevated temperature. It is not usually necessary to exceed the Curie temperature. For example, PZT ceramics with T_C in the range 230–350 °C can be poled by applying 3–5 kV/cm (depending on the type of PZT – see below – soft PZTs need lower poling fields than hard PZTs) at about 150 °C, with the field kept applied while the workpiece is cooled to room temperature. It is usual to immerse the ceramic in a heated bath of oil (mineral or silicone) during the process. One disadvantage of this is that the ceramic then needs to be carefully cleaned after poling. Silicone oil can be very hard to remove completely and its presence as a residue will compromise electrode solderability. For this reason, some workers have developed a process whereby the ceramic is poled under SF_6 gas. There is a rapid decay of properties after poling. This decay stabilises after a few hours, so it usual to wait at least 24 hours before electrical properties are measured.

There are many variations on the above basic process route which have been researched. One of these is the use of solution techniques to prepare the oxide powders. The basic principle here is that if the cations are mixed in solution, then they will be mixed on the interatomic scale without the need for milling processes that can introduce impurities. (This is a matter of discussion, as the act of precipitation and decomposition can lead to separation of the components.) Many routes have been explored, including the use of inorganic precursors such as nitrates and metalorganic precursors such as oxalates, citrates or alkoxides and acetates. Some of these have achieved a degree of commercial success, although the use of solution routes is more complex and the raw materials much more expensive than the mixed oxide routes. The use of mixed barium titanium oxalates has been very successful in producing high-quality fine-grained barium titanate powder for use in the capacitor industry. The use of metal citrates (frequently called the Pechini process) has been successfully used on a research scale to prepare many different types of ferroelectric oxide, but this type of process has not been applied on a commercial scale. Metal alkoxides, such as titanium isopropoxide, are readily soluble in alcohols and will react quickly with water to precipitate a hydroxide gel. Workers have used mixtures of titanium and zirconium alkoxides with lead and lanthanum acetates to coprecipitate a mixed hydroxide gel that could be calcined and sintered to make transparent lead lanthanum zirconate titanate ceramics for electro-optical applications.

26.3.3 Thick Films

There has been considerable interest in the integration of thick (10–50 μm thick) films of ferroelectric materials with alumina and other types of substrates such as silicon, to complement the wide range of other thick-film processes that are available, covering conductors, dielectrics, magnetic materials etc. There are many potential advantages to thick film processing for making certain types of sensor, especially the ability to use screen-printing for depositing the patterns of the materials required. Screen-printing involves using a sheet of wire mesh (the *screen*) that is coated with a photosensitive polymer. Exposure and development of the polymer allows selected areas to be removed, opening regions through which a paste of the required material can be pushed using a rubber blade or squeegee. The principle is simple, but there is a considerable amount of know-how in the formulation of the paste, which consists of the active material (in this case a ferroelectric powder such as PZT), an organic vehicle (a mixture of a solvent and polymer) and a glass frit. The screen is stretched over a former, and held close to, but not in contact with, the surface onto which the print is required. The paste is placed on the screen, and then spread over the screen with the squeegee, which prints the paste onto the substrate. Successive layers of different materials can be printed and cofired, provided there is good compatibility between them. The process has been well developed for piezoelectric films and adequate properties have been obtained from the films, although they are still well below the values that could be expected

from a bulk ceramic material. (See review by *Dorey* and *Whatmore* [26.28] for further details.)

26.3.4 Thin Films

The integration of high-quality thin films ($< 0.1-5\,\mu m$ thick) of ferroelectric materials onto substrates such as silicon has excited considerable interest for potential applications ranging from nonvolatile information storage to their use as active materials in micro-electromechanical systems (MEMS), where they can potentially be used for microsensors and actuators. Almost all the interest has been in the use of oxide ferroelectrics, but there has been some interest in the use of P(VDF-TrFE) copolymer films. These can be spun onto electroded substrates from methyl ethyl ketone solution. They are dried at relatively low temperatures ($< 100\,^{\circ}C$) and crystallised by annealing at $180\,^{\circ}C$ for several hours [26.29]. Such films have been applied to pyroelectric devices. However, the activity coefficients which can be obtained from such films are much lower than those that can be obtained from oxide materials. A range of deposition techniques have been developed for growing ferroelectric oxide films, which are summarised below:

- Chemical solution deposition (CSD): This term is applied to a wide range of processes that involve taking the metal ions into metalorganic solution, which is then deposited on the substrate by spinning, followed by drying and annealing processes to remove the volatile and organic components and convert the layer into a crystalline oxide. There are two broad classes of CSD process: metalorganic deposition (MOD) and sol gel. The MOD processes usually involve dissolving metal complexes with long-chain carboxylic acids in relatively heavy solvents such as toluene. The carbon content of the precursors is quite high, so there is a good deal of thickness shrinkage during firing. MOD solutions tend to be quite stable with time and resistant to hydrolysis. Sol-gel processes use precursors such as metal alkoxides, acetates and β-diketonates in alcohol solution. (For example, a set of precursors to deposit PZT would be Ti isopropoxide, Zr n-propoxide and lead acetate). Alkoxide precursors are very susceptible to hydrolysis, and so careful control of moisture content during the sol synthesis is essential and stabilisers such as ethylene glycol are usually added to the solutions to extend the useable lifetimes of the sols. Whereas MOD solutions are true solutions, the sols are actually stable dispersions of metal oxide/organic ligand particles with a size of about 4–6 nm. Sols possess lower viscosities and tend to produce the oxide layer at a somewhat lower temperature than the MOD processes, but the individual layer thicknesses produced by a single spin tend to be lower. Single crack-free layers tend to be in the range 100–200 nm thick. CSD processes have the advantages that they are low cost, the composition can be easily changed and they produce very smooth layers. The processes are planarising and will not follow underlying surface topology, which can be a disadvantage. Also, the processes are not industry standard in that they are wet and tend to have many variations.
- Metalorganic chemical vapour deposition (MOCVD): This is a variation on the process that has been very successfully applied to the growth of group III–V semiconductor layers. The principle is simple: volatile metalorganic compounds are passed over a heated substrate, where they decompose to form a layer of the desired compound. The problem with the growth of ferroelectric oxides is that most of the available metal organic precursors are relatively nonvolatile at room temperature. There has, therefore, been a great deal of research into the available precursors for the compounds that are required. Metal alkyls (such as lead tetraethyl) are very volatile, but only available for relatively few of the metal ions of interest (Pb being the main one). They are also pyrophoric and highly toxic. Some metal alkoxides, such as Ti isopropoxide, are suitable MOCVD precursors. Metal β-diketonates and related compounds such as tetramethyl heptane dionates (THDs) have received considerable attention as Ba and Sr precursors. All of these precursors need to be heated to give them suitable volatility and this means that the lines connecting the precursor source to the growth chamber need to be heated as well. Some of the precursors (especially THDs) are solids, which means that they are not really suitable for use in conventional bubbler-type sources. There has been considerable success in using solutions of these compounds in tetrahydrofuran (THF). The solutions are sprayed into a vaporiser that consists of a cylinder, containing wire wool or ball bearings, heated to a temperature at which the solution will flash-evaporate. A carrier gas is passed through the cylinder and this carries the precursor vapour into the growth chamber. The growth chamber is usually held at reduced pressure and a certain amount of oxygen is introduced to aid the oxide deposition. Frequently a radio-frequency (RF) or microwave plasma is also introduced to aid the growth of a high density film and reduce the required substrate temperature. The major problem

with the MOCVD process for complex ferroelectric oxides which have many cation components is finding the right combination of precursors that will all decompose at the same substrate temperature (usually ca. 550–650 °C) at a rate that will give the desired composition in the film. The process has been very successful in growing thin films of materials such as $(Ba, Sr)TiO_3$, with potential applications in the dynamic random-access memory (DRAM). The major advantage of MOCVD is that it is a truly conformal growth technique, with major advantages for semiconductor devices with complex topologies, but is a very expensive technique to set up because of the complex growth and control systems needed. Also, precursor availability is still a problem for many systems.

- Sputtering: A range of sputtering processes have been applied to the growth of ferroelectric thin films, including RF magnetron sputtering, direct-current (DC) sputtering and dual ion-beam sputtering. The RF magnetron process is probably the most popular. With all the processes, the major problem is one of obtaining the correct balance of cations in the growing film. Many different solutions have been found to this problem. In reactive sputtering, a composite metal target can be used. This can be made of segments of the metals to be sputtered (for example, Pb, Zr and Ti for PZT) and their relative areas changed to obtain the right composition in the film. Alternatively, multiple targets can be used and the substrate exposed to each one for different lengths of time, or the power applied to each one can be varied. In reactive sputtering, it is necessary to have an amount of oxygen in the sputtering gas (usually Ar). It is possible to sputter ferroelectric thin films from ceramic or mixed powder targets, but it is necessary to adjust the target composition to allow for different yields for each element. In any sputtering process there are many variables to adjust to optimise the process, including the sputtering power, and RF or DC substrate bias, which will affect the ion bombardment of the growing film, the sputtering atmosphere pressure and gas mixture and the substrate temperature. All of these can affect the film growth rate, composition, crystallite size and crystalline phases that are deposited, and the stress in the growing film. For this reason, the development of a sputtering process for a complex ferroelectric oxide can be a time-consuming business, and once a set of conditions has been arrived at for one particular composition, it cannot quickly be changed to accommodate new compositions. Dual ion-beam sputtering differs from the RF and DC processes in that a much lower background pressure is used, the material is sputtered from the target using an ion beam and a second lower-energy ion beam is used to stimulate and densify the growing film. The sputtering processes have the advantage of being well accepted industrially, as they are dry and can readily coat large-area substrates.
- Laser ablation: This process involves bombarding a ceramic target with a pulsed, focussed laser beam, usually from an ArF excimer source. The target is held under vacuum. A plasma plume is produced and the products ablated from the target are allowed to fall on a heated substrate. The advantages of the process are that there is usually good correspondence between the composition of the target and the growing film. Relatively small ceramic targets are acceptable for the process, and it is thus a good method for getting a rapid assessment of the properties of thin films of a given material. The disadvantages of the process are that the plasma plume will only coat a relatively small area of substrate, although there are now systems which use substrate translation to coat large areas, and particles can be ablated from the target, causing defects in the growing film.

In all the techniques used for the growth of ferroelectric oxide thin films, the key issues are control of composition and the formation of the desired crystalline phase (usually perovskite) with the desired crystallinity (crystallite size, morphology and orientation). All of the ferroelectric perovskites have a tendency to crystallise into a non-ferroelectric fluorite-like pyrochlore phase at low temperatures. In the case of the CSD processes, this means that as the film is heated from room temperature, after it loses the organic components, it first forms an amorphous oxide which then crystallises into a nanocrystalline pyrochlore phase, finally forming the desired perovskite phase. The temperatures at which this will occur depend very much on the ferroelectric oxide that is being grown and the precise composition. In the case of PZT, the pyrochlore phase will form above about 300–350 °C. The perovskite phase will start to form above about 420 °C, depending upon the ratio of Zr : Ti in the solid solution. The compositions close to $PbTiO_3$ will crystallise into perovskite much more readily than those close to $PbZrO_3$. In the case of a complex perovskite, such as $Pb(Mg_{1/3}Nb_{2/3})O_3$ or $PbSc_{1/2}Ta_{1/2}O_3$, the pyrochlore phase is much more stable and much higher temperatures (> 550 °C) are needed to convert it to perovskite. Excesses of PbO will tend to favour the formation of perovskite, and deficiencies favour pyrochlore. Higher annealing temperatures will promote PbO loss and it is

possible to get into a position, through PbO loss, where the pyrochlore becomes the most stable phase, even at high annealing temperatures. Residual pyrochlore phase invariably compromises the electrical properties of the films through reduced permittivity and piezoelectric/pyroelectric coefficients. The other growth techniques have the advantage that the films can be deposited onto heated substrates, at temperatures where they will grow directly into the perovskite phase (at least in principle), although there are many examples in the literature of films being deposited (e.g., by sputtering) at low substrate temperatures and converted to the desired perovskite phase by post-deposition annealing, in which case the same problems of pyrochlore formation apply. The control of film crystallinity (crystallite orientation and size) is important as it has a direct effect on the electrical properties. This is usually achieved through control of the crystallite nucleation. Sputtered Pt is frequently used as a substrate onto which ferroelectric thin films are grown. Like many metals, this will naturally grow with a (111) preferred orientation. It is face-centred cubic (fcc), with a lattice parameter of about 3.92 Å, which matches the lattice parameters of many of the ferroelectric perovskites, which are about 4 Å. This means that, with appropriate process control, it is quite possible to get a highly orientated (111) ferroelectric film on Pt, with a crystallite size of about 100 nm. Changing the underlying nucleation layer can allow other orientations to be grown. For example, the use of thin films of TiO_2 or PbO on top of the Pt electrode can induce a (100) orientation. It is important to control the nucleation density of the perovskite phase. If this is allowed to become too low, than large circular grains several microns in diameter (called *rosettes*) can form, which tend to induce defects at their boundaries. (Further details on thin-film ferroelectric growth techniques can be found in the book by *Pas de Araujo* et al. [26.30]).

26.4 Ferroelectric Applications

26.4.1 Dielectrics

As noted above, $BaTiO_3$ possesses a very high permittivity close to T_C. The inclusion of selected dopants can reduce T_C and optimise the properties of the material for capacitor applications. The substitution of Sr^{2+} for Ba^{2+}, for example, will reduce it, so that at about 15% substitution it will occur at about 20 °C. Substitution of Zr^{4+} for Ti^{4+} has a similar effect. The use of off-valent substitutions will have an effect upon resistivity and degradation characteristics, so that substituting La^{3+} on the A site will reduce resistivity at low concentrations. Substituting Nb^{5+} on the B site at the 5% level has been shown to confer resistance to degradation. The addition of Mn to the lattice has been shown to have a positive effect on dielectric loss. Control of grain size has also been shown to have a marked effect on dielectric properties. Reducing the grain size has the effect of increasing the concentration of domain walls per unit volume of ceramic, and thus increasing the domain-wall contribution to the permittivity. A reduction in grain size also brings about an increase in the unrelieved stress on the grains, which further increases the permittivity. Once the grain size falls below about 0.5 μm, the stress on the grains reduces their tetragonality and the permittivity falls. It is also possible to use heterogeneity in the ceramic grains to flatten the curve of dielectric constant versus temperature. A range of Electronic Industries Alliance (EIA) codes have been introduced to define the variation of capacitance with temperature. The X7 temperature range, for example, is −55 to +125 °C, while following this with the letter R would specify a capacitance change of no more than ±15%. Other dielectrics have been developed based on PMN that have higher peak permittivities, but worse temperature characteristics, capable of meeting a Z5U specification (+22 to −56% capacitance variation over the range of 10−85 °C) but not much better. There has been huge progress in MLC technology, with capacitors now available with > 100 layers of submicron thickness and capacitances ranging from a few hundred pF to 100 μF. The majority of these now have base-metal (Ni) electrodes. The fabrication of the ceramic dielectrics for these is a very complex area and a great deal of technology has been developed to service a very large MLC capacitor market (615 Bn units in 2002, worth an estimated $ 8.4 Bn).

26.4.2 Computer Memories

The storage of digital information is of great interest and growing technological importance. The ability of ferroelectrics to store information via the sense of the spontaneous polarisation has made them candidates for this application ever since the advent of the electronic computer. However, it is only with the development of the techniques for the growth of ferroelectric thin films onto silicon at relatively low temperatures (see above) that their use has become a reality. There are two ways in which ferroelectrics can be used in com-

puter memories. The first is the replacement of the dielectric layer in dynamic random-access memories (DRAMs). Here, the motivation is to exploit the high permittivities exhibited by ferroelectric oxides to reduce the area of silicon required to store a single bit of information. The favoured materials here are based on (Ba, Sr)TiO_3, while the favoured technique for thin-film growth is plasma-enhanced MOCVD because of its ability to deposit conformal coatings. No DRAM devices using these thin films have yet reached the market place. The use of the switchable polarisation for the nonvolatile storage of information has received a great deal of research. A good deal of this work has been based upon the use of PZT thin films, grown by a variety of techniques, including CSD, RF sputtering and MOCVD. Initially, there were many problems to solve, including the decay of the switchable polarisation (ca. 35 $\mu C/cm^2$) as the number of switching cycles increased (fatigue), the tendency for the polarisation in a bit to become less easily switched with time and the number of times it is switched in the same sense (imprint) and the tendency for the switchable polarisation in any one bit to decay with time (retention). The fatigue and imprint problems in PZT have been ascribed to the motion of oxygen vacancies addressed by the use of oxide, rather than metallic, electrodes and doping to reduce the vacancy concentration. It has also been demonstrated that Aurivillius compounds such as strontium bismuth tantalate (SBT) and La-doped $Bi_4Ti_3O_{12}$ do not show fatigue behaviour, although they do require rather higher growth temperatures than the PZT family and have a significantly smaller switchable polarisation (ca. 7 $\mu C/cm^2$). A number of companies are now making chips with nonvolatile on-board embedded ferroelectric random-access memory (FRAM). These are going into applications such as smart cards and computer games.

26.4.3 Piezoelectrics

This represents the most diverse range of applications for ferroelectric materials, as noted in the introductory section, covering a wide range of sensors, actuators and acoustic wave components for frequency control and filtering applications. The key properties determining the performance of a piezoelectric material in a particular application are the piezoelectric coefficients, elastic and dielectric constants. These are all tensor properties (see *Nye* [26.31]) and are combined through the constitutive equations

$$D_i = d_{ijk}X_{jk} + \varepsilon_{il}E_l \,,$$
$$x_{ij} = s^E_{ijkl}X_{kl} + d_{mij}E_m \,,$$

where D_i = electric displacement, E_i = electric field, X_{ij} = stress, x_{ij} = strain, d_{ijk} = piezoelectric coefficients, ε^X_{ij} = dielectric permittivity coefficients at constant stress, and s^E_{ijkl} = elastic compliance coefficients at constant electric field.

The above equations use the Einstein repeated-suffix tensor notation. Note that most piezoelectric equations use the reduced-suffix notation for the piezoelectric and elastic constants [26.31].

There are two basic modes of operation for piezoelectric sensors and actuators. The first is well away from any mechanical resonance – usually at low frequency or quasi-DC. Here, the electric signals generated by an applied stress, or the strain generated by an applied electric field can be easily calculated using the constitutive equations of the type quoted above. The commonly used modes for operating piezoelectric ceramic actuators are longitudinal, where the electric field is applied parallel to the direction in which strain is required and the coupling is via the piezoelectric d_{33} coefficient, and transverse, where a field is applied along the polar axis, but the exploited strain is perpendicular to this direction and coupling is via the d_{31} and d_{32} coefficients. Corresponding modes are frequently used for piezoelectric sensors. The two most-commonly used actuator structures are illustrated in Fig. 26.15. Figure 26.15a shows a piezoelectric bimorph, in which two pieces of piezoelectric ceramic are bonded to a central metal shim. Fields are applied parallel to the polar axes of the ceramic elements so that their transverse extensions are equal and opposite. Usually one end of the bimorph is clamped so that the other end bends. In a variation on this structure, a single piece of ceramic is bonded to a piece of metal, making a unimorph. Bimorph and unimorph devices can provide high displacements (several tens to thousands of microns at up to 200-V drive) but relatively low forces (typically up to 1 N), depending on the size of the bimorph. The second widely used actuator structure is the multilayer

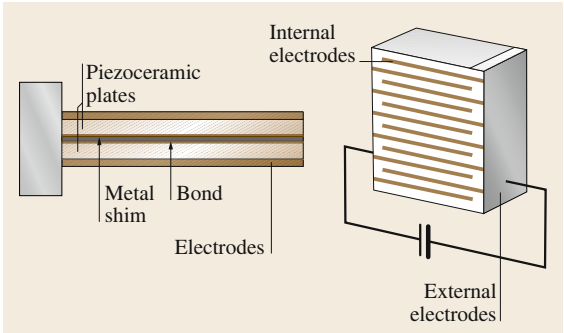


Fig. 26.15 (a) Piezoelectric bimorph and (b) MLC actuator structures

ceramic actuator (Fig. 26.15b). This device operates in longitudinal mode, with applied field and extension parallel to the polar axis. The division of the structure into many thin layers reduces the voltage required to obtain a given extension. Typical devices can produce a few to a few tens of microns extension at a few hundred Volts (depending on dimensions) but high forces (in the kiloNewtons range). It is also possible to use a shear mode of operation for both sensing and actuation. In this actuation mode, the field is applied perpendicular to the polar axis and a shear strain is generated in the plane containing the vectors parallel to the applied field and the polar axis.

Many piezoelectric devices are operated under an alternating electric field whose frequency is tuned to match a mechanical resonance. A piezoelectric element operated well away from resonance will exhibit a capacitative impedance characteristic, As a mechanical resonance is approached, the electrical behaviour of the element can be characterised using an equivalent circuit as shown in Fig. 26.16a. C_0 is the static capacitance. The L_m, R_m, and C_m form the motional arm of the circuit. If the admittance of such a circuit is plotted as a function of frequency, a behaviour is obtained as shown in Fig. 26.16b. The maximum in the admittance is very close to the mechanical resonant frequency (f_r), while the minimum in the admittance is close to antiresonance (f_a); f_r is determined by the elastic properties of the piezoelectric, its density and the dimensions relevant to the particular resonant mode. It is useful to define an effective electromechanical coupling coefficient k_{eff} which is defined by

$$k_{eff}^2 = \frac{\text{mechanical energy converted to electrical energy}}{\text{input mechanical energy}}$$

or

$$k_{eff}^2 = \frac{\text{electrical energy converted to mechanical energy}}{\text{input electrical energy}} ,$$

k_{eff} can be determined from f_r and f_a

$$k_{eff}^2 \approx \frac{f_a^2 - f_r^2}{f_a^2} .$$

It can be shown that, for a simple resonant mode of a piezoelectric element, k_{eff} takes the form

$$k_{eff}^2 = \frac{d^2}{s^E \varepsilon^X}$$

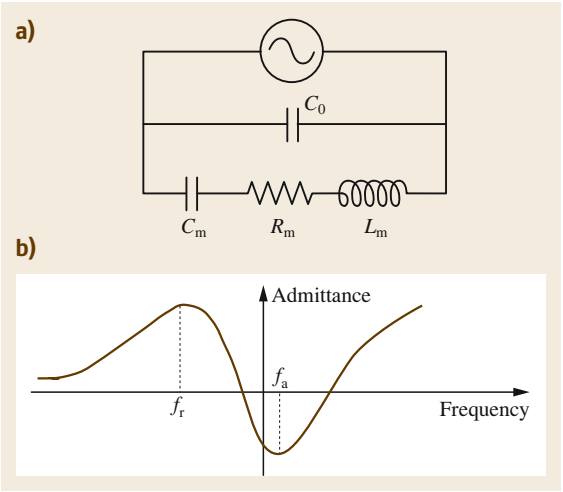


Fig. 26.16 (a) Piezoelectric element equivalent circuit (b) Admittance versus frequency for a typical piezoelectric resonator

where the d, s and ε coefficients in this equation are combinations of the piezoelectric, elastic and dielectric permittivity coefficients. The precise combination of the coefficients will depend on the shape of the piezoelectric element and the mode being excited into resonance. A list of the more commonly used modes and the dependence of k_{eff} on the relevant materials constants is given in Table 26.4. For more complex structures, such as those which involve a piezoelectric element (or elements) bonded to a metal component, it is still possible to define and measure a k_{eff}, but its value will be a much more complex combination of the materials constants, and will also involve the properties of the metal parts and the relative dimensions of the piezoelectric and metal components. (For more details refer to the IEEE standards on piezoelectricity).

By far the most widely exploited piezoelectric materials system is based on the $PbZr_xTi_{1-x}O_3$ ceramic solid solution series. Most of the compositions of-interest are centred on the morphotropic phase boundary (MPB) region (Fig. 26.7) with $x = 0.52$. In the undoped series, the piezoelectric coefficients and dielectric constants peak at this composition. However, their properties can be greatly enhanced by the inclusion of selected cation dopants. These can be classified as follows:

- Isovalent: These are divalent cations that will substitute for Pb^{2+} on the A site (e.g., Sr^{2+}) or tetravalent ions (e.g., Sn^{4+}) substituting for Ti/Zr^{4+} on the B site.
- Donors: These are cations with a valency greater than the site onto which they substitute (e.g., La^{3+} for Pb^{2+} or Nb^{5+} for Ti/Zr^{4+}).

Table 26.4 A list of the most commonly used piezoelectric resonances and the dependence of the k_{eff} for each on the piezoelectric, elastic and dielectric constants. (After [26.32])

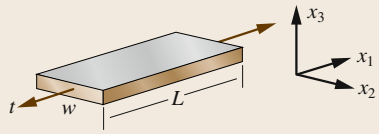
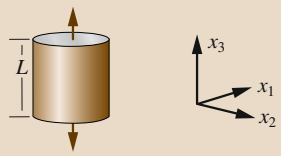
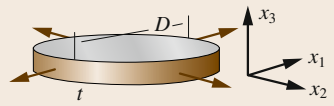
Mode type	Mode shape	Coupling factor
Length extensional mode of a bar using transverse excitation $L > 5w$, $L > 10t$		$k_{31}^l = \frac{d_{31}}{\sqrt{\epsilon_{33}^{\mathrm{T}} s_{11}^{\mathrm{E}}}}$
Length extensional mode of a cylindrical rod using logitudinal excitation $L > 10 \times$ diameter		$k_{33}^l = \frac{d_{33}}{\sqrt{\epsilon_{33}^{\mathrm{T}} s_{33}^{\mathrm{E}}}}$
Radial mode of a round plate using longitudinal excitation $D > 20 \times$ thickness		$k_p = \frac{k_{31}}{\sqrt{2/(1-\sigma^p)}}$
	Grey areas denote the electroded surfaces	

Table 26.5 The effects of substituents in the PZT system. (After [26.11])

Isovalent substituents, e.g., Sr^{2+} for Pb^{2+}, Sn^{4+} for Zr^{4+}	**Donor substituents, e.g., La^{3+} for Pb^{2+}, Nb^{5+} for Zr/Ti^{4+}**
Lower Curie point	No change in Curie point
Increase permittivity	Higher permittivity
Small improvement in linearity at high drive	Higher electromechanical coupling
No change in coupling factor, aging, volume resistivity or low amplitude mechanical or dielectric loss	Poorer high drive linearity
	Lower aging
	Increase resistivity
	Reduced Q_E and Q_M
Acceptor substituents, e.g., K^+ for Pb^{2+}; Fe^{3+} for Zr/Ti^{4+}	**Variable valence substituents, e.g., Mn – can be 2+, 3+, 4+ – for Zr^{4+}**
No change in Curie point	No change in Curie point
Lower permittivity	Range of permittivity
Lower electromechanical coupling	Range of electromechanical coupling – usually lower
Improved drive linearity	High drive/linearity not changed
Little change in aging	Reduced aging
Volume resistivity reduced	
Increased Q_E and Q_M	

- Acceptors: These are cations with a valency lower than the site onto which they substitute (e.g., K^+ for Pb^{2+} or Ni^{3+} for Ti/Zr^{4+}).
- Multivalent: These are ions that take multiple valency (e.g., Mn substituted on the B site).

The broad effects of each class of dopants are listed in Table 26.5.

The resulting PZT ceramics are broadly classified into *hard* or *soft* according to whether they are doped with acceptors or donors, respectively. Each manufacturer has its own nomenclature scheme for its piezoceramic products. However, the US Navy produced a set of specifications under MIL-STD-1376B which are universally recognised. Hard ceramics (e.g., US Navy type III) have relative permittivities in the range of 1000 with high Qs (> 500) and d_{33} of about 215 pC/N. Soft ceramics (e.g., US Navy type VI) have much higher permittivities (3250) and piezoelectric coefficients (d_{33} = 575 pC/N) but lower Q's (65). Hard ceramics have much better drive-voltage stability and higher Curie temperatures than soft ceramics. The soft compositions tend to be used for actuator applications, where the maximum piezoelectric displacement for a given drive

Table 26.6 Typical values for the piezoelectric, dielectric and elastic properties of some selected piezoelectric ceramic materials. (Taken from [26.33])

Material type	I	III	V	VI	Units
Stress-free relative permittivity (ε_{33})	1275	1025	2500	3250	
Dielectric loss tangent ($\tan\delta$)	≤ 0.006	≤ 0.004	≤ 0.025	≤ 0.025	
d_{33}	290	215	495	575	pC/N
k_p	0.58	0.50	0.63	0.64	
Mechanical quality factor Q_m	≥ 500	≥ 800	≥ 70	≥ 65	
Curie temperature (T_C)	325	325	240	180	°C

voltage is sought, while soft ceramics tend to be used for resonant and sound-generation applications. There are, however, exceptions to this. Medical ultrasound transducer arrays tend to use soft, high-permittivity ceramics, because they need to have the maximum charge sensitivity in the reception of the sound waves reflected from tissues within the body, and high permittivities because the small elements usually need to drive transmission lines with significant capacitance.

There are other types of piezoelectric ceramics that are used for specialist applications. Modified $PbTiO_3$ ceramics possess very low values of d_{31} and tend to be used for ultrasound array applications where small cross-coupling between adjacent elements is important. They also possess much higher response to hydrostatic stress than MPB PZT ceramics. Because of this, arrays of small blocks of this type of ceramic are used in flank array sensors in submarines. As noted above, lead metaniobate ceramics are used in high-temperature applications. Table 26.6 lists the piezoelectric and other properties of some selected piezoelectric ceramics.

Single crystals of PZN-PT and PMN-PT have received considerable attention because of the very high piezoelectric d_{33} coefficients and k_{33} coupling factors that can be obtained and single-crystal plates of these are now commercially available. Table 26.7 lists some of the properties that have been measured from these materials, and compares them with a commercial soft PZT ceramic.

Single crystals of $LiNbO_3$ are widely used in surface acoustic wave devices. In these, interdigitated electrodes (IDEs) are applied to one polished face of a crystal. Excitation of these with an RF field will cause the generation of piezoelectrically excited Rayleigh waves, which will propagate with very little attenuation. These can be detected by further sets of IDE placed on the same face of the crystal. Such devices are widely used in signal processing and filtering applications in mobile telecommunication applications.

Thin films of piezoelectric materials, excited into thickness-mode resonance, are being explored for very-high-frequency (> 1 GHz) resonant filter applications. Thin films of PZT are now widely used in inkjet printers.

Table 26.7 Piezoelectric properties of PMN-PT single crystal materials [26.34] compared with those of a commercial soft-piezoelectric ceramic HD3203 [26.35]

Property	Composition	
	PMN-PT X2B	HD3203
$\varepsilon_{33}^T/\varepsilon_o$	5000–7500	3370
T_{RT} (°C)	85–98	NA
T_C (°C)	> 142	195
k_t (%)	55–59	0.536
k_{33} (%)	91–95	0.763
N_{33} (Hz.m)	599	2000
d_{33} (pC/N)	1500–2300	564

26.4.4 Pyroelectrics

The pyroelectric effect is now widely used in uncooled detectors of long-wavelength infrared radiation (IR). The principle behind the operation of these devices is very simple. The radiation to be detected is allowed to fall upon a thin chip of the pyroelectric material. The energy absorbed causes a change in temperature and the generation of a pyroelectric charge, which will cause the flow of current in an external circuit. This can be amplified and used, for example, to switch an alarm. The basic circuit of a pyroelectric detector is shown

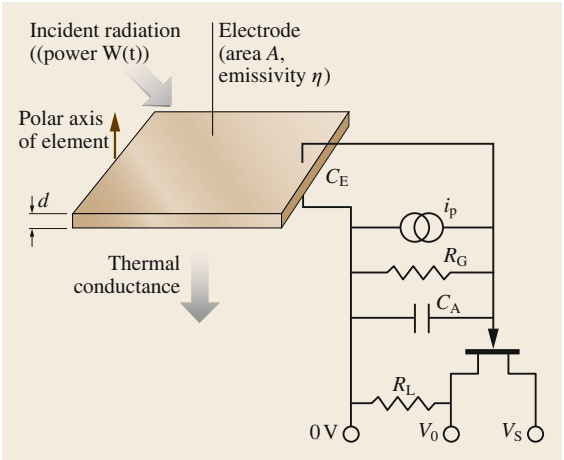


Fig. 26.17 Schematic diagram of a pyroelectric IR sensor

schematically in Fig. 26.17. The field-effect transistor (FET) amplifier needs to be close to the pyroelectric element, ideally in the same package. This is because the latter has a very high output impedance and thus is very noise sensitive. The pyroelectric signal is represented in this circuit as a current source, i_p. i_p is proportional to the rate of change of the element temperature with time. Pyroelectric devices do not have a DC response. They only see changes in the intensity of the radiation with time. This is a major advantage in many applications, where usually there is a requirement to detect changes in the radiation coming from a scene, rather than the average or unchanging background intensity. An example of this is the requirement to detect the radiation from a person coming into the field of view of an IR detector.

Pyroelectric devices have several advantages over other radiation detectors. Their response is independent of the wavelength of the incident radiation, provided there is some means to absorb the radiation. Hence, pyroelectric radiation detectors have been used across the full range of the electromagnetic spectrum, from microwaves to x-rays. Basically a single pyroelectric detector design can be used for different wavelengths of radiation, simply by equipping the package with different windows coated with filters according to the radiation it is desired to sense. Such devices are widely used in spectroscopic sensors for, e.g., automotive exhaust-gas analysis. However, by far the widest application of pyroelectric detectors is in the sensing of IR in the wavelength range $8-14\,\mu m$. The reasons for this are that there is an atmospheric window of low absorption and that objects with temperatures in the range of room temperature (300 K) emit most of their radiative energy in that waveband. However, all of the semiconductor detectors of $-8-14\,\mu m$ IR need to be cooled to about 77 K to work efficiently. On the other hand, pyroelectric detectors work perfectly well uncooled. Cheap detectors are readily available for under $\$1$, and have thus found their way into a host of applications, notably detectors of people in such things as intruder alarms and remote light switches. Pyroelectric arrays have been developed in which a plane of pyroelectric material is interfaced to a silicon chip bearing a two-dimensional array of FET amplifiers and a set of switches which can be used to multiplex the pyroelectric signals onto a single output. These arrays have been used for uncooled thermal imagers with performances comparable to those that can be achieved with cooled semiconductor devices [26.36].

The requirements for the active materials in a pyroelectric device has been described by, for example, *Whatmore* [26.22], *Whatmore* and *Watton* [26.36]. There are two time constants that determine the frequency response of a pyroelectric device: the electrical time constant determined by $R_G(C_E + C_A)$ (see Fig. 26.17) and the thermal time constant determined by H/G where H is the heat capacity of the pyroelectric element and G the thermal conductance from the pyroelectric element to the environment. The capacitance of the element (C_E) is determined by the permittivity (ε) of the pyroelectric material while the heat capacity is determined by its specific heat per unit volume (c'). The pyroelectric current is proportional to the material's pyroelectric coefficient ($p = dP_s/dT$). A full treatment of the physics of device operation shows that detector performance (as determined by specific detectivity) is proportional to one of three basic figures-of-merit (FOMs), which combine the pyroelectric, dielectric and thermal properties of the pyroelectric material together. If the pyroelectric element capacitance is small in comparison with the amplifier capacitance (C_A), then the appropriate FOM is $F_I = p/c'$. If $C_E \gg C_A$ and the noise in the device is dominated by the amplifier noise sources, then the appropriate FOM is $F_V = p/c'\varepsilon$. If the noise is dominated by alternating-current (AC) Johnson noise in the pyroelectric element then the appropriate FOM is $F_D = p/c'\sqrt{\varepsilon \tan\delta}$, where $\tan\delta$ is the dielectric loss tangent of the pyroelectric material. Note that it is very important to measure the dielectric constant and loss at a frequency relevant to the device use. As most pyroelectric devices are used in the frequency range $0.1-100\,Hz$, and dielectric loss is usually much greater at $< 100\,Hz$ than at 1 kHz, a low-frequency measurement is essential. Table 26.8 lists the pyroelectric properties of several ferroelectric materials. Note that many of the literature papers on the subject only quote 1 kHz dielectric measurements. The differences between the FOM for the materials where high- and low-frequency data are available are clear. It can be seen that that the DTGS (deuterated triglycerine sulfate) family gives the largest figures of merit. However, these materials are water soluble and have relatively low Curie temperatures and so they are only still used in the high-performance devices for instruments such as in Fourier-transform infra-red (FTIR) spectrometers. $LiTaO_3$ (LT) is a good pyroelectric material with relatively low permittivity. It is inert and relatively easy to handle. It is used in many single-element detectors. Ceramic materials such as those based on modified $PbZrO_3$ (Mod PZ) or modified $PbTiO_3$ (Mod PT) are widely used in low-cost detectors. Note that, although their FOM are worse than LT, they are low cost and their performance is perfectly adequate for many applications. The PVDF family (represented here by a P(VDF/TrFE70/30 copolymer) have good F_V, but relatively low F_D values compared with the other materials listed. They tend to be used in large-area detectors because of their low permittivi-

Table 26.8 Pyroelectric properties of several ferroelectric materials (taken from *Whatmore* and *Watton* [26.36])

Material	Meas. T (°C)	p (10^{-4} C m^{-2}K^{-1})	Dielectric properties (ε)	(tan δ)	Freq. (Hz)	c' (10^6 J m^{-3}K^{-1})	T_C (°C)	F_V (m^{2}C^{-1})	F_D (10^{-5} Pa$^{-1/2}$)
DTGS (Crystal)	40	5.5	43	0.02	1000	2.6	61	0.53	8.3
$LiTaO_3$ (Crystal)	25	2.3	47	10^{-3}	1000	3.2	665	0.17	11.1
P(VDF/TrFE) 70/30 (Polymer)	25	0.33	7.4	0.017	1000	2.3	121	0.22	1.4
Mod. PZ (Ceramic)	25	4.0	290 300	0.003 0.014	1000 33	2.8	230	0.054 0.054	5.2 2.3
Mod. PT (Ceramic)	25	3.5	220 220	0.01 0.03	1000 33	2.8	> 250	0.063 0.063	2.9 1.6

ties, low cost and the fact that they are readily made in very thin films with low thermal mass, which is an advantage in some circumstances. They have been demonstrated in linear arrays. It is advantageous for very small-area detectors (say $< 100\,\mu$m square), such as those used in arrays, for the pyroelectric material to have a relatively high permittivity (a few hundred) so that the detector can have a capacitance that matches the input capacitance of the FET amplifier (usually ca. 1 pF). The ceramic materials are well suited to this application for that reason. There has been considerable research into the use in thermal imaging arrays of ferroelectrics with T_C close to room temperature under an applied bias field, which will provide an induced pyroelectric effect well above the normal T_C. This has been called dielectric bolometer mode of operation and the best materials researched for this have been $Pb(Sc_{0.5}Ta_{0.5})O_3$ (PST) and $(Ba_xSr_{1-x})TiO_3$ solid solutions with $x \approx 0.35$. These materials have relative permittivities of > 1000 under the operational conditions and very high pyroelectric coefficients which can give an effective F_D some three times greater than can be achieved using conventional pyroelectric ceramics [26.37].

References

26.1 G. Busch: Ferroelectrics **74**, 267 (1987)
26.2 A.F. Devonshire: Phil. Mag. **40**, 1040 (1949)
26.3 M.E. Lines, A.M. Glass: *Principles and Applications of Ferroelectric Materials* (Clarendon, Oxford 1977)
26.4 V.M. Goldschmidt: *Geochemistry* (Oxford Univ. Press, Oxford 1958)
26.5 R.D. Shannon, C.T. Prewitt: Acta Cryst. B **25**, 925 (1969)
26.6 A.M. Glazer: Acta Cryst. B **28**, 3384 (1972)
26.7 W.J. Merz: Phys. Rev. **76**, 1221 (1949)
26.8 J.M. Herbert: *Ceramic Dielectrics and Capacitors* (Gordon Breach, Philadelphia 1985)
26.9 R.W. Whatmore, A.M. Glazer: J. Phys. C **12**, 1505 (1979)
26.10 B. Noheda, J.A. Gonzalo, A.C. Caballero, C. Moure, D.E. Cox, G. Sirane: Ferroelectrics **237**, 541 (2000)
26.11 B. Jaffe, W.R. Cook Jr., H. Jaffe: *Piezoelectric Ceramics* (Academic, New York 1971)
26.12 R.W. Whatmore: Structural Phase Transitions in Modified Lead Zirconate, Ph.D. Thesis (Cambridge Univ., Cambridge 1977)
26.13 G.A. Smolenskii, V.A. Isupov, A.A. Agranovskaya, S.N. Popov: Fiz. Tverd. Tela **2**, 2906 (1960)
26.14 N.-H. Chan, D.M. Smyth: J. Am. Ceram. Soc. **67**, 285 (1984)
26.15 N.-H. Chan, R.K. Sharma, D.M. Smyth: J. Am. Ceram. Soc. **65**, 168 (1981)
26.16 S.C. Abrahams, E. Buehler, W.C. Hamilton, S.J. Laplaca: J. Phys. Chem. Solids **34**, 521 (1973)
26.17 P.B. Jamieson, S.C. Abrahams, J.L. Bernstein: J. Chem. Phys. **48**, 5048 (1968)
26.18 L. Keeney, T. Maity, M. Schmidt, A. Amann, N. Deepak, N. Petkov, S. Roy, M.E. Pemble, R.W. Whatmore: J. Am. Ceram. Soc. **96**, 2339 (2013)
26.19 M.E. Hagerman, K.R. Poeppelmeier: Chem. Mater. **7**, 602 (1995)
26.20 Y.J. Ding, X.D. Mu, X.H. Gu: J. Non-Lin. Opt. Phys. Mater. **9**, 21 (2000)
26.21 D. Madgy, S.F. Ahsan, D. Kest, I. Stein: Arch. Otolaryng. Head Neck Surg. **127**, 47 (2001)
26.22 R.W. Whatmore: Rep. Prog. Phys. **49**, 1335 (1986)
26.23 H. Kawai: Jpn. J. Appl. Phys. **8**, 967 (1969)
26.24 J.G. Bergman, J.H. McFee, G.R. Crane: Appl. Phys. Lett. **18**, 203 (1971)
26.25 J.A. Ghambaryan, R. Guo, R.K. Hovsepyan, A.R. Poghosyan, E.S. Vardanyan, V.G. Lazaryan: J. Optoelectron. Adv. Mater. **5**, 61 (2003)
26.26 R. Clarke, R.W. Whatmore: J. Cryst. Growth **33**, 29 (1976)
26.27 A. Dabkowski, H.A. Dabkowska, J.E. Greedan, W. Ren, B.K. Mukherjee: J. Cryst. Growth **265**, 204 (2004)
26.28 R.A. Dorey, R.W. Whatmore: J. Electroceram. **12**, 19 (2004)

26.29 N. Neuman, R. Köhler: Proc. SPIE **2021**, 35 (1993)
26.30 C. Pas de Araujo, J.F. Scott, G.W. Taylor: *Ferroelectric Thin Films: Synthesis and Basic Properties* (Gordon, Breach, Princeton 1996)
26.31 J.F. Nye: *Physical Properties of Crystals, Their Representation by Tensors and Matrices* (Oxford Univ. Press, Oxford 1957)
26.32 ANSI/IEEE Standard on Piezoelectricity: IEEE Trans. **UFFC 43**, 717 (1996)
26.33 MIL-STD-1376B(SH) (1995) US Navy Military Standard for Piezoelectric Ceramics
26.34 http://www.trstechnologies.com/Materials/High-Performance-PMN-PT-Piezoelectric-Single-Crystal
26.35 S. Sherrit, H.D. Wiederick, B.K. Mukherjee: Proc. SPIE **3037**, 158 (1997)
26.36 R.W. Whatmore, R. Watton: Pyroelectric Materials and Devices. In: *InInfrared Detectors and Emitters: Materials and Devices*, ed. by P. Capper, C.T. Elliott (Chapman Hall, New York 2000) p. 99
26.37 R.W. Whatmore, P.C. Osbond, N.M. Shorrocks: Ferroelectrics **76**, 351 (1987)

27. Dielectric Materials for Microelectronics

Robert M. Wallace

Dielectrics are an important class of thin-film electronic materials for microelectronics. Applications include a wide swathe of device applications, including active devices such as transistors and their electrical isolation, as well as passive devices, such as capacitors. In a world dominated by Si-based device technologies, the properties of thin-film dielectric materials span several areas. Most recently, these include high-permittivity applications, such as transistor gate and capacitor dielectrics, as well as low-permittivity materials, such as inter-level metal dielectrics, operating at switching frequencies in the gigahertz regime for the most demanding applications.

This chapter provides a survey of the various dielectric material systems employed to address the very substantial challenge associated with the scaling Si-based integrated circuit technology. A synopsis of the challenge of device scaling is followed by an examination of the dielectric materials employed for transistors, device isolation, memory and interconnect technologies. This is presented in view of the industry roadmap which captures the consensus for device scaling (and the underlying economics) – the International Technology Roadmap for Semiconductors.

27.1 **Overview** 615
27.1.1 The Scaling of Integrated Circuits 615
27.1.2 Role of Dielectrics for ICs 619
27.2 **Gate Dielectrics** 620
27.2.1 Transistor Structure 620
27.2.2 Transistor Dielectric Requirements in View of Scaling 621
27.2.3 Silicon Dioxide 625
27.2.4 Silicon Oxynitride: SiO_xN_y 631
27.2.5 High-k Dielectrics 633
27.3 **Isolation Dielectrics** 635
27.4 **Capacitor Dielectrics** 636
27.4.1 Types of IC Memory 636
27.4.2 Capacitor Dielectric Requirements in View of Scaling 636
27.4.3 Dielectrics for Volatile Memory Capacitors 637
27.4.4 Dielectrics for Nonvolatile Memory 637
27.5 **Interconnect Dielectrics** 639
27.5.1 Tetraethoxysilane (TEOS) 640
27.5.2 Low-k Dielectrics 640
27.6 **Summary** 641
References 641

27.1 Overview

This chapter considers the role of dielectric materials in microelectronic devices and circuits and provides a survey of the various materials employed in their fabrication. We will examine the impact of scaling on these materials, and the various materials utilized for their dielectric behavior. Extensive reviews are available on the device characteristics for the reader to consult [27.1–4]. We will primarily confine the discussion here to Si-based microelectronic circuits, which continue to dominate the industry.

Dielectric materials are an integral element of all microelectronic circuits. In addition to their primary function of electrical isolation of circuit and device components, these materials also provide useful chemical and interfacial properties. The material (and resulting electrical) properties of dielectrics must also be considered in the context of the thin films used in semiconductor microelectronics, as compared to bulk properties. The dimensions of these dielectric thin films are determined by the device design of the associated integrated circuit technology, and these dimensions decrease due to a calculated design process called *scaling*.

27.1.1 The Scaling of Integrated Circuits

The ability to reduce the size of the components of integrated circuits (ICs), and therefore the circuits

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_27

themselves, has resulted in substantial improvements in device and circuit speeds over the last 40 years. Equally important, this calculated size reduction permits the fabrication of a higher density of circuits per unit area on semiconductor substrates. The economic implication of this scaling was captured by *G. Moore* more than 50 years ago [27.5].

Moore's Law

Moore observed that the minimum cost of manufacturing integrated circuits per component actually decreases with increasing number of IC components, and thus with greater circuit functionality and computing power. This is obviously an important economic driving force, as the ability to increase the number of circuits per unit area would lead to a lower minimum cost, and thus higher market demand and more potential profit. Moreover, Moore noted that the rate of increase in the number of components for a given circuit function roughly doubled each year in the early 1960s, and predicted that it would continue to do so through 1975. In 1975, *Moore* revised this estimate of doubling time to 24 months due to the anticipated complexity of circuits [27.6].

The semiconductor industry has generally confirmed (and pursued) these extrapolations, often referred to as Moore's law, over the last 40 years. The extrapolation is often analyzed in the semiconductor industry, and the doubling period, which has varied between 17 and 32 months over the life of the industry, is now roughly 24 months, with projections to lengthen [27.8]. Indeed the cost per transistor has decreased from $\approx 5\,\$/\text{transistor}$ in 1965 to less than $10^{-9}\,\$/\text{transistor}$ today [27.9]. Current advanced Si IC production technology results in the fabrication of well over 1 billion transistors on a microprocessor chip.

Technology Roadmap

The contemporary industry analysis encompassing this observation is presented in the International Technology Roadmap for Semiconductors (ITRS) where the extrapolations of the future technological (and economic) targets for the industry are annually updated [27.7]. The current scaling trends indicate that the compound annual reduction rate (CARR) in device dimensions is currently consistent with the following equation

$$\text{CARR}(T) = \frac{1}{2}^{\left(\frac{1}{2T}\right)} - 1\,, \tag{27.1}$$

where T is the technology node cycle time measured in years. Thus in two years, the rate of reduction is -15.9%. This corresponds to a scaling factor of $\approx 0.7\times$ from a given technology node to the next, or a factor of $\approx 0.5\times$ over the time of two technology nodes. This pace has been enabled through the introduction of new materials technologies such as strained Si, SiGe, and high-k/metal gate technologies. Analysis of the limits of Si-based scaled integrated circuit technology are available, including alternative channel materials, such as Ge and III-V semiconductors integrated on Si, have been discussed [27.10, 11]. According to the most recent ITRS Roadmap, non-Si channel materials may be in transistor production soon [27.7]. Scaling has also driven the transistion from planar complementary metal-oxide-semiconductor (CMOS) technologies in production to three-dimensional (3-D) (vertical) transistor component geometries.

Table 27.1 provides selected scaling targets from the ITRS through 2020 [27.7]. As may be seen, the industry roadmap now has scaling targets which are driven by controlling power consumption. Higher-performance technologies, such as MPU/ASIC applications, require more aggressive scaling, while low power applications require less-aggressive scaling. A key criteria to enable these technologies is the associated power management in the *on* and *off* state, hence leakage current (I_{off}) remains an important distinction amongst the various roadmap applications.

Table 27.1 Selected scaling targets from the ITRS roadmap [27.7]

Year of production	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Technology node label (nm)	**"16/14"**		**"11/10"**		**"8/7"**		**"6/5"**	
High-Performance applications								
Physical gate length (nm)	20	18	16.7	15.2	13.9	12.7	11.6	10.6
Equivalent oxide thickness, T_{ox} (nm)	0.8	0.77	0.73	0.70	0.67	0.64	0.61	0.59
Drive voltage V_{dd} (V)	0.86	0.85	0.83	0.81	0.80	0.78	0.77	0.74
I_{off} (*Leakage*) (nA/μm)	100	100	100	100	100	100	100	100
Low Power applications								
Physical gate length (nm)	23	21	19.0	18.0	16.0	14.6	13.3	12.2
T_{ox} (nm)	0.8	0.77	0.73	0.70	0.67	0.64	0.61	0.59
V_{dd} (V)	0.86	0.85	0.83	0.81	0.80	0.78	0.77	0.74
I_{off} (nA/μm)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2

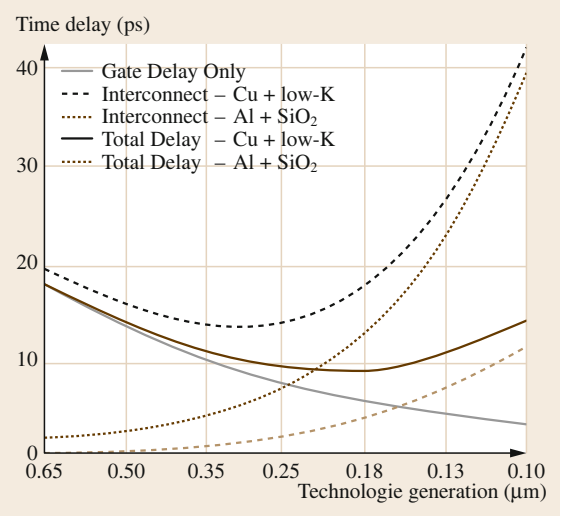


Fig. 27.1 Time delay contributions for integrated circuits. The resultant delay is dependent upon the circuit architecture as well as the material constituents. (After [27.12], © 2004 IEEE, with permission)

Performance and Scaling

The concomitant reduction in device dimensions, such as transistor channel length, associated with increasing the number of components per unit area has resulted in a significant increase in processing performance – the speed at which computations can be done. For example, as scaling reduces the distance that carriers must travel in a transistor channel, the response time of the transistor as a digital switch also decreases (as long as sufficient mobility is maintained). Scaling has resulted in the speeds of microprocessors increasing from 25–50 MHz clock frequencies in the early 1990s ($\approx 10^6$ transistors) to 2.2 GHz in 2003 ($\approx 10^8$ transistors) [27.13]. This speed has essentially leveled off since 2003 due to power consumption considerations and the resultant voltage scaling invoked, as well as interconnect limitations [27.9].

The various dielectric materials associated with the components impacts the overall performance of the corresponding IC technology. In the case of transistors, the gate dielectric is integral to the performance of transistor electrical characteristics such as the drive current I_d. The interconnection performance of circuit elements is influenced by the dielectric material that isolates the various metal interconnection lines through the line-to-line capacitance. Memory elements incorporate very-low-leakage dielectric materials for charge-storage purposes. The behavior of these materials with component size reduction is thus an important design consideration in IC technology.

Figure 27.1 shows the dependence of the delay time as a function of technology node (scaling) associated with conventional CMOS transistor gates interconnected with metal lines that are isolated with different dielectrics [27.12]. The total delay time associated with the circuit response for a given technology node is a combination of the temporal response of the transistor gate (influenced by the gate capacitance) as well as the propagation through the interconnecting wire lines associated with the circuit (influenced by the wire resistance and the line-to-line capacitance, i. e., a resistive–capacitive (*RC*) delay) [27.12, 15]. Changing both the interconnect wire material (from Al to Cu) to decrease the line resistivity as well as decreasing the dielectric constant in the material isolating these lines (from SiO_2 to a lower dielectric constant material) to decrease line-to-line parasitic capacitance minimizes the overall *RC* delay time response. Of course, the challenge in adopting this solution was the successful integration of these material changes into a high-volume fabrication line, which took the investment of considerable time and effort.

In order to compare the performance of different CMOS technologies, one can consider a performance metric that captures the dynamic response (i. e., charging and discharging) of the transistors associated with a specific circuit element to the supply voltage provided to the element at a representative (clock) frequency [27.14, 16, 17]. A common element employed to examine such switching-time effects in a new transistor design technology node is a CMOS inverter. This circuit element is shown in Fig. 27.2 where the input signal is attached to the gates and the output signal is connected to both the n-type metal/oxide/semiconductor (MOS) (NMOS) and p-type MOS (PMOS) transistors associated with the CMOS stage. The switching time is limited by both the fall time required to discharge the load capacitance by the n-type field-effect transistor (n-FET) drive current and the rise time required to charge the load capacitance by the p-FET drive current. That is, the switching response times τ are given by

$$\tau = \frac{C_{\mathrm{LOAD}} V_{\mathrm{dd}}}{I_{\mathrm{d}}}\,, \quad \text{where}$$
$$C_{\mathrm{LOAD}} = FC_{\mathrm{GATE}} + C_{\mathrm{j}} + C_{\mathrm{i}}\,, \tag{27.2}$$

and C_j and C_i are the parasitic junction and local interconnection capacitances, respectively. Of course, as one considers the combination of transistors to produce various circuit logic elements to enable computations (viz. a *processor*), it is also clear that the manner in which the transistors are interconnected also plays an impor-

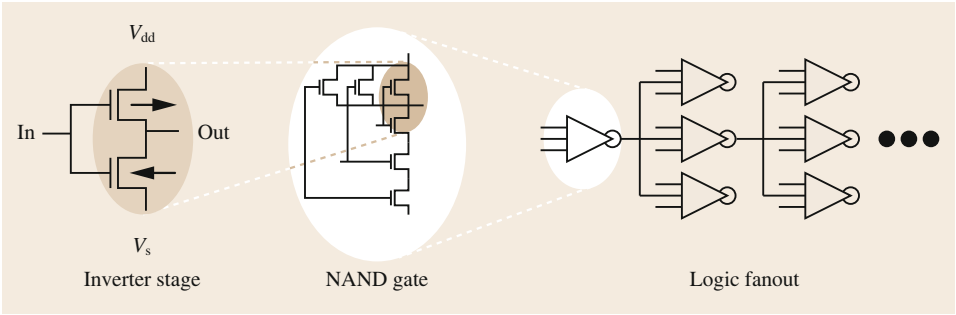


Fig. 27.2 Example circuits employed to evaluate the performance of a transistor technology. For example, elements are combined to simulate processor performance adequately. (After [27.14], AIP, with permission © 2001)

tant role on the overall performance of the circuit. The *fan out* for interconnected devices is given by the factor F. Ignoring delay in the gate electrode response, as $\tau_{GATE} \ll \tau_{n,p}$, the average switching time is therefore

$$\bar{\tau} = \frac{\tau_p + \tau_n}{2} = C_{LOAD} V_{dd} \left(\frac{1}{I_d^n + I_d^p} \right) . \qquad (27.3)$$

The load capacitance in the case of a single CMOS inverter is simply the gate capacitance if one ignores parasitic contributions such as junction and interconnect capacitance. Hence, an increase in I_d is desirable to reduce switching speeds. For more realistic estimates of microprocessor performance, the load capacitance is connected (*fanned out*) to other inverter elements in a predetermined fashion. When coupled with other NMOS/PMOS transistor pairs in the configuration shown in Fig. 27.2, one can create a logic NAND gate, which can be used to investigate the dynamic response of the transistors and thus examine their performance under such configurations. For example, a fan out $F = 3$ can be employed in microprocessor performance estimates, as shown in Fig. 27.2.

One can then characterize the performance of a circuit technology (based on a particular transistor structure) through this switching time. To do this, various *figures of merit* (FOMs) have been proposed that incorporate parasitic capacitance as well as the influence of gate sheet resistance on the switching time [27.16–18]. For example, a common FOM employed is related to (27.3) simply by

$$\text{FOM} \cong \frac{1}{\bar{\tau}} = \frac{2}{\tau_p + \tau_n} . \qquad (27.4)$$

In the case where parasitic capacitance and resistance effects are ignored, it is easily seen then that an increase in the device drive current I_d results in a decrease in the switching time and thus an increase in the FOM value (performance). However, the incorporation of parasitic effects results in the limitation of FOM

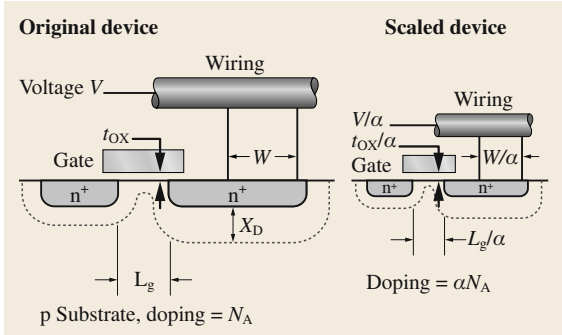


Fig. 27.3 *Classical* scaling methodology for integrated circuit technology. (After [27.21], © 2004 IEEE, with permission)

improvement, despite an increase in the gate dielectric capacitance. All of these issues depend critically upon the materials constituents of the integrated circuit, and dielectrics are an important component [27.14].

Impact of Scaling on Dielectrics

The scaling associated with the device dimensions that constitute an integrated circuit impacts on virtually all film dimensions such as thickness, including the dielectrics employed in the circuit. Various scaling methodologies have been applied since the pioneering work by *Dennard* and coworkers [27.19, 20]. These various methods have been recently summarized [27.2, 21]. For example, following the work by *Dennard* et al., the scaling of the transistor from one node N to the next $N + 1$ under a constant-electric-field condition would result in a relationship between the *new* transistor gate (or channel) length $L_{N+1} = L_N/\alpha$, where $\alpha > 1$ is the constant scaling factor (Fig. 27.3). Other dimensions, such as the gate dielectric thickness and the gate width, scale similarly. The voltage would also scale according to $V_{N+1} = V_N/\alpha$ in this scenario as well. Table 27.1 provides the drive voltage scaling (V_{dd}) anticipated by the ITRS roadmap, depending upon the application.

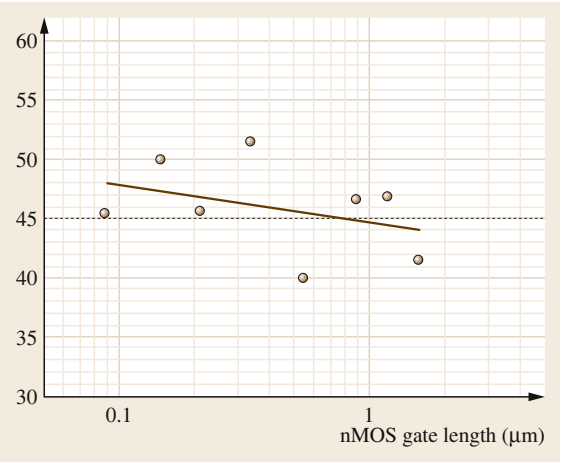


Fig. 27.4 Scaling trends of the planar transistor gate length and the gate insulator thickness for various generations of integrated technologies. The *solid line* is a linear fit to the data. The *dashed line* represents the approximate scaling trend between the gate dielectric thickness and gate length observed over several technology generations: $L_G = 45 \times t_{ox}$. (After [27.22], with permission)

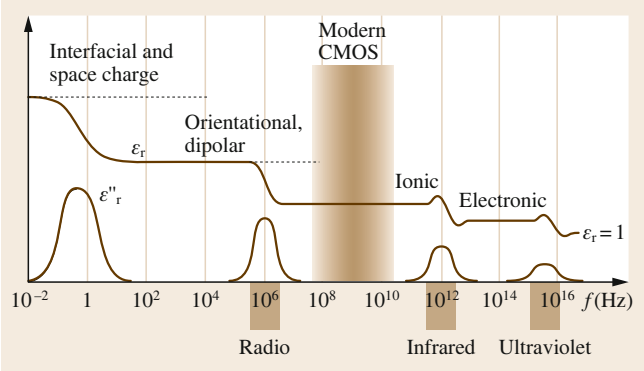


Fig. 27.5 Schematic of the dependence of dielectric permittivity on frequency. (After [27.23], with permission)

This scaling approach has several shortcomings, including the reduction of power-supply voltages with each node, raising problems with circuit power-supply compatibility. So, for earlier technology nodes ($L \geq 0.35\,\mu m$), the power supply was kept constant while the device dimensions were scaled as described above. Eventually, the resultant higher electric fields associated with this scaling approach led to reliability concerns for the gate dielectric, mobility degradation and hot-carrier effects. For further scaling, the power supply was scaled at a different factor (such as $\alpha^{1/2}$) to compensate for such effects. For modern devices with *deep-submicron* gate length, other techniques are used to scale from node to node. Nevertheless, for planar CMOS technology through the 90 nm node, it has been noted that the gate dielectric thickness can be related to the channel length as shown in Fig. 27.4 for a wide variety of device technologies as the depletion depth was scaled with the dielectric [27.22]. As seen in the ITRS Roadmap [27.7], scaling of the gate dielectric has now reached a value just below a $T_{ox} \approx 0.8\,nm$ ($k \approx 12.5$), and projected to decrease to $T_{ox} \approx 0.6\,nm$ by 2020, implying a higher dielectric constant will be needed ($k \approx 16$).

The critical point in the context of dielectric films is that the dimensions and geometry of the devices, and therefore the dimensions (thicknesses) of the dielectric layers, are significantly reduced as scaling proceeds (Table 27.1). At some point, the materials properties associated with the dielectric in a component will be predicted to no longer provide the desirable electrical behavior attained in previous nodes, thus stimulating research on new dielectric materials. In the case of gate dielectrics, scaling results in thinner layers of a given material to a point where quantum-mechanical tunneling becomes an issue. For memory capacitor dielectrics, a specific capacitance density is required, regardless of capacitor dimensions, for a reliable memory element. Indeed, scaling impacts on the selection and integration of numerous constituents of electronic devices including gate electrode materials, source/drain junction regions, contacts, interconnect dielectrics.

27.1.2 Role of Dielectrics for ICs

Dielectrics are pervasive throughout the structure of an integrated circuit. Applications include gate dielectrics, tunneling oxides in memory devices, capacitor dielectrics, interconnect dielectrics and isolation dielectrics, as well as sacrificial or masking applications during the circuit fabrication process.

The basic properties utilized in dielectric materials include their structure and the resultant polarizability behavior. As noted in Fig. 27.5, there are several mechanisms to consider in the polarization of materials in the presence of an electric field. For nonmetallic (insulating) solids in CMOS applications, there are two main contributions of interest to the dielectric constant which give rise to the polarizability: electronic and ionic dipoles. Figure 27.5 illustrates the frequency ranges where each contribution to the real part of the complex relative dielectric permittivity ε_r dominates, as well as the imaginary part of the complex permittivity (dielectric losses), ε''_r [27.23]. The region of interest to CMOS applications includes the shaded area where switching frequencies are in the GHz range at the present time.

In general, atoms with a large ionic radius (i. e., high atomic number, Z) exhibit more electron dipole response to an external electric field, because there are more electrons to respond to the field (electron screening effects also play a role in this response). This electronic contribution tends to increase the permittivity of materials with higher-Z atoms.

The ionic contribution to the permittivity can be much larger than the electronic portion in cases such as perovskite crystals of $(Ba,Sr)TiO_3$ (BST) and $(Pb,Zr)TiO_3$ (PZT), which exhibit ferroelectric behavior below the Curie point. In these cases, Ti ions in unit cells throughout the crystal are uniformly displaced in response to an applied electric field (for the case of ferroelectric materials, the Ti ions reside in one of two stable, non-isosymmetric positions about the center of the Ti−O octahedra). This displacement of Ti ions causes an enormous polarization in the material, and thus can give rise to very large dielectric constants in bulk films of 2000−3000, and has therefore been considered for dynamic random-access memory (DRAM) capacitor applications [27.24, 25]. Since ions respond more slowly than electrons to an applied field, the ionic contribution begins to decrease at very high frequencies, in the infrared range of $\approx 10^{12}$ Hz, as shown in Fig. 27.5.

27.2 Gate Dielectrics

27.2.1 Transistor Structure

A cross-sectional schematic of the structure of a planar metal–oxide–semiconductor (MOS) field-effect transistor (FET) and a modern CMOS transistor, consisting of the n-FET and p-FET pair, is shown in Fig. 27.6. The source- and drain-region dopant profiles reflect modern planar CMOS technologies where an extended, doped region is produced under the gate region in the channel. Additionally, so-called *halo* or *pocket* dopant implantation regions are also shown. These dopant profile approaches have been incorporated in recent years in an effort to permit transistor channel scaling and yet maintain performance [27.27].

As can be seen, there exist several regions of the device which require dielectric materials to enable useful transistor operation. Both n-MOS and p-MOS transistors are isolated with a dielectric, typically deposited SiO_2 in most modern device technologies, which is deposited in trench structures. This isolation technique is called *shallow* trench isolation (STI) as the associated trench depth is $\leq 0.5\,\mu m$. Earlier generations of CMOS (and often devices under research) also utilized the so-called local oxidation of silicon (LOCOS) isolation approach [27.28, 29]. So-called spacer dielectrics (typically SiO_2, SiO_xN_y, or SiN_x) are also used around the transistor *gate stack* for isolation and implantation-profile control. The gate stack is defined here as the films and interfaces comprising the gate electrode, the underlying gate dielectric, and the channel region. The gate dielectric, typically SiO_2 or SiO_xN_y for transistors currently in production, electrically isolates the gate electrode from the underlying Si channel region while allowing the modulation of the carrier flow in the channel. As we shall see, the interface between the gate dielectric and the channel regions is particularly important in regard to device performance.

Transistor structures have further evolved (Fig. 27.7) to a non-planar configuration to enable adequate electrostatics. Drive current and characteristics control with scaling. Examples include the multigate (*FinFET*) and *gate-all-around* transistor

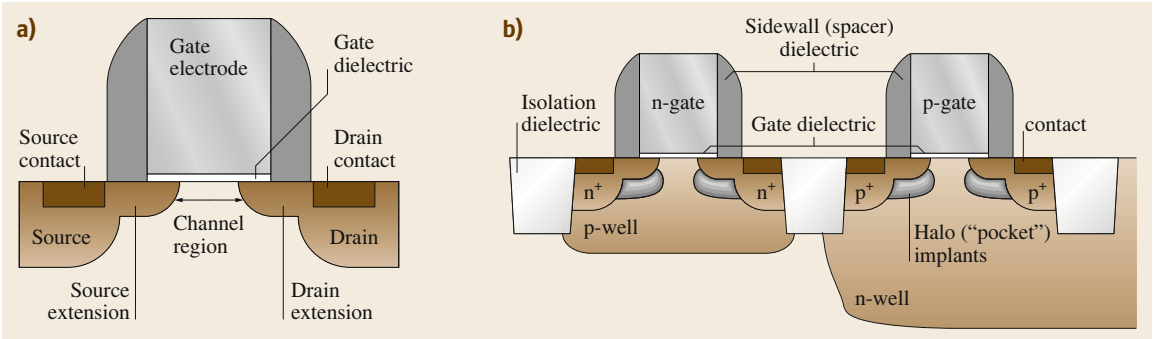


Fig. 27.6a,b Important regions of **(a)** a metal–oxide–semiconductor (MOS) field-effect transistor, and **(b)** a planar CMOS transistor structure. (After [27.26], © 2004 IEEE, with permission)

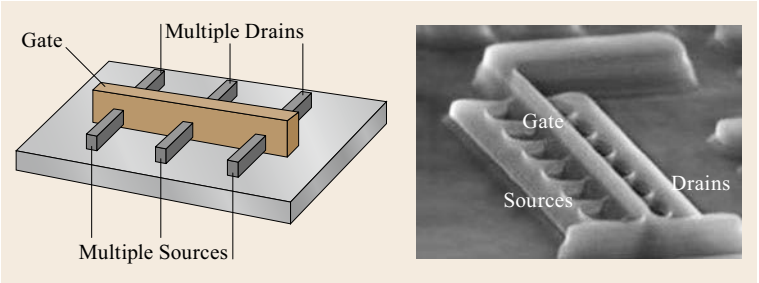


Fig. 27.7 Schematic and image of a *trigate* transistor structure

structures. All of such structures still include a gate dielectric to isolate the gate from the channel, but require conformal deposition processes to enable reproducible and reliable performance.

27.2.2 Transistor Dielectric Requirements in View of Scaling

In addition to the thickness reduction associated with scaling, the gate dielectric must adhere to several requirements that are dependent upon the specific product application, including adequate drive current, suitable capacitance, minimal leakage current, and reliable performance. Table 27.1 provides some selected values describing the ITRS 2013 roadmap.

Drive Current

The improved performance associated with the scaling of logic device dimensions can be seen by considering a simple model for the drive current associated with a FET [27.14, 16, 18, 30]. In the gradual channel approximation, the drive current can be written as

$$I_{\mathrm{d}} = \frac{W}{L}\mu C_{\mathrm{inv}}\left(V_{\mathrm{G}} - V_{\mathrm{T}} - \frac{V_{\mathrm{D}}}{2}\right)V_{\mathrm{D}}\,, \qquad (27.5)$$

where W is the width of the transistor channel, L is the channel length, μ is the channel carrier mobility (assumed to be constant in this analysis), C_{inv} is the capacitance density associated with the gate dielectric when the underlying channel is in the inverted state, V_{G} and V_{D} are the voltages applied to the transistor gate and drain, respectively, and the threshold voltage is given by V_{T}. It can be seen that in this approximation the drain current is proportional to the average charge across the channel (with a potential $V_{\mathrm{D}}/2$) and the average electric field (V_{D}/L) along the channel direction. Initially, I_{D} increases linearly with V_{D} and then eventually saturates to a maximum when $V_{\mathrm{D,sat}} = V_{\mathrm{G}} - V_{\mathrm{T}}$ to yield:

$$I_{\mathrm{D,sat}} = \frac{W}{L}\mu C_{\mathrm{inv}}\frac{(V_{\mathrm{G}} - V_{\mathrm{T}})^2}{2}\,. \qquad (27.6)$$

A goal here is to increase the saturation drive current as much as possible for a given transistor technology design. The term $(V_{\mathrm{G}} - V_{\mathrm{T}})$ is limited in range due to reliability concerns as too large a V_{G} would create an undesirable, high electric field across the gate dielectric. Moreover, practical operation of the device at room temperature (or above) requires that V_{T} cannot easily be reduced below about $8 \times k_{\mathrm{B}}T \cong 200\,\mathrm{meV}$ ($k_{\mathrm{B}}T \approx 25\,\mathrm{meV}$ at room temperature). Elevated operating temperatures could therefore cause statistical fluctuations in thermal energy, which would adversely affect the desired V_{T} value. Thus, even in this simplified approximation, one is left to pursue a reduction in the channel length, increasing the width, increasing the gate dielectric capacitance, or a carefully engineered combination of all of these in order to increase $I_{\mathrm{D,sat}}$. Note that none of these dimensions are, in reality, independent and so substantial resources are utilized to simulate device performance in order to maximize the performance prior to the fabrication of a representative transistor [27.2, 29, 31].

Gate Capacitance

In the case of increasing the inversion capacitance to improve I_{D}, consider the gate stack structure idealized as a parallel-plate capacitor (ignoring quantum-mechanical and depletion effects from a Si substrate and gate [27.32]),

$$C = \frac{k\varepsilon_0 A}{t}\,, \qquad (27.7)$$

where k is the dielectric constant of the material, ε_0 is the electric constant ($= 8.85 \times 10^{-3}\,\mathrm{fF/\mu m}$), A is the area of the capacitor, and t is the thickness of the dielectric between the capacitor electrode plates. (Note that the relative permittivity of a material is often given by ε or ε_{r}, such as with the expression $C = \varepsilon\varepsilon_0 A/t$. Note that the relation between k and ε varies depending on the choice of units (e.g., when $\varepsilon_0 = 1$), but since it is always the case that $k \approx \varepsilon$, we shall assume here that $k = \varepsilon$.) It is clear that transistor designs utilizing planar CMOS scaling approaches discourage increasing the area of the capacitor, and so one is left to considering the reduction of the dielectric thickness in order to increase the capacitance.

In order to compare alternative dielectric materials which exhibit a dielectric constant higher than the standard for the industry, SiO_2, this expression for C can also be rewritten in terms of t_{eq} (i. e., equivalent oxide thickness) and the k_{ox} of the capacitor ($k_{ox} = 3.9$ for the low-frequency dielectric constant of SiO_2). The term t_{eq} then represents the theoretical thickness of SiO_2 that would be required to achieve the same capacitance density as the dielectric (ignoring important issues such as leakage current and reliability). For example, if the capacitor dielectric is SiO_2, $t_{eq} = 3.9\,\varepsilon_0(A/C)$, and so a capacitance density of $C/A = 34.5\,\mathrm{fF}/\mu\mathrm{m}^2$ corresponds to $t_{eq} = 10\,\text{Å}$.

Thus, the physical thickness of an alternative high-k dielectric employed to achieve the equivalent capacitance density of $t_{eq} = 10\,\text{Å}$ can be obtained from the expression

$$\frac{t_{eq}}{k_{ox}} = \frac{t_{high\text{-}k}}{k_{high\text{-}k}} \quad \text{or simply,}$$

$$t_{high\text{-}k} = \frac{k_{high\text{-}k}}{k_{ox}} t_{eq} = \frac{k_{high\text{-}k}}{3.9} t_{eq}\,. \tag{27.8}$$

From this expression, a dielectric with a relative permittivity of 16 therefore results in a physical thickness of $\approx 40\,\text{Å}$, to obtain $t_{eq} = 10\,\text{Å}$. The increase in the physical thickness of the dielectric impacts on properties such as the tunneling (leakage) current through the dielectric, and is discussed further below.

The industry scaling process clearly presents a major challenge for the core transistor gate dielectric as predictions call for a much thinner effective thickness for future alternative gate dielectrics: $t_{eq} \leq 1\,\text{nm}$ [27.7]. The interfacial regions between the gate electrode, dielectric and channel (in totality termed the metal–insulator–semiconductor (MIS) *gate stack* as shown schematically in Fig. 27.8) require careful attention, as they are particularly important in regard to transistor performance. These regions, less than approximately 0.5 nm thick, serve as a transition between the atoms associated with the materials in the gate electrode, gate dielectric and Si channel, and can alter the overall capacitance of the gate stack, particularly if they have a thickness that is substantial relative to the gate dielectric. Additionally, these interfacial regions can be exploited to obtain desirable properties. The upper interface, for example, can be engineered in order to block boron outdiffusion from the poly-Si gate. The lower interface, which is in direct contact with the CMOS channel region, must be engineered to provide low interface trap densities (e.g., dangling bonds) and minimize carrier scattering (maximize channel carrier mobility) in order to obtain reliable, high-performance device. Mobility degradation, relative to that obtained using SiO_2 (or SiON) gate dielectrics, associated with the incorporation of high-k gate dielectrics is an important issue currently under investigation and discussed further below.

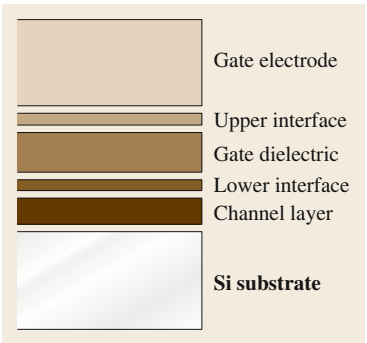


Fig. 27.8 Schematic of the important regions of the transistor gate stack. (After [27.14], AIP, with permission © 2001)

Reactions at either of these interfaces during the device fabrication process can result in the formation of a significant interfacial layer that will likely reduce the desired gate stack capacitance. Additionally, any suitable interfacial layer near the channel must result in a low density of electrically active defects ($\lessapprox 10^{11}/\mathrm{cm}^2\,\mathrm{eV}$ is often obtained for SiO_2) and avoid degradation of carrier mobility in the region near the surface channel.

The reduced capacitance can be seen by noting that the dielectric film that has undergone interfacial reactions results in a structure that essentially consists of several dielectric layers in series. If we suppose that the dielectric consists of two layers, then from electrostatics the total capacitance of two dielectrics in series is given by $1/C_{tot} = 1/C_1 + 1/C_2$, where C_1 and C_2 are the capacitances of the two layers. Thus, the lowest-capacitance layer will dominate the overall capacitance and also set a limit on the minimum achievable t_{eq} value. If we consider a dielectric stack structure such that the bottom layer (layer 1) of the stack is SiO_2, and the top layer (layer 2) is the high-k alternative gate dielectric, (27.8) is expanded (assuming equal areas) to

$$t_{eq} = t_{SiO_2} + \frac{k_{SiO_2}}{k_{high\text{-}k}} t_{high\text{-}k}\,. \tag{27.9}$$

From (27.9), it is clear that the layer with the lower dielectric constant (SiO_2 in this case) limits the ultimate capacitance of the MIS gate stack.

Of course, the actual capacitance of a CMOS gate stack for ultra-large-scale integration (ULSI) devices does not scale simply with $1/t$ due to parasitics, quantum-mechanical confinement of carriers, and depletion effects. Indeed, consideration of these important effects can result in much confusion on the definition of dielectric thickness as extracted from electrical measurements.

Parasitic resistances and capacitances associated with various portions of the transistor structure can result in an overall degradation of performance as defined by a delay time figure of merit [27.14, 17]. There exist several materials issues associated with the control of such parasitics including source/drain dopant-profile control, gate/contact sheet resistance minimization, etc. which are beyond the scope of this chapter.

Quantum-mechanical confinement effects on carriers in the channel region occur as a result of the large electrical fields in the vicinity of the Si substrate surface. These fields quantize the available energy levels resulting in the displacement of the charge centroid from the interface into the Si and at energies above the Si conduction-band edge. The extent of the penetration of the charge centroid is dependent upon the biasing conditions employed for the metal–insulator–semiconductor (MIS) structure and, for accurate estimates of transistor drain-current performance, the inversion capacitance measurement provides accurate determination of the equivalent oxide thickness [27.32–34].

Earlier MIS gate-stack structures employ heavily doped polycrystalline Si (poly-Si) as the *metal* gate electrode for CMOS transistors. Poly-Si gates were introduced as a replacement for Al metal gates in the 1970s for CMOS integrated circuits and have the very desirable property of a tunable work function for both n-MOS and p-MOS transistors through the implantation of the appropriate dopant into the poly-Si gate electrode.

However, in deep-submicron scaled devices, the *poly-Si depletion effect* impacts the overall gate-stack capacitance significantly, as the scaling results in an increased sensitivity to the effective *electrical* thickness resulting from all of the gate-stack component films, including the gate electrode. The depletion effect (which occurs in any doped semiconductor) is a result of the decrease in the density of majority carriers near the poly-Si/dielectric interface upon biasing, resulting in an increased depletion width in the poly-Si [27.27, 35]. This results in a voltage drop across the depletion region, rendering a smaller voltage drop across the remaining portions of the gate stack (i. e., the dielectric and the substrate). As a result, the (low-frequency) capacitance under strong inversion biasing conditions is significantly lower than that obtained under accumulation (Fig. 27.9). Moreover, the inversion-layer charge in the substrate is reduced from that possible in the ideal case where no depletion layer exists in the gate (i. e., an ideal metal conductor for the gate electrode). This reduction in the inversion-layer charge in turn reduces the ideal transistor drive current. As the dielectric thickness is reduced in the scaling process, this effect becomes a serious limitation for transistor drive current.

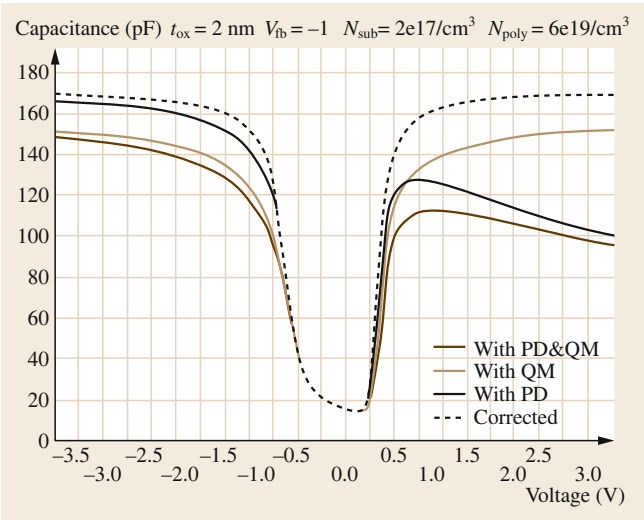


Fig. 27.9 Capacitance–voltage curves demonstrating the effects of poly-Si depletion and quantum-mechanical behavior for scaled transistors. N_{sub} and N_{poly} denote the dopant density for the substrate and poly-Si gate, respectively. (After [27.36])

To solve this problem, a return to metal gate electrode materials has been invoked to eliminate any depletion region in the gate, and thus the interaction of any dielectric with various metal gate candidates is of interest. Compatibility of dielectric materials with metal gate materials is discussed later in this chapter.

When a capacitance–voltage curve is properly corrected for these quantum-mechanical and dopant depletion effects, as seen in Fig. 27.9, one can accurately extract the equivalent oxide thickness (EOT), t_{eq} [27.36]. This thickness definition is in contrast to that derived directly from the raw data in a capacitance–voltage (C–V) measurement, which is termed the *capacitance equivalent thickness* (CET > EOT), or the *physical thickness*, which can be determined by non-electrical characterization methods such as ellipsometry or high-resolution transmission electron microscopy, as seen in Fig. 27.10 [27.36, 37]. Please note that thicknesses derived from quantum-mechanical-corrected accumulation capacitance data are also often abbreviated as t_{qm}.

Leakage Current

Of course, physically thinning the dielectric layer to the nanometer regime raises the prospect of quantum-mechanical tunneling through the dielectric – often referred to as *leakage*. To minimize gate leakage (tunneling) currents, quantum mechanics indicates that the gate dielectric must be sufficiently thick and have a sufficient energy barrier (band offset) to minimize the

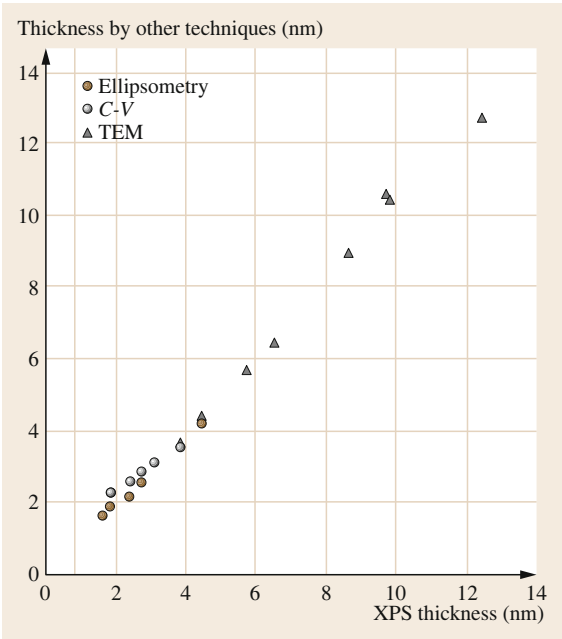


Fig. 27.10 Metrology results for various measurement techniques of the thickness of thin SiO_2 films. X-ray photoelectron spectroscopy (XPS) measured thicknesses are compared to thicknesses derived from high-resolution transmission electron microscopy (HRTEM), *C–V* and ellipsometry. (After [27.37], AIP, with permission © 1997)

resultant tunneling current. Consider the band diagram shown in Fig. 27.11, where the electron affinity (χ) and gate electrode work function (Φ_M) are defined (q is the charge). For electrons traveling from the Si substrate to the gate, this barrier is the conduction-band offset, $\Delta E_C \cong q\,[\chi - (\Phi_M - \Phi_B)]$; for electrons traveling from the gate to the Si substrate, this barrier is Φ_B. The effect of the height of the energy barrier and the film thickness on tunneling current can be seen by considering the expression for the direct tunneling current where electron transport occurs through a trapezoidal energy barrier [27.16]

$$J_{dt} = \frac{A}{t_{diel}^2}\exp\left(-Bt_{diel}\right) \times \left\{\left(\Phi_B - \frac{V_{diel}}{2}\right)\exp\left[\sqrt{\left(\Phi_B - \frac{V_{diel}}{2}\right)}\right] - \left(\Phi_B + \frac{V_{diel}}{2}\right)\exp\sqrt{\left(\Phi_B + \frac{V_{diel}}{2}\right)}\right\} . \quad (27.10)$$

Here A is a constant, $B = 4\pi/h(2m^*q)^{1/2}$, Φ_B is the potential-energy barrier associated with the tunneling

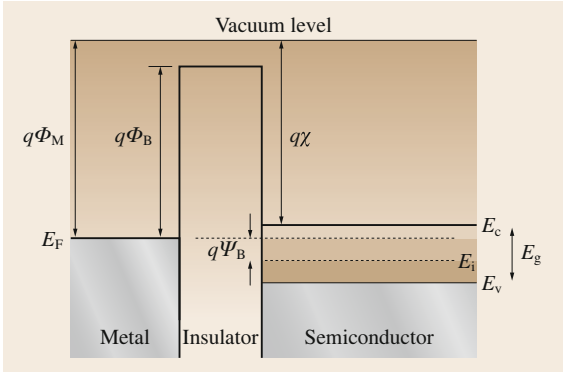


Fig. 27.11 Energy-band diagram for an MIS stack structure

process, t_{diel} is the physical thickness of the dielectric, V_{diel} is the voltage drop across the dielectric, and m^* is the electron effective mass in the dielectric. From (27.10), one observes that the tunneling (leakage) current increases exponentially with decreasing barrier height and thickness for electron direct-tunneling transport. In the context of charge transport through the dielectric, electrically active defects (electron or hole *traps*) can result in charging of the dielectric, which in turn deleteriously affects the electric field in the channel region and therefore mobility. Conduction mechanisms through such fixed charge traps can also be evaluated through electrical characterization techniques [27.16]. In the case of dielectrics layers, tunneling transport has also been previously examined [27.38].

The introduction of a dielectric material that exhibits a suitable energy barrier and thickness, while performing electrically as a thin SiO_2 layer, is now a key research and development goal for the industry to continue the scaling trend. It is also clear that the gate dielectric must exhibit adequate thickness uniformity and integrity over the surface of the wafer. The reduction of this leakage current impacts important circuit properties such as stand-by power consumption, as shown in Fig. 27.12. With the minimization of leakage current as a key driving force, alternate gate dielectric materials appear to be required for low-power CMOS device technologies in the near future and are therefore the subject of intense research.

Reliability

Any gate dielectric material must possess the ability to enable a specified, stable operational lifetime for the transistor. Key areas of concern in this regard includes defect generation/mitigation resulting in dielectric breakdown phenomenon, charge formation within the dielectric and at the associated interfaces, and resistance to energetic (*hot*) carrier interactions. This is often

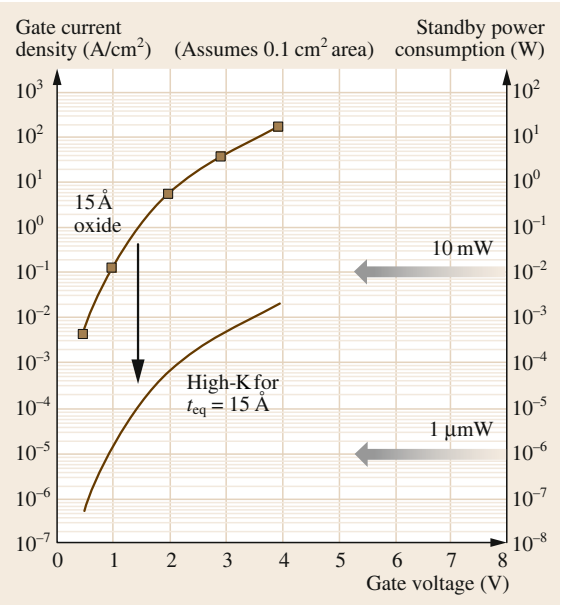


Fig. 27.12 Reduction of leakage (tunneling current) from the incorporation of a higher permittivity gate dielectric. (After [27.14], AIP, with permission © 2001)

Table 27.2 Properties of SiO_2

Geometrical parameters	
Si–Si bond length	3.12 Å
Si–O bond length	1.62 Å
O–O bond length	2.27 Å
Mean bond angle	144° (tetrahedral bonding)
Bond angle range	Si–O–Si: 110 to 180°
Density (g/cm³)	
Thermally grown (fused silica)	2.20
Quartz	2.65
Index of refraction (optical frequencies)	
Thermally grown (fused silica)	1.460
Quartz	1.544
Quasi-static dielectric constant (≤ 1 kHz)	
Thermally grown (fused silica)	3.84
Quartz	3.85
Band gap	
Thermally grown (fused silica)	8.9 eV
Quartz	≈ 9.0 eV
Electrical breakdown strength	
Thermally grown (fused silica)	10 MV/cm
Quartz	≈ 10 MV/cm

evaluated in the context of accelerated lifetime testing through electrical stressing in combination with thermal stressing of devices. Reviews on these topics can be found in this handbook as well as other sources [27.3, 27, 39, 40].

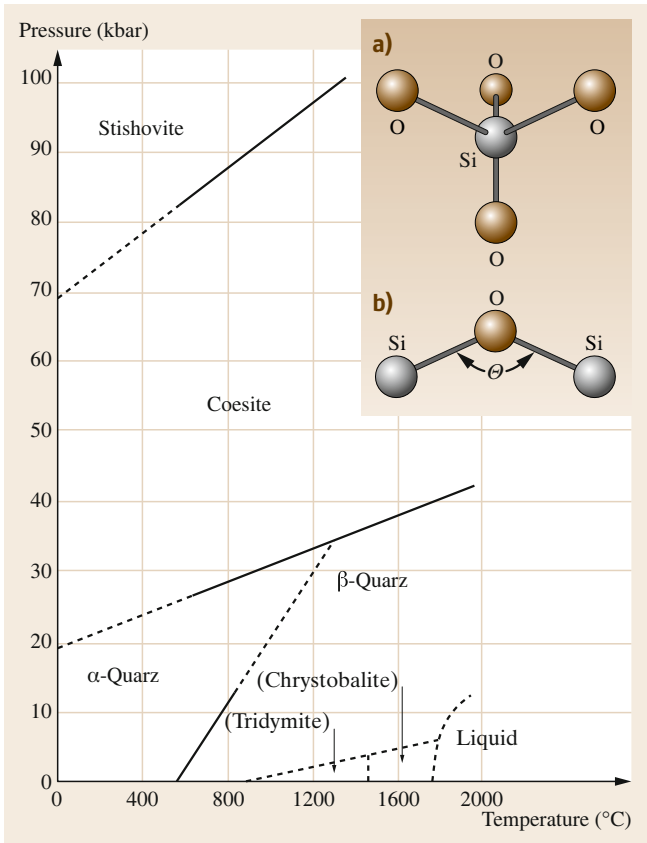


Fig. 27.13a,b Phase diagram for SiO_2. *Inset*: bonding configurations for the O–Si–O system. **(a)** Tetrahedral unit for the Si–O bonding arrangement. **(b)** Bond angle θ defined for O–Si–O bonds. (After [27.41], IOP Publishing © 1994, with permission)

27.2.3 Silicon Dioxide

As articulated by *Hummel*, silicon has an amazing number of desirable materials (and therefore electronic) properties that no other electronic material can rival [27.42]. Silicon constitutes 28% of the Earth's crust, it is nontoxic, exhibits a useful band gap ($E_G = 1.12$ eV) for applications at room temperature, and can be grown in crystalline form (from abundant sand, i. e., SiO_2). However, it can be argued that the ability of Si to form the stable, high-quality insulator, SiO_2, may be the primary reason that Si-based transistor technology has dominated the industry since the 1960s [27.5, 43]. This compound has a number of material properties that result in outstanding electrical performance. Some of these properties are summarized in Table 27.2. Over the last 25 years, abundant compendia of the properties of SiO_2 with particular emphasis on materials and electrical properties have been published, as well as

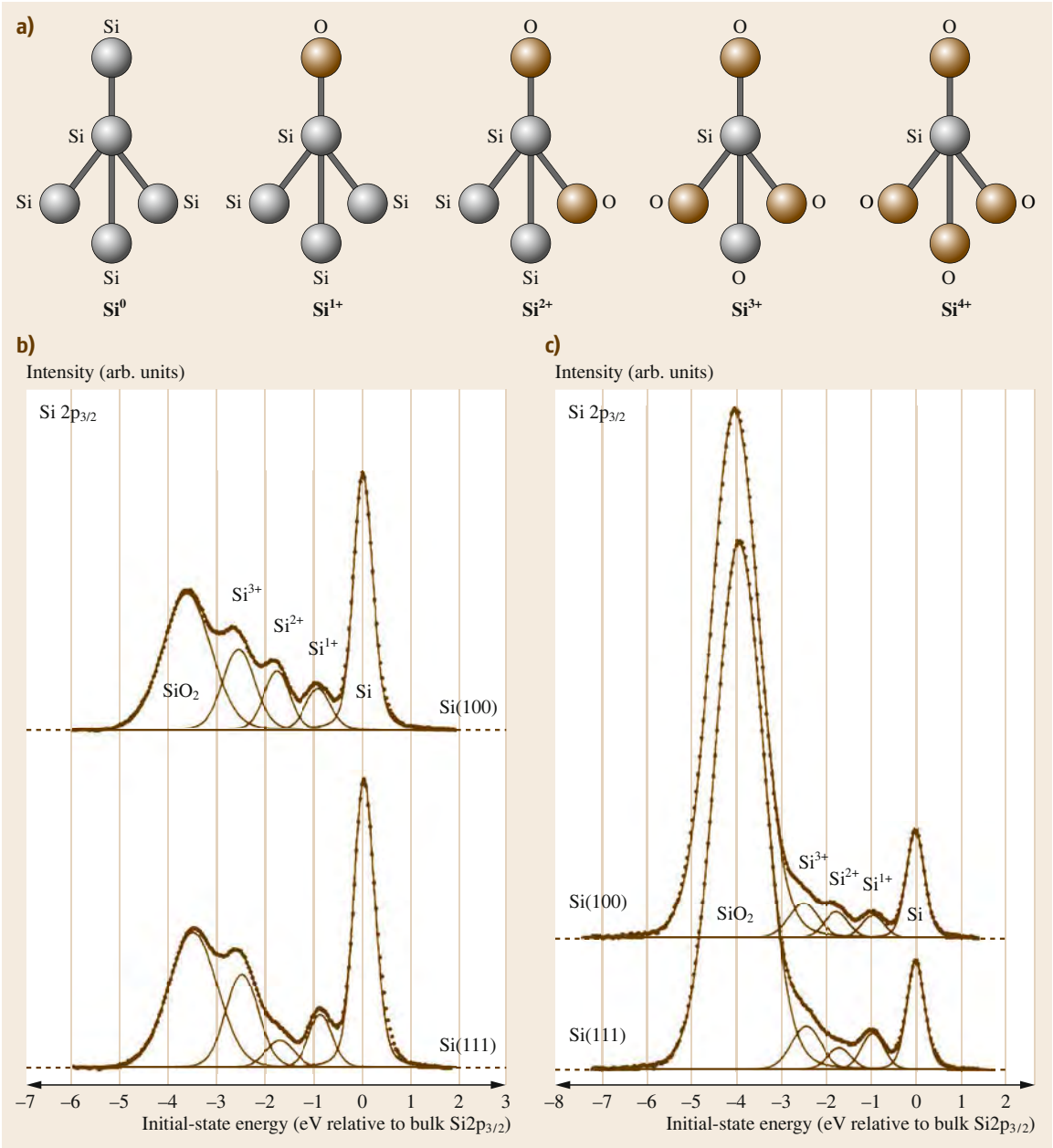


Fig. 27.14 **(a)** Oxidation states and local chemical bonding variations associated with the Si−SiO_2 interface. After [27.41]. High-resolution XPS results demonstrating the existence of such bonding variations at the interface for substrate orientation and with **(b)** 5 Å and **(c)** 14 Å SiO_2 film thickness. Photon energy $h\nu = 130$ eV. (After [27.44], APS, with permission © 1988)

conferences which have focused almost exclusively on SiO_2 [27.45–54].

It is noteworthy that the early development of transistor technology utilized the semiconductor Ge, which has a rather unstable oxide, GeO_2, that is water soluble. The utilization of crystalline Si and its stable oxide presented a superior solution for the practical manufacturing of the planar transistor and the associated integrated circuit of the era [27.43]. It was the utilization of the properties of SiO_2 that resulted in the dominance of Si-based device technology over Ge-based technology in the 1960s.

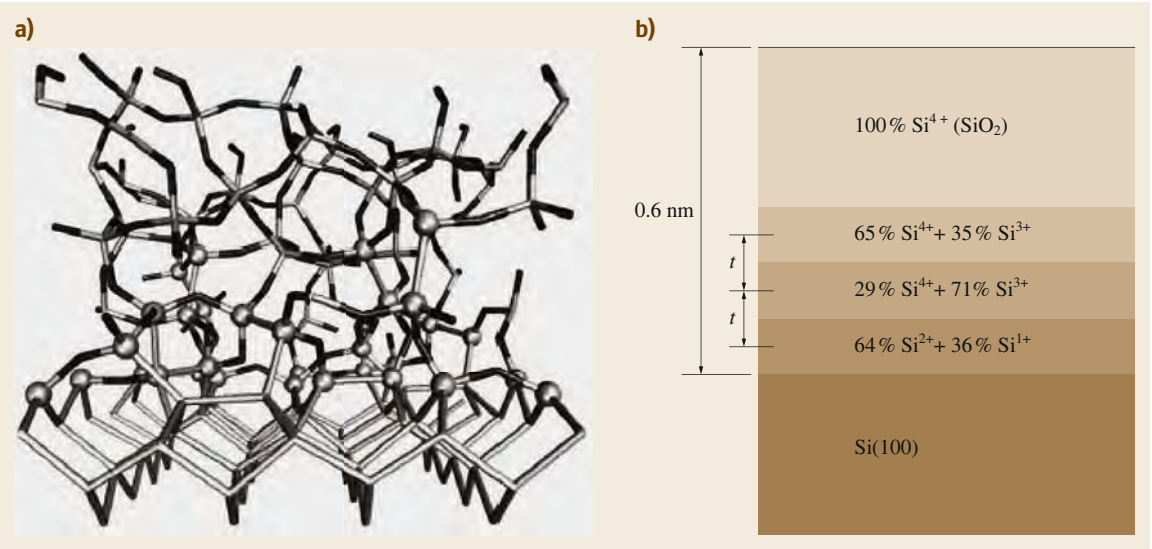


Fig. 27.15 **(a)** First-principles simulations of the SiO_2/Si interface. After [27.55] © 2002 Elsevier. **(b)** Angle-resolved XPS results for the SiO_2/Si interface showing the detected O−Si−O bonding environments. The single layer thickness $t = 0.137$ nm associated with the bonding distance in Si in this evaluation. (After [27.56], © 2001 by APS, with permission)

Bonding Arrangements for Thin SiO_2

As can be seen in Fig. 27.13, the phase diagram for SiO_2 has a number of crystalline phases and associated polymorphisms [27.41, 53]. In the context of accessible phase space appropriate for typical CMOS processes (pressure ≈ 1 bar and $T \leq 1100\,^{\circ}\mathrm{C}$), the quartz and tridymite phases are stable [27.57]. High-pressure phases are also well studied [27.58].

The bulk SiO_2 tetrahedral bonding coordination (inset, Fig. 27.13) results in an average Si−O−Si bond angle of $\theta \approx 145^{\circ}$ [27.59] comparable to that observed for vitreous silica [27.41, 60]. As noted originally in the classic work by *Zachariasen*, the bonding interconnections among the various SiO_4 tetrahedra can result in an amorphous SiO_2 film consisting of locally ordered ring structures commonly observed in silica glasses [27.51, 53, 61]. The numerous available variations of the ring structure provide substantial flexibility and minimize bond strain and the density of defects at interfaces [27.62].

For the purposes of microelectronic applications, we first consider so-called *thermally grown* SiO_2. As the name implies, a SiO_2 layer is formed on Si from a thermally activated oxidation process (often referred to as a *thermal oxide*) rather than simply a (low-temperature) deposition process. Thermal oxide films are stable enough to endure high-temperature ($\approx 1000\,^{\circ}\mathrm{C}$) post-implantation/dopant activation anneals. These films have useful electrical properties including a stable (thermal and electrical) interface with minimal electrically-active-defect densities ($\approx 10^{10}\,/\mathrm{cm}^2$) as well as high breakdown strength ($\approx 10\,\mathrm{MV/cm}$) over a large area ($\approx 1\,\mathrm{cm}^2$) [27.16]. The breakdown strength is important for high reliability of devices fabricated with a dielectric, as the scaling of devices typically results in somewhat higher electric fields across the oxide.

Recent experimental results on thin thermal SiO_2 [27.36, 39, 63] combined with modern computational modeling [27.55, 59] of the defects associated with the SiO_2/Si(001) interface, which is the dominant microelectronic materials system, and *bulk* SiO_2 [27.64] has provided a reasonably consistent picture of the known defects and interface structure.

High resolution x-ray photoelectron spectroscopy (XPS) evidence, coupled with other experimental and theoretical modeling results, appears to support the presence of SiO_x ($x = 4$) species (often called *suboxides*) detected as a chemical shift in the associated Si oxidation states (Fig. 27.14) [27.44, 65–67].

More recent high-resolution angle-resolved XPS analysis of ultrathin (6 Å) SiO_2 has resulted in a further understanding of the average oxidation state observed with depth along the interfacial transition region [27.56, 68]. From these experimental results, models of the interface have been constructed from first-principles methods that reproduce the experimental results, as shown in Fig. 27.15 [27.55, 59]. As seen in Fig. 27.15, the transition region between the Si substrate and the SiO_2 consists of interfacial suboxide species of varying abundance (and therefore density). These results appear

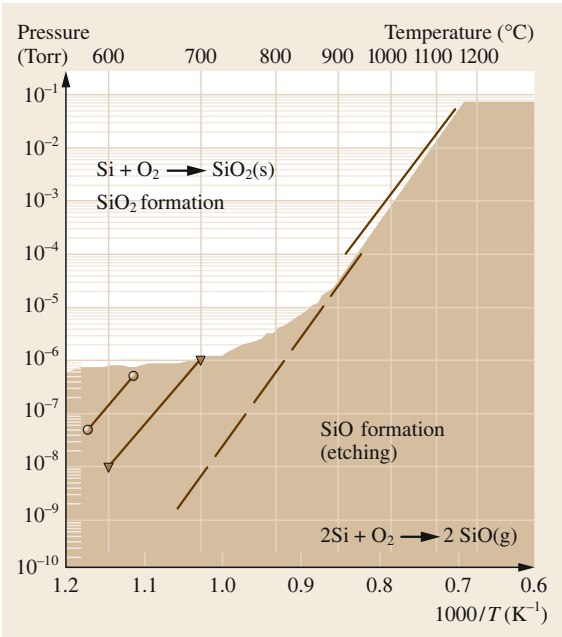


Fig. 27.16 Pressure–temperature phase diagram for thin SiO_2/Si. (After [27.76], © 1997 AIP, with permission)

to be in general agreement with recent high-resolution transmission electron microscopy as well [27.69].

Oxidation Kinetics of Si

The oxidation kinetics of Si have also been exhaustively studied and reviewed [27.27, 39, 70]. Two regimes for oxidation are generally characterized as *passive* and *active* in nature. The passive oxidation regime refers to the low-temperature/high-pressure region of the pressure–temperature (p–T) phase space shown in Fig. 27.16 [27.39, 71–76]. In this portion of the p–T phase space, the reaction between Si and O_2 proceeds as a surface reaction resulting in a thin (≤ 2 nm) SiO_x layer, where $0.5 \leq x \leq 2$. Decreasing pressure and increasing temperature during the reaction results in both Si etching and oxidation – the so-called *active* regime – and the formation of thicker SiO_2 films.

The clever utilization of isotopic labeling techniques coupled with depth profiling from ion scattering [27.77] and nuclear reaction analysis [27.70] has generally confirmed that molecular oxygen (O_2) is the mobile species that interstitially diffuses through a growing oxide and only reacts at the Si interface to form the SiO_2 on the Si surface, as first empirically proposed by *Deal* and *Grove* in 1965 for SiO_2 thicker than about 4 nm [27.78, 79]. The details of the oxidation process for thin films deviates from the Deal–Grove model, and is discussed thoroughly in the review by *Green* et al., for example [27.39].

Defects for Thin SiO_2

In addition to the electrical breakdown strength associated with SiO_2, the physicochemical properties of the interface between the substrate Si and SiO_2 plays a major role in establishing the observed electrical properties [27.63, 67]. Like all interfaces of solids, the interface between bulk Si and bulk SiO_2 naturally results in the presence of point defects (Fig. 27.17). Examples include strained bonds, atoms that are not fully bonded to neighboring atoms (so-called *dangling*), bonds and hydrogenic species bonded among the Si—O bonding network. These defects can result in energy levels in the band structure associated with an MIS device, and serve as *traps* for carriers (electrons and holes).

The dominant orientation for contemporary Si-based integrated circuit technology substrates is Si(100). This is, of course, no accident and the choice of the substrate orientation has much to do with the resultant SiO_2/Si interface. Early in the development of Si-based transistors, it was observed that the density of interface states for SiO_2 grown on Si(100) was significantly lower than that observed on Si(111) [27.80–84]. This difference in interface state behavior is attributed to the density of unsatisfied (*dangling*) bonds at the SiO_2/Si interface.

Techniques sensitive to the presence of such point defects, such as electron (paramagnetic) spin resonance (EPR or ESR) [27.85] have provided valuable insight on the atomic nature of the various defects in the bulk and interfacial regions of SiO_2 [27.86–91]. For such defects, although the electrical manifestation of significant densities ($\geq 10^{10}$ /cm^2 eV) of such *trap states* is easily detected by conventional electrical characterization methods, the unambiguous identification of the moiety responsible for the observed electrical behavior is very difficult using only electrical characterization. From techniques like ESR employed over the last 30 years, point-defect structural models have been established in the bulk regions of SiO_2 and Si, as well as at the interface (including various substrate orientations). A schematic illustration of various known point defects for the Si/SiO_2 system is shown in Fig. 27.17. The major classes of defects that have been examined over the last 35 years by ESR have included interfacial dangling bonds (P_b centers) and bulk defects in SiO_2 (E′ centers) [27.88, 91].

The early discovery and study of P_b centers [27.92–94] eventually led to the interpretation of their ESR signals as due to dangling bonds and the realization of their importance to MOS device physics [27.87, 88, 95, 96]. It has also been noted that P_b-center defects are generated for unannealed, thermally stressed and radiation-damaged MOS structures, and moreover can account for roughly 50% of the density of interface de-

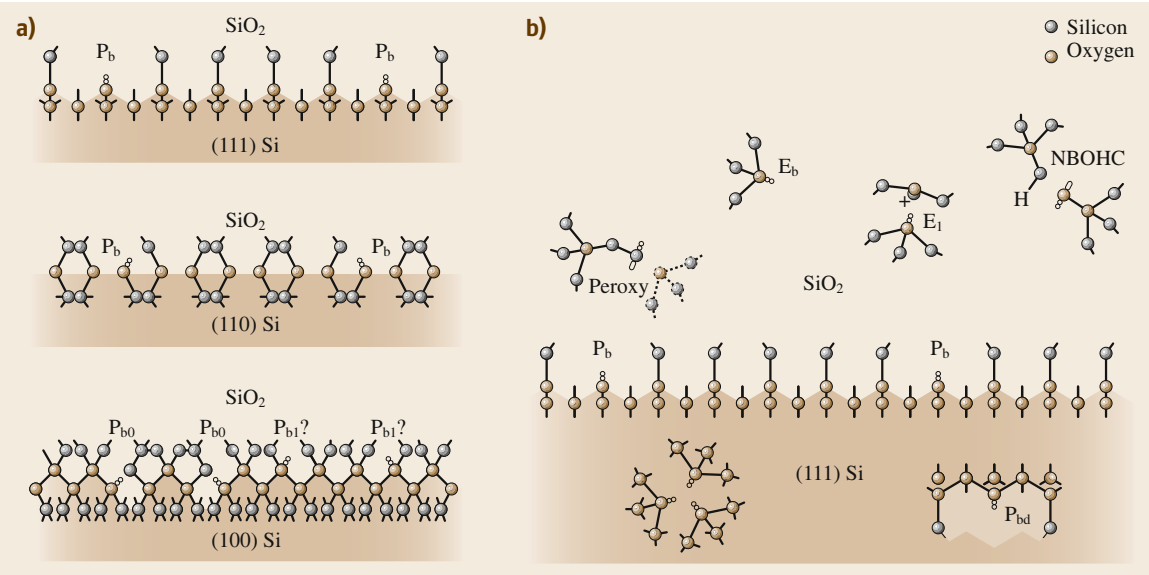


Fig. 27.17a,b Schematic representations of **(a)** the various point defects for the SiO_2/Si interface and **(b)** bulk SiO_2. (After [27.41], IOP, with permission © 1994)

fect states (D_{it}) for Si/SiO_2 interfaces [27.41, 97, 98]. It has been demonstrated that the interface state density can be directly proportional to the density of P_b centers (dangling bonds) for the SiO_2/Si(111) and SiO_2/Si(100) interfaces, as seen in Fig. 27.18. Recently, further ESR work has been performed to establish the detailed structure of the defect on Si(100) and Si(111) [27.99–106].

Mitigation of Defects for Thin SiO_2

Synopses of the early MOS transistor work [27.108–110] examining the importance of the dielectric–semiconductor interface are available [27.3, 43, 111]. Indeed, the close connection between interfacial chemical behavior and electrical device performance was investigated and realized in pioneering surface science work by *Law* and coworkers on the reaction of gaseous species with atomically clean Ge [27.112] and Si surfaces [27.113–115]. In particular, the relative interface state (dangling-bond) density, as measured by the areal density of surface reactions with technologically important species such as H_2O, H_2, O_2, CO, CO_2, provided important clues on the control of the Si/SiO_2 interface and the resultant electrical properties reported 5–10 years later.

The control of the density of interface defects through the chemical reaction of species, introduced mainly through gaseous exposure at elevated temperatures, has proven to be fruitful. As noted above, early surface science work [27.113] indicated the rapid reaction of H_2 with the atomically clean Si surface. By definition, the atomically clean Si surface is saturated

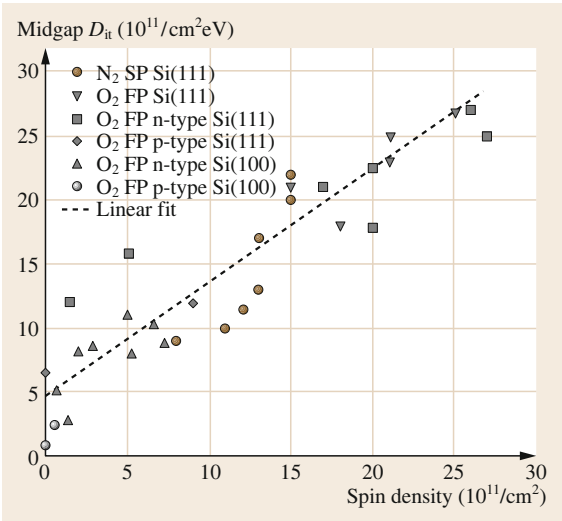


Fig. 27.18 Correlation of mid-gap interface state density with detected spin density from ESR measurements for SiO_2/Si(111) and SiO_2/Si(100). FP and SP denote *fast* and *slow* pullout conditions for the wafers from the furnace after thermal treatments. (After [27.96, 107], with permission)

with dangling bonds. The reaction of these bonds with H_2 results in the chemical passivation of the surface – that is, the reaction of the surface to eliminate the dangling bonds and produce a H-terminated Si surface. In the context of the MOS structure [27.48, 116], it was realized early that annealing the structure in

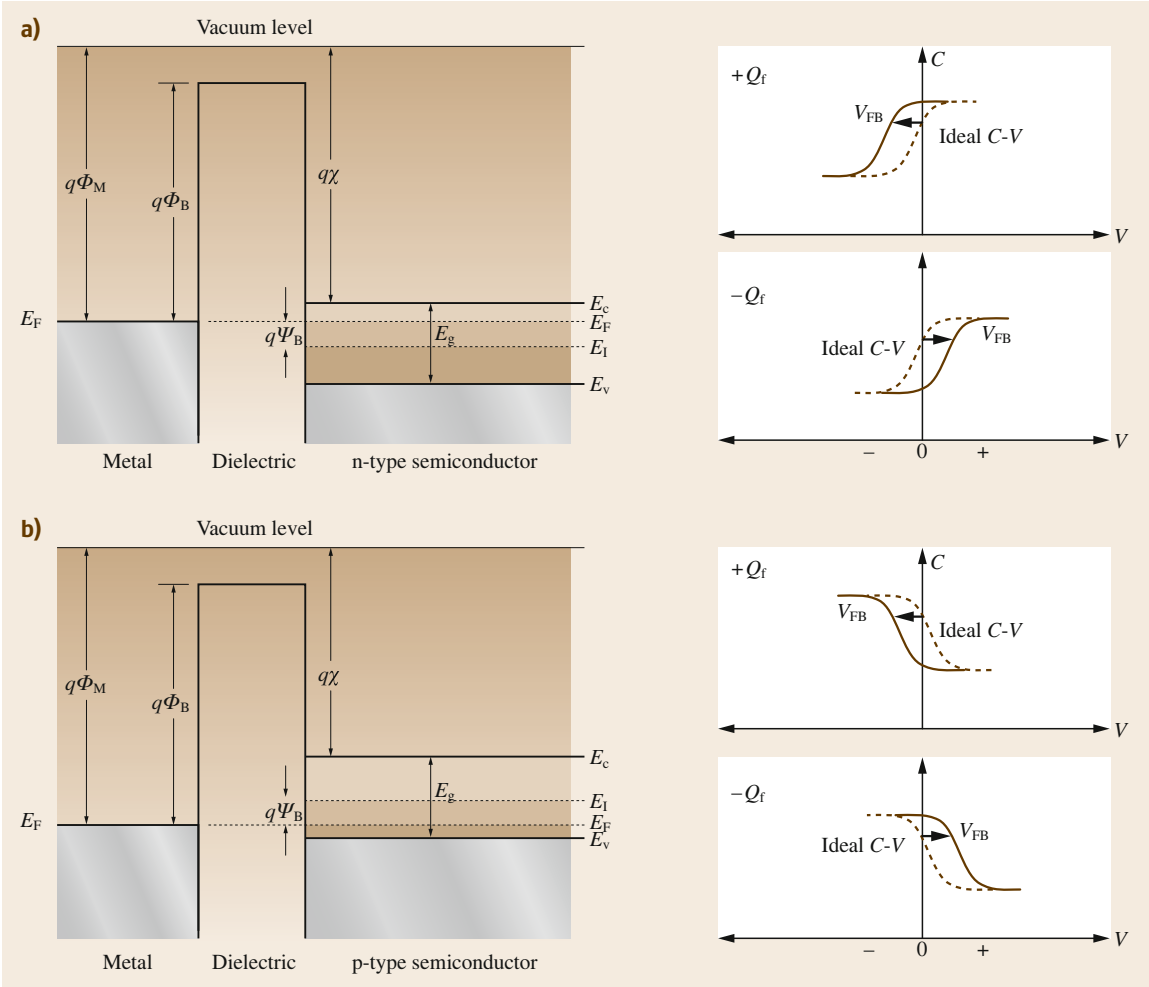


Fig. 27.19a,b The effect of fixed charging of the capacitor dielectric, demonstrating the shift of the associated flat-band voltage for MIS structures with **(a)** n-type and **(b)** p-type substrates. (After [27.14], © 2001 AIP, with permission)

ambient H_2 resulted in beneficial transconductance performance [27.117–120]. The use of anneals in forming gas ($N_2 : H_2$ of various mixtures, typically 90–95% N_2 : 10–5% H_2) were originally developed to improve electrical contacts for the gate and source/drain regions of the metal/oxide/semiconductor field effect transistor MOSFET. *Balk* and *Kooi* established the effect of hydrogen ambients on the reduction of fixed charge in the Al/SiO_2/Si MOS structure [27.119, 121].

Subsequent studies demonstrated that anneals of the MOS structure in hydrogenic environments, typically at 400–500 °C for 30–60 min, results in the passivation of dangling bonds at the interface [27.96, 99–104, 122–125].

Hydrogen incorporation into the bulk of SiO_2 can, however, also be detrimental to dielectric performance in MOS capacitors and FETs [27.41, 126]. For example, silicon bonded to hydroxyl (silanol) species have been identified with fixed charge in the oxide, resulting in undesirable, irreversible voltage shifts. This charge induced shift is shown in Fig. 27.19 for n-type and p-type MIS diode structures from their associated capacitance–voltage response [27.3, 4, 14]. (An analogous threshold-voltage shift would be observed in a transistor turn-on characteristic.) More complicated effects such as negative-bias temperature instability (NBTI), where an increase of the density of fixed charge (Q_f) and interface trap (Q_{it}) density is noted with time upon thermal stress and/or under negative bias, has been attributed to H_2O-induced depassivation of Si dangling bonds (i. e., generation of P_b centers) at the Si/SiO_2 interface [27.41, 127]. Defects generated by radiation damage have also been extensively investigated [27.41, 128] as well as interactions

with annealing ambients such as vacuum [27.129] or SiO [27.123].

Dielectric Breakdown and Reliability of SiO_2

As noted in Sect. 27.2.2, the reliability of dielectrics is obviously an important phenomenon to control. The scaling of microelectronic devices necessarily results in increased stress on the dielectric due to the higher electric fields placed across the dielectric film. As a result, power-supply voltages are also scaled to minimize the likelihood of catastrophic (*hard*) breakdown, which would generally result in complete failure in the associated integrated circuit. As noted by *Hori*, such a catastrophic breakdown phenomenon is dependent upon the presence of defects in the dielectric, and thus requires a statistical analysis of many devices (and therefore films) to enable a reliability prediction for the dielectric layer [27.16]. Evaluation of the breakdown is often performed under the conditions of constant-field stress until a time at which breakdown is observed (called time-dependent dielectric breakdown (TDDB)). Breakdown fields for thermally grown SiO_2 with thicknesses larger than $\approx 10\,\text{nm}$ often exceed $10\,\text{MV/cm}$, providing outstanding insulator prosperities for microelectronic applications. Various models have also been developed to accelerate such testing to perform reliability predictions. A concise review of these is offered in the literature as well [27.16, 27]. Defects such as film nonuniformity, bond stress, surface asperities, contamination, and particulates embedded in the film or the Si substrate are a few examples of potential causes of catastrophic breakdown phenomenon.

With the scaling of the SiO_2 gate dielectric layer to thicknesses well below 10 nm, extensive evaluations have been made of the various models that attempt to predict the reliability of SiO_2 in regard to breakdown. Often, a *soft* breakdown can be observed, where the film exhibits a sudden increase in conductivity, but not to the same degree as that observed in hard breakdown phenomenon. In spite of the many years of research on SiO_2 thin films, the nuances of the dependence of voltage acceleration extrapolation on dielectric thickness and the improvement of reliability projection arising from improved oxide thickness uniformity, have only recently become understood, despite decades of research on SiO_2 [27.130].

27.2.4 Silicon Oxynitride: SiO_xN_y

The introduction of nitrogen in SiO_2, often described as *silicon-oxynitride* (or more simply SiON) has provided the opportunity to maintain the scaling expectations for integrated circuits while minimizing the process variations associated with the modification of the gate

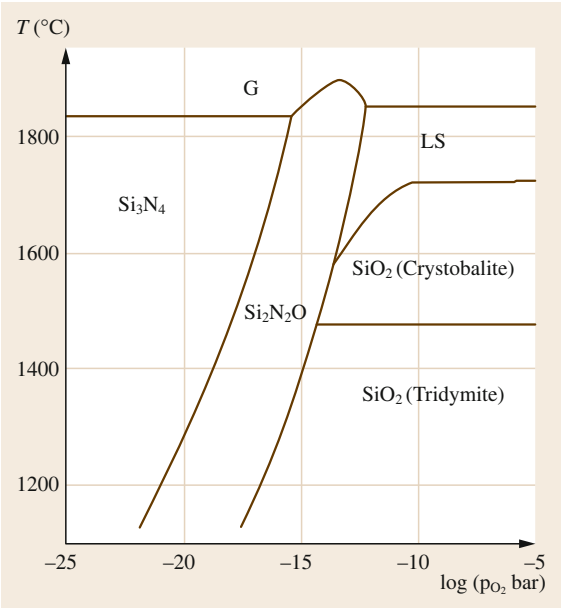


Fig. 27.20 Calculated phase diagram for the growth of Si–O–N for various oxygen partial pressures. (After [27.133], with permission)

dielectric material [27.16, 39, 70, 131]. (It should be noted that engineers and technologists often loosely refer to SiON films simply as oxide or nitride – frequently leading to confusion on the detailed chemical composition of the film.) Introduction of N into SiO_2 has been accomplished by thermal treatments (in N_2, N_2O, NO and NH_3) as well as plasma treatments [27.16, 132]. The bulk phase diagram for the SiON system is presented in Fig. 27.20, which indicates that, under equilibrium conditions, Si_2N_2O is the only stable SiON species and SiO_2 and Si_3N_4 would not coexist [27.133]. As noted by *Green* et al., N would not be expected to incorporate into SiO_2 whenever an even very small partial pressure of oxygen $P_{O_2} > 10^{-20}\,\text{atm}$ is present. However, the nonequilibrium surface reaction kinetics and/or a reduction in interfacial bond strain likely plays a key role in the tendency for interfacial N incorporation for SiON films [27.39].

The dielectric constant and band gap as a function of the SiO_2 content are presented in Fig. 27.21 [27.134]. As noted in Sect. 27.2.2, it is the increase in the dielectric constant that increases the capacitance of an associated MOS structure. Thus, for a SiON film with $k = 7$, a 2 nm SiON film would ideally exhibit electrical behavior (such as capacitance and leakage-current behavior) equivalent to a 1.1 nm SiO_2 film, according to (27.8). The data in Fig. 27.21 implies that the index of refraction (at 635 nm) varies between $n(SiO_2) = 1.46$ and $n(Si_3N_4) = 2$, indicating that optical techniques

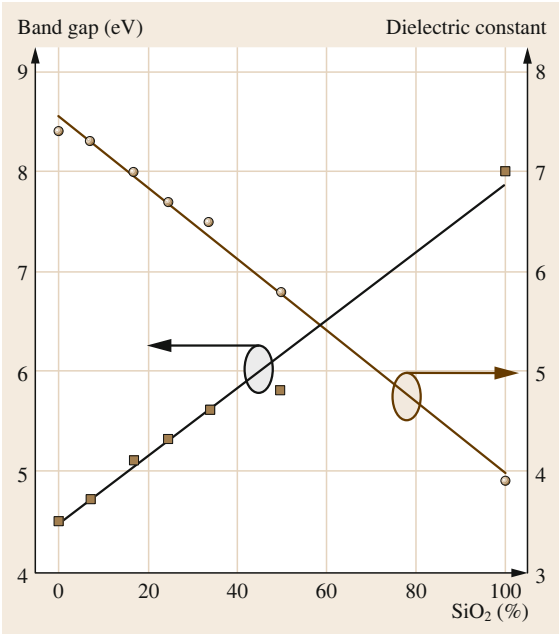


Fig. 27.21 Dependence of permittivity and insulator band gap of SiO_xN_y with SiO_2 content. (After [27.134], with permission)

such as ellipsometry can provide very useful information for such films.

Like the SiO_2/Si system, point defects have also been studied for SiON films on silicon [27.135]. *Lenahan* and coworkers examined the role of strain at the interface and the impact on the associated dangling bonds [27.136]. More recently, work on the effects of hydrogen annealing has also examined the SiON system [27.137].

It is also noted that the SiON material system also exhibits useful diffusion-barrier properties. This property is utilized to inhibit uncontrolled dopant diffusion from the polysilicon gate electrode through the dielectric layer and into the channel region of transistors, for example. The formation of SiON films from the reaction of NO or N_2O with Si, as well as nitridation of Si and SiO_2 from N_2 and NH_3 exposure, has been examined in detail and has been summarized in recent reviews [27.39, 70].

For transistor gate dielectric applications, a controlled variation in the depth concentration profile for N is often required throughout the dielectric to maintain the improved device reliability and mobility associated with SiON thin films [27.27]. An example of such N profiles is shown in Fig. 27.22, where a somewhat reduced N content is noted in the vicinity of the Si channel/dielectric interface relative to that near the

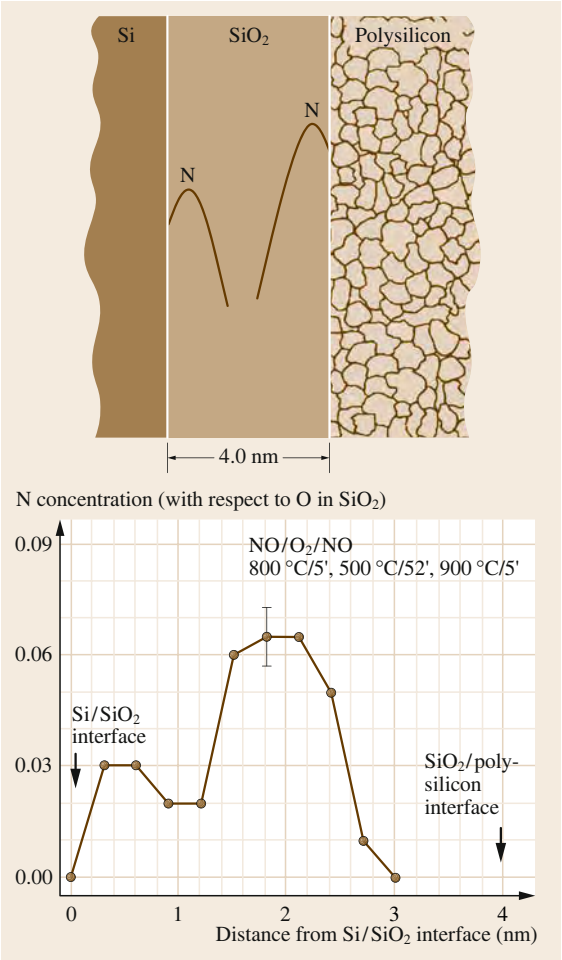


Fig. 27.22 N concentration depth profiles for tailoring the overall dielectric constant and materials properties of a SiON layer. (After [27.36], © 1999 Elsevier, with permission)

dielectric/poly-Si gate interface [27.36]. Such profiles are thought to control effects such as interfacial strain and roughness (which can degrade mobility) and inhibit dopant out-diffusion from the gate electrode. Incorporation of N in thin dielectric films using plasma processes has been studied extensively [27.132, 138–142].

Given the various leakage-current constraints afforded by power limitations and reliability, the scaling of SiON appears to be limited to $t_{eq} \approx 1$ nm, corresponding to the 65 nm node for high-performance microprocessor products [27.7]. As a result, the incorporation of alternative high-k dielectric materials has been achieved for scaled transistors over the last several years.

Table 27.3 Desirable properties for high-k gate dielectrics (65 nm node)

Physical property	Value/criteria	Electrical property	Value/criteria
Permittivity	15–25	Equivalent oxide thickness	< 1 nm
Band gap	> 5 eV	Gate leakage current (low power)	$< 0.03\ A/cm^2$
Band offset	> 1.5 eV	Gate leakage current (high performance)	$< 10^3\ A/cm^2$
Thermodynamic stability to 1000 °C	When in direct contact with Si channel	CV frequency dispersion	Minimal (meV)
Compatibility with metal electrodes	Mimimized extrinsic (pinning) defects	CV hysteresis	Minimal (meV)
Morphology control	Resistance to interdiffusion of constituents, dopants, and capping metals	V_T (V_{FB}) shift (fixed charge, defects, trapping, etc.)	Minimal
Deposition process	Suitable for high-volume production	Channel mobility	Near SiO_2 *universal* curve
Etching	Suitable control for patterning after processing/annealing	Interface quality	Near SiO_2 $D_{it} \approx 5 \times 10^{10}\ /cm^2\ eV$

Table 27.4 Comparison of relevant properties for selected high-k candidates. Key: mono. = monoclinic; tetrag. = tetragonal

Material	Dielectric constant (k)	Band gap E_G (eV)	ΔE_C (eV) to Si	Crystal structure(s) (400–1050 °C)
SiO_2	3.9	8.9–9.0	3.2–3.5[b]	Amorphous
Si_3N_4	7	4.8[a] – 5.3	2.4[b]	Amorphous
Al_2O_3	9	6.7[h] – 8.7	2.1[a] – 2.8[b]	Amorphous*
Y_2O_3	11[d] – 15	5.6–6.1[d]	2.3[b]	Cubic
Sc_2O_3	13[d]	6.0[d]		Cubic
ZrO_2	22[d]	5.5[a] – 5.8[d]	1.2[a] – 1.4[b]	Mono., tetrag., cubic
HfO_2	22[d]	5.5[d] – 6.0	1.5[b] – 1.9[c]	Mono., tetrag., cubic
La_2O_3	30	6.0	2.3[b]	Hexagonal, cubic
Ta_2O_5	26	4.6[a]	0.3[a,b]	Orthorhombic
TiO_2	80	3.05–3.3	≈ 0.05[b]	Tetrag. (rutile, anatase)
$ZrSiO_4$	12[d]	6[d] – 6.5	1.5[b]	Tetrag.
$HfSiO_4$	12	6.5	1.5[b]	Tetrag.
$YAlO_3$	16 – 17[d]	7.5[d]		**
$HfAlO_3$	10[e] – 18[g]	5.5–6.4[f]	2–2.3[f]	**
$LaAlO_3$	25[d]	5.7[d]		**
$SrZrO_3$	30[d]	5.5[d]		**
HfSiON	12–17[i,j]	6.9[k]	2.9[k]	Amorphous

* (γ-Al_2O_3 phase) has been recently reported [27.143], ** Onset of crystallization depends upon Al content,
[a] [27.144], [b] [27.145–147], [c] [27.148], [d] [27.149], [e] [27.150], [f] [27.148], [g] [27.151], [h] [27.152], [i] [27.153], [j] [27.154], [k] [27.155]

27.2.5 High-k Dielectrics

High-k dielectrics have now been incoprorated to further enable scaling of transistor (and memory capacitor) technology. In the context of the industry roadmap, the term high-k dielectric more generally refers to materials which exhibit dielectric constants higher than the SiON films ($k \approx 7$) described above.

Desirable Properties for High-k Dielectrics

Many materials systems were originally under consideration as potential replacements for SiO_2 and SiO_xN_y as the gate dielectric material for sub-100 nm (CMOS) technology. A systematic consideration of the required properties of gate dielectrics indicates that the key desirable properties for selecting an alternative gate dielectric include: permittivity and the associated band gap/alignment to silicon as well as mobility, thermodynamic stability, film morphology, interface quality, compatibility with the current or expected materials to be used in processing for CMOS devices, process compatibility, and reliability. The desirable properties are summarized in Table 27.3. Reviews of the developing field and recent research on high-k dielectric materials candidates are available [27.14, 156, 157].

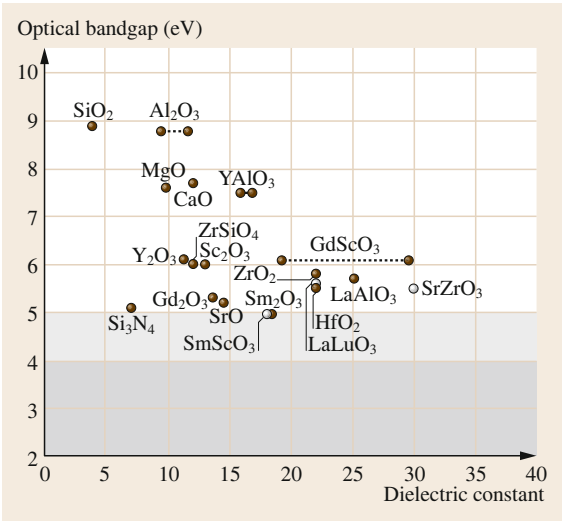


Fig. 27.23 Band gap versus dielectric constant for a number of dielectric materials. (After [27.149], MRS Bulletin, with permission)

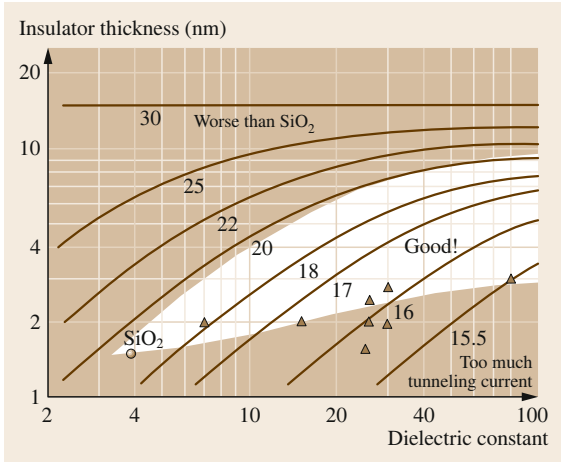


Fig. 27.24 Contours of constant scale length, which is proportional to the transistor gate length, versus dielectric constant and insulator thickness, showing the useful design space for high-k gate dielectrics. Data points are rough estimates of the tunneling constraints for various high-k insulators. The depletion depth is assumed to be 15 nm. The useful design space will shrink with decreasing depletion depth. (After [27.21], © 2004 IEEE, with permission)

Many gate dielectric materials appear favorable in some of these areas, but very few materials are promising with respect to all of these properties. Indeed, many of these desirable properties are interrelated. Many of these materials have been examined over the last 20 years for capacitor applications, for example. With the introduction of Hf-based (and Zr-based) gate dielectrics in research in the mid to late 1990s [27.158–163] and then announced by the industry for 45 nm node production in 2007 [27.164, 165], it is well known that the industry has widely adopted Hf-based oxides, largely based upon the desirable properties mentioned above. Table 27.4 summarizes some of the measured or calculated properties associated with various high-k dielectric material constituents. One notes the tradeoff between the band gap (and therefore the conduction/valence-band offsets to those of Si) for these dielectrics and the associated dielectric constant, as seen in Fig. 27.23 [27.149].

The compatibility of alternative dielectrics with metal gate electrodes is also an important consideration [27.14, 30, 166]. The interfacial reactions between some gate electrode metals and gate dielectric is thought to lead to extrinsic states due to Fermi-level pinning, which shifts the threshold voltages for transistors to a fixed value. Another clear challenge is compatibility with other *gate-first* process steps that entail substantial thermal budgets ($\approx 1000\,^{\circ}\mathrm{C}$, ≤ 10 s) for conventional CMOS. Indeed, it was found that the development of a *replacement gate last* process was key for high-k dielectric integration in high-volume manufacturing [27.1]. A key innovation in the deposition process for the high-k dielectric was also employed: atomic layer deposition (ALD). This deposition process enables atomic layer conformaility over device structures – a key requirement for scaled planar and 3-D transistor (and trech capacitor) geometries utilized currently [27.1].

Mobility Degradation for Transistors with High-k Dielectrics

In the case of many of the alternate gate dielectric materials (mainly metal oxides) currently under consideration, the polarizability of the metal–oxygen (*ionic*) bond is responsible for the observed low-frequency permittivity enhancement. Such highly polarizable bonds are described to be *soft* relative to the less polarizable *stiff* Si−O bonds associated with SiO_2. Unlike the relatively stiff Si−O bond, the polarization frequency dependence of the M−O bond is predicted to result in an enhanced scattering coupling strength for electrons with the associated low-energy and surface optical phonons. This scattering mechanism can therefore degrade the electron mobility in the inversion layer associated with MOSFET devices. A theoretical examination of this effect is provided by *Fischetti* et al. where calculations of the magnitude of the effect indicate that pure metal–oxide systems, such as ZrO_2 and HfO_2, suffer the worst degradation, whereas materials which incorporate Si−O bonds, such as silicates, fare better [27.167]. In that work, it is also noted that the

presence of a thin SiO_2 interfacial layer between the Si substrate and the high-k dielectric can help boost the resultant mobility by screening this effect, although the maximum attainable effective mobility is still below that for the ideal SiO_2/Si system. These researchers further suggest that the effect is also minimized by the incorporation of Si−O in the dielectric, as in the case of pseudo-binary systems such as silicates. Indeed, the utilization of a thin, Si-oxide rich interface has demonstrated several advantages, including a higher quality high-k/Si interface [27.1].

At this time, work on *higher-k* dielectric materials are under exploration to enable the roadmap targets presented in Table 27.1. Combinations of ALD high-k layers are often employed to increase the overall gate stack capacitance, while minimizing the thickness within the constraints for leakage current [27.1].

It is also important to note that most of the high-k metal oxides listed in Table 27.4 crystallize at relatively low temperatures ($T \approx 500$–$600\,^{\circ}C$). An exception is Al_2O_3 where crystallization in thin films is observed at $T \approx 800\,^{\circ}C$ [27.168]. In the case of Hf–silicates, film morphology (viz. suppression of crystallization) and diffusion-barrier properties can be controlled through the incorporation of N, as in the case of SiO_2 [27.153, 160–162, 169, 170].

Suppression of crystallization is also observed for aluminates (depending upon the Al content) [27.171], but there are fewer studies of interdiffusion with the Si substrate available at this time. As noted above however for the case of Al_2O_3 thin films, Al penetration into the Si substrate has been reported upon high temperature thermal annealing budgets appropriate for dopant activation [27.172]. Such penetration could be a concern for any high-k dielectric incorporating Al as well if a high temperature gate stack processs is employed.

The morphology of the dielectric film may also impact on the propensity for capping layer constituents (from metal gate material layers) to diffuse through the stack as well. Metal gate materials can include well-known minority-lifetime killers such as Ni (in NiSi) as well as tunable-work-function alloys and layers [27.30, 166, 173, 174]. The potential for the interdiffusion of metals into the channel region is clearly undesirable.

The limits of a useful high-k dielectric constant value have also been examined. For example, the study by *Frank* et al. (Fig. 27.24) suggests that there exist a range of useful permittivity values as planar CMOS transistor structures are scaled [27.21]. It is noted that the physical thickness of the gate insulator should be less than the Si depletion length under the channel region, and that this results in a limited design space that enables further device scaling. Again, one notes that the dielectric constant cannot be arbitrarily increased without careful consideration of the complete transistor design.

27.3 Isolation Dielectrics

Dielectrics are also employed to electrically isolate various regions of an integrated-circuit technology. The dominant material for these applications is SiO_2. A method commonly employed to obtain such isolation in larger-scale IC technologies (generally with gate lengths larger than $0.25\,\mu m$) is the so-called local oxidation of silicon (LOCOS) technique, where regions of Si between various components are preferentially (thermally) oxidized [27.16, 27]. Such isolation oxides are typically several hundred nanometers thick and the tapered shape of the edge of the LOCOS isolation oxide near the transistor gate dielectric and source/drain region (often referred to as the *bird's beak*) is an important region to control during the device fabrication at these gate lengths. Thinning of the dielectrics in this region results in breakdown reliability concerns, as the electric fields in this region can be quite high. A *re-oxidation process* is often performed to improve the thickness and reliability of the SiO_2 in this region as well.

Scaling ICs beyond this gate-length regime however has required the placement of isolated regions utilizing deposited SiO_2 trench structures, as seen in Fig. 27.6. As summarized by *Wolf*, extensive work has been done to control the shape of the trench walls and the SiO_2 layer initially formed on these walls. Chemical vapor deposition methods are normally utilized for the dielectric deposition as this approach provides superior conformality in the trench structure. Filling the trench without void formation, controlling film stress, and chemical mechanical polishing properties are also important aspects that must be addressed in the fabrication process [27.7, 27]. Scaling of integrate circuits will result in a decrease of the area available for these isolation trenches, and so the increasing aspect ratio of the trench depth to width requires considerable attention in regard to trench filling. The control of the shape of the top corner regions of the trench structures is also an area of concern due to high-field reliability as well as the formation of essentially a parasitic

edge transistor. Initially, thermal oxidation methods were employed to round off the shape of the corner, but further scaling will require etch processing methods.

The fabrication of transistor source/drain regions utilizing implantation processes also utilizes a dielectric spacer layer that surrounds the transistor gate region (Fig. 27.6). This spacer provides isolation between the gate and source/drain regions, and also permits the control of the depth (and therefore profile) of the implanted dopant species – a so-called self-aligned dopant implantation process. Dielectrics used for spacer technologies include deposited SiO_2 and Si_3N_4, and the extent of the spacer dimensions from the gate is an important device parameter to control [27.16]. Further scaling of transistor structures will likely result in the need for elevated source/drain regions, and so process compatibility of the spacer material with the source/drain formation processes and high-k dielectrics will become an important consideration [27.7].

27.4 Capacitor Dielectrics

Capacitors are employed in a variety of integrated circuits including storage (memory) circuit elements and input/output coupling circuitry. Clearly, dielectrics are critically important in this application. In contrast to the MIS structure associated with transistors, capacitors are passive devices incorporating a metal–insulator–metal (MIM) structure. Early electrodes were composed of degenerately doped Si while more recent work focuses on integrating metals for modern devices.

27.4.1 Types of IC Memory

A dominant memory technology for the IC industry includes dynamic random-access memory (DRAM) in which capacitors play the essential role of storing charge, and thereby useful information. This type of memory element requires refreshing in order to maintain useful information, and the rate of refresh is fundamentally related to the dielectric associated with the capacitor structure and leakage of charge from that capacitor. As a result, this class of memory devices is called volatile. In contrast, a static random-access memory (SRAM) nonvolatile memory element required a capacitor which stores charge without the refresh requirement, and can preserve charge for many years. Again capacitor design must include the consideration of the dielectric [27.25].

In addition to DRAM capacitors, other types of capacitors are used in the back end of the transistor flow. These devices are MIM capacitors, and they typically reside between levels of metal interconnects, and serve as decoupling or radio-frequency (RF) capacitors for input/output functions. For these applications, the capacitance density is still of critical importance, but other factors must be considered, such as the linearity of the capacitance as a function of voltage and temperature. Memory is also more frequently becoming *embedded* with logic, enabling faster on-chip storage.

27.4.2 Capacitor Dielectric Requirements in View of Scaling

In brief, scaling the capacitor dimensions requires a tradeoff between the amount of stored charge required for a reliable memory element and the area occupied by the capacitor (and the associated transistors for the memory cell). From generation to generation of devices, the amount of stored charge has been kept roughly constant ($\approx 30\,\mathrm{fF/cell}$), even though dimensional scaling occurs, for bit detection (sense amplifier), retention and reliability reasons [27.27, 175].

As noted in (27.6) for a simple parallel-plate capacitor, the area associated with the capacitor as well as the dielectric permittivity play an important role. Maintaining the required capacitance with scaling was (and will be) accomplished by increasing the capacitor area, for a suitable dielectric thickness where leakage currents are severely limited; this can be accomplished by utilizing three-dimensional capacitor structures produced in the Si bulk, such as deep-trench capacitors [27.175, 176]. Clever and intricate processing methods to increase the capacitor area were adopted for several generations for well-established SiO_2 and Si–oxynitride dielectric capacitor materials, such as the implementation of hemispherical-grain (HSG) poly-Si and *crown* structures (Fig. 27.25). Eventually, practical concerns about cost and device yield make manufacturing such structures problematic. Thus, substantially increasing the dielectric constant of the insulating material was considered, and less-complicated structures can be fabricated. At this point, both stacked and trench capacitor structures are dominant in the industry. Indeed, the ability to fabricate deep-trench structures will help guide the need for the incorporation of alternative dielectrics. These structures often require the use of chemical vapor deposition (CVD) methods to ensure conformality over the stacked structure topography or deep within trenches that have large aspect ratios. More recently,

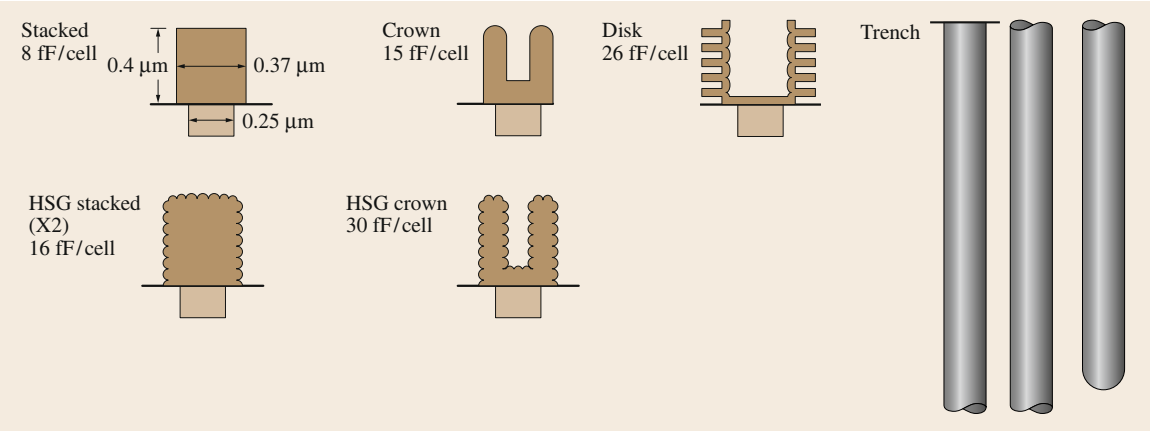


Fig. 27.25 Assorted capacitor structures employed in IC technology. (Courtesy of S. Summerfelt)

atomic-layer CVD processes have been examined for this purpose as well. Although capacitors do not have requirements regarding lateral transport of carriers, as transistors do, capacitors have much more rigorous charge-storage-capacity requirements when they serve as memory elements. This additional constraint results in projected limitations.

27.4.3 Dielectrics for Volatile Memory Capacitors

Capacitor dielectric properties for dynamic random-access memory (DRAM) applications have been important since these devices were introduced commercially in the early 1970s for volatile memory [27.177]. *Fazan* provides some rules of thumb that provide guidance for the selection of desirable materials (and electrical) properties for DRAM capacitor dielectrics [27.175]. In addition to the desire to maintain the storage capacitance around 30 fF/cell (in spite of shrinking dimensions), leakage of charge from the storage capacitor must be kept to a current density of less than $0.1\,\mu A/cm^2$ so that less than $\approx 10\%$ of the capacitor charge is lost during the associated refresh (recharge) cycle. Scaling the DRAM cell size has necessarily resulted in the consideration of a variety of dielectric materials for some time [27.7, 25, 27].

SiO_2 and SiO_xN_y

The earliest DRAM planar capacitors utilized SiO_2 as the dielectric. Subsequent scaling into the Mbit regime required an increased capacitance density, and thus a higher dielectric constant. So-called oxide/nitride (ON) or oxide/nitride/oxide (ONO) dielectric stacks were incorporated [27.16]. As the names imply, the dielectric consisted of a stack of SiO_2 and Si_3N_4 layers. The incorporation of the Si nitride layer, typically by a deposition method such as chemical vapor deposition, results in an overall increase in the dielectric constant of the stack. Scaling has required the consideration of high-k materials in currrent DRAM technologies.

$ZrO_2/Al_2O_3/ZrO_2$

Scaling metal-insulator-metal capacitors required the development of alternative dielectrics (e.g., Al_2O_3 and ZrO_2) and their integration into DRAM processes [27.7, 178]. The stack provides useful dielectric constants in the range of $k \approx 20-50$ for adequate charge storage and leakage performance.

Perovskites

Considerable effort has also been exerted in the search for CMOS-compatible *ultra high-k* dielectrics. Such materials are envisioned to be required for capacitor scaling at and beyond the 10 nm node [27.7]. Perovskite materials, with a $k > 50-100$ have been a focus for this application. The movement of the Ti atom (ion) in the perovskite lattice structure results in a substantial contribution to the polarization of this materials system. Utilization of perovskite has required the use of noble-metal electrodes including Pt, Ru (RuO_2) and Ir to control interfacial reactions. As a result, considerable process complexity is introduced into the manufacturing process.

27.4.4 Dielectrics for Nonvolatile Memory

As the name implies, nonvolatile memory devices retain their state whether power is applied to the device or not. A thorough description of such devices is provided by *Hori* [27.16]. For example, the electrically erasable programmable read-only memory (EEPROM) device requires an erase operation prior to programming (*writing*) new data to the device. To accomplish this,

Table 27.5 Comparison of non-volatile memory technologies. (After [27.179])

Specifications	FRAM	SRAM	EEPROM	Flash
Non-volatile Retains data w/o power	Yes	No	Yes	Yes
Write speed (13 KB)	10 ms	< 10 ms	2 s	1 s
Average active power (μA/MHz) 16 bit word access by the CPU	100	< 60	> 50 000	230
Write endurance	10^{15}	Unlimited	100 000	10 000
Soft errors	Below measurable limits	Yes	Yes	Yes
Bit-wise programmable	Yes	Yes	No	No
Unified memory Flexible code and data partitioning	Yes	No		

a stacked-gate MOS structure is utilized where the intermediate gate is embedded in a dielectric – a so-called *floating gate*. The placement of a higher electric field to permit Fowler–Nordheim tunneling through such dielectrics to the floating gate is utilized to program arrays of these elements – called *flash memory*. Scaling and the reliability required for such nonvolatile memory devices has required the evolution from using SiO_2 to SiO_xN_y.

SiO_2 and SiO_xN_y

For flash memory elements, the formation of nitrided SiO_2 enables a suitable dielectric for reliable operation. The nitridation process, as described earlier, entails the exposure of an SiO_2 dielectric to anneals with N_2, NH_3, or N_2O (or combinations of these), often under rapid thermal annealing conditions. Details of this process have been summarized by *Hori* [27.16]. Subsequent reoxidation of the dielectric is also employed to improve reliability properties. So-called silicon oxide nitride oxide silicon (SONOS) structures are under examination to improve the scaling limitations of the dielectric thickness and tunneling properties [27.181].

Further scaling of transistors presents significant challenges for flash memory. The thickness of the dielectric is again a key constraint given the need for reliable operation under high-electric-field conditions. Memories based on Si nanocrystals embedded within SiO_2 are currently under development to address this challenge [27.182].

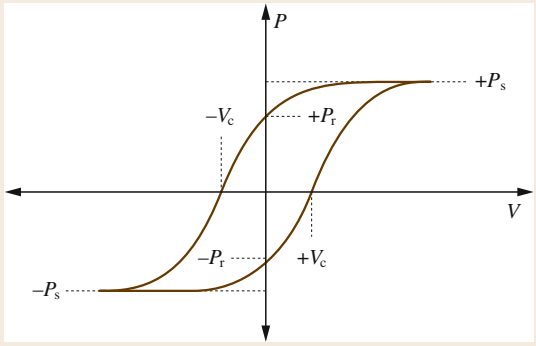


Fig. 27.26 Polarization–voltage curve for a ferroelectric material. (After [27.180], © 2004 IEEE, with permission)

Ferroelectrics

Ferroelectric materials, such as $PbZr_{1-x}Ti_xO_3$ (PZT) or $SrBi_2Ta_2O_9$ (SBT) have been utilized for some time as dielectrics for non-volatile memory elements based upon the ferroelectric effect [27.24]. In storage capacitors incorporating such materials, the polarization state of the ferroelectric is preserved (once poled) without the presence of an electric field [27.23]. This state is then sensed by associated circuitry and can therefore be used as a memory device. The technology has been successfully incorporated in smart card applications as well as microcontroller embeded random access memory – so called ferroelectric RAM (FeRAM or FRAM) [27.179]. A comparison of various memory technologies is provided in Table 27.5.

Principles. Ferroelectricity describes the spontaneous alignment of dipoles in a dielectric as the result of an externally applied electric field. This alignment behavior has thermal constraints in that heating the dielectric above the Curie temperature results in a phase transformation to a paraelectric state. Therefore, capacitors and circuits which utilize the ferroelectric effect must contain materials that have a relatively high Curie temperature compared to that experienced during operation and reliability testing (generally > 200 °C). As shown in Fig. 27.26, the polarization response of these materials to an externally applied electric field exhibits a hysteresis behavior [27.180]. (This behavior is analogous to that observed in ferromagnetic materials, hence the name. Note however that ferroelectric materials contain no iron.) Upon increasing the voltage across the ferroelectric above a coercive value (V_c), the polarization of the material is limited to the spontaneous value (P_s), corresponding to maximum domain align-

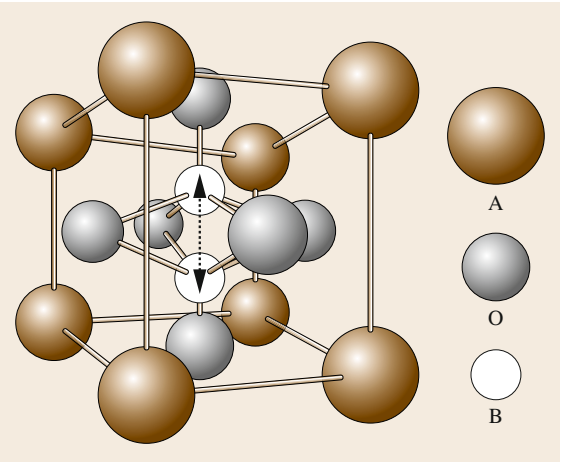


Fig. 27.27 Perovskite ABO unit cell showing motion of the B ion in the cell among the equilibrium positions, resulting in ferroelectric polarization behavior. (After [27.180], © 2004 IEEE, with permission)

ment. Removal of the electric field results in a decrease of the polarization to the remnant value (P_r), which can be sensed accordingly. The time required for the polarization to be switched is on the order of nanoseconds for these devices. Moreover, this can be accomplished in low-power circuits, making these materials attractive, in principle, for scaled CMOS.

Materials. In addition to the perovskite PZT and SBT materials mentioned above, other materials investigated for nonvolatile memory applications include $BaTiO_3$, $PbTiO_3$, $Pb_{1-x}La_xZr_{1-y}Ti_yO_3$ (PLZT), $PbMg_{1-x}Nb_xO_3$ (PMN), $SrBi_2Nb_2O_9$ (SBN), and $SrBi_2(Ta_{1-x}Nb_x)_2O_9$ (SBTN). These materials generally have the cubic structure ABO_3 shown in Fig. 27.27, with the larger (A) cations in the corner of the cubic unit cell, and the smaller (B) cations (e.g., Ti, Zr, Mg, Nb, Ta) located in the center of the unit cell. The oxygen anions are located at the face-centered-cube positions. The distortion of the unit cell in response to the externally applied electric field results in the observed polarization. In particular, the movement of the B cation (e.g., Ti) in the associated lattice between the equilibrium positions results in substantial distortion and therefore polarization [27.24, 180].

Issues for Ferroelectric Materials. Substantial research has been performed to understand the reliability issues associated with ferroelectric memory devices, as their integration tends to occur at the back end of (fabrication) line (BEOL) stage of the typical chip fabrication process. The retention of the polarization state is one area of investigation. It is observed that the polarization state decreases slowly over time (*log-time* decay behavior), even in the absence of an electric field, and the reasons are still not understood. Another area of concern is *imprint*, where a polarization state, if repeatedly poled to the same state, becomes preferred. Subsequent switching to the opposite state can result in relatively poor retention times. Again, the physical mechanism associated with this phenomenon is poorly understood. The role of hydrogen exposure is another area of investigation, where exposure of ferroelectric random-access memory (FeRAM) elements to hydrogen, commonly from forming gas during CMOS back-end processing, can result in the suppression of the remnant polarization [27.183–185]. Barriers for hydrogen (and moisture) permeation into the ferroelectric, such as ALD Al_2O_3, are important for reproducible and reliable operation [27.186–188].

27.5 Interconnect Dielectrics

As noted in Fig. 27.1, the performance of the integrated circuit, as measured by the time delay for signal propagation, also depends upon the interconnections between circuit elements. The scaling of CMOS has resulted in a substantial increase in interconnect metal lines throughout the IC chip, which make a major contribution to the delay time. The industry segments these levels into *local* (interconnection between neighboring devices), *intermediate* (metal 1 interconnection between neighboring circuits), and *global* (interconnection across the chip), as shown in Fig. 27.28. A cross section of a 65 nm-node IC is shown in Fig. 27.29 where eight layers of metallization and the associated dielectric isolation are evident.

As discussed in Sect. 27.1.1, the *RC* time delay for the interconnect contribution to performance can be attributed to the metal lines and their isolation dielectrics (see Fig. 27.1). The resistivity of the lines has been reduced in the industry by recently adopting copper metallization processes in lieu of aluminium metallization in 1998. Further reductions in the delay time then require the consideration of the dielectric between the lines, as these essentially form a parasitic capacitor structure, and therefore low-*k* dielectrics are required. Examples of such interconnect dielectrics are presented in Table 27.6. Reviews of interconnect technology for CMOS are available [27.189]. It is noted that for the global interconnect level, new

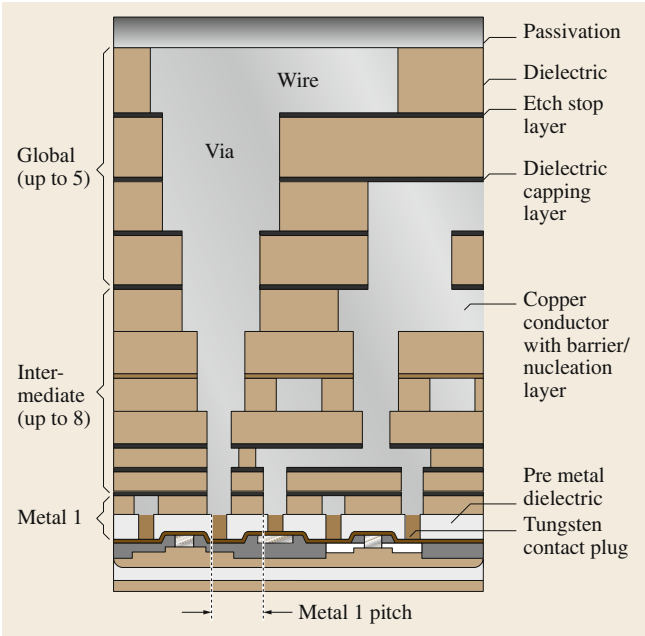


Fig. 27.28 Schematic of a typical chip cross section showing key interconnect metallization and dielectric layers. (From [27.7])

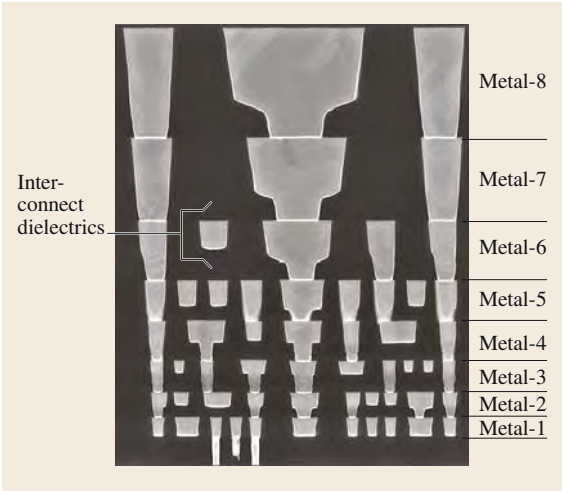


Fig. 27.29 Cross section of a 65 nm node IC chip from 2005 showing eight levels of metal interconnects isolated by low-k dielectrics. (Courtesy of Intel)

concepts such as RF or optical communication, will likely be needed to address extreme CMOS scaling.

27.5.1 Tetraethoxysilane (TEOS)

For many years, CVD-deposited SiO_2 provided adequate isolation for interconnection of ICs. This was frequently accomplished through the deposition of tetraethoxysilane (TEOS) and subsequent densification thermal treatments to render a dielectric constant of $k \approx 4$. Films produced in this manner were relatively easy to process and provided good mechanical strength. The incorporation of fluorine into these films (fluorinated silicate glass or FSG) succeeded in a reduction of the dielectric constant to $k \approx 3.5–3.7$ after considerable efforts. As F is among the most electronegative elements, the incorporation of F into the silica matrix renders the film less polarizable due to Si−F bond formation, and therefore results in a lower permittivity. Scaling CMOS however has driven the industry to consider interconnect dielectric materials with $k \ll 4$.

Table 27.6 Interconnect dielectric materials. (After [27.7, 27])

Dielectric constant	Material
3.5–3.7	FSG
3.0–3.6	Polyimides
2.7–3.1	Spin-on glass
2.6–2.9	Organo-silicate glasses (OSGs)
2.6–2.8	Parylene-based polymers
2.5–3.2	Methyl/hydrogen silsesquioxane (MSQ/HSQ)
1.8–2.4	Porous MSQ, parylene-based polymers
1.1–2.2	Silica aerogels
1.5–2.2	Silica xerogels

27.5.2 Low-k Dielectrics

As seen in Figures 27.28 and 27.29, several levels of dielectrics must be incorporated with the metallization schemes associated with ICs. As can be seen, dielectrics utilized in this back end of (fabrication) line (BEOL) portion of the IC fabrication process are also segmented into pre-, inter- and intra-metal dielectrics. Additionally, dielectric layers are employed to facilitate etching control for Cu metal patterning of the various interconnection lines. This requires materials integration with barrier layers (that control Cu diffusion), etch stop layers, CMP (chemical mechanical polishing) stop layers as well as overall mechanical stability. Additionally, for the local and metal 1 layers, the future incorporation of metal gate and/or NiSi materials may require extensive limitations on processing temperatures ($< 500\,^{\circ}C$).

Several materials have been explored for the low-k application, and some of these are summarized in Table 27.5 [27.7, 27]. Most consist of polymeric (low-Z) materials with varying porosity to enable a sufficiently low dielectric constant for interconnect applications. At this point, films are deposited by either CVD or spin-on methods. Desirable properties for low-k dielectrics include thermal stability (to $\approx 500\,^{\circ}C$), mechanical

stability (to withstand packaging), electrical isolation stability/reliability (similar to SiO_2), chemical stability (minimal moisture absorption, resistance to process chemicals, resist compatibility), compatibility with BEOL materials (e.g., diffusion barriers) and processes (etching, cleaning, post-metallization and forming-gas anneals, chemical mechanical polishing, etc.) and of course low cost. As noted by Wolf, thermal properties of low-*k* materials are important for effective power dissipation in high-performance applications. According to the 2013 ITRS Roadmap, a lack of progress on low-*k* dielectric scaling presents a significant challenge to the industry where innovative solutions would be very welcome [27.7].

27.6 Summary

It should be evident that IC technology is critically dependent upon suitable dielectrics throughout the entire chip. Materials properties, and their resultant electrical properties, must be carefully evaluated throughout the research and development process associated with integrated circuit technology.

Recent years and roadmap predictions clearly place an emphasis on the development of new materials for the various dielectrics employed to achieve scaling. Researchers and technologists engaged in this endeavor must be able to span several disciplines to enable the successful integration of these new dielectric materials.

Acknowledgments. RMW gratefully acknowledges the many discussions and hard work of his colleagues and students engaged in gate-stack research. This work is supported in part by the Erik Jonsson Distinguished Chair at the University of Texas at Dallas.

References

27.1 H.-S. P. Wong: In: *ULSI Devices*, ed. by C.Y. Chang, S.M. Sze (Wiley, New York 2000) Chap. 3

27.2 S. Wolf: *Silicon Processing for the VLSI Era*, Vol. 3 (Lattice, Sunset Beach 1995)

27.3 E. Nicollian, J. Brews: *MOS Physics and Technology* (Wiley, New York 1982)

27.4 S.M. Sze: *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York 1981)

27.5 G. Moore: Electronics **38**, 8 (1965), Also see: http://www.intel.com/labs/index.htm

27.6 G. Moore: Tech. Dig. Int. Electron. Dev. Meet. (IEEE, Washington DC 1975) p. 11

27.7 ITRS: http://itrs2.net/2013-itrs.html/ (2013)

27.8 A. Huang: IEEE Spectrum **52**, 42 (2015)

27.9 D. Hutcheson: Transistors by the numbers, IEEE Spectrum **52**(4), 33 (2015)

27.10 R.F. Service: Science **323**, 1000 (2010)

27.11 C.L. Hinkle, E.M. Vogel, P.D. Ye, R.M. Wallace: Curr. Opin. Solid State Mater. Sci. **15**, 188 (2011)

27.12 S.C. Sun: IEEE Tech. Dig. Int. Electron. Dev. Meet., Washington DC (1997) p. 765

27.13 W. Haensch, E.J. Nowack, R.H. Dennard, P.M. Solomon, A. Bryant, O.H. Dokumaci, A. Kumar, X. Wang, J.B. Johnson, M.V. Fischetti: IBM J. Res. Dev. **50**(4/5), 339 (2006)

27.14 G. Wilk, R.M. Wallace, J.M. Anthony: J. Appl. Phys. **89**(10), 5243 (2001)

27.15 M.T. Bohr: IEEE Tech. Dig. Int. Electron. Dev. Meet., Washington DC (1995) p. 241

27.16 T. Hori: *Gate Dielectrics and MOS ULSI's*, Series in Electronics and Photonics, Vol. 34 (Springer, Berlin 1997)

27.17 A. Chatterjee, M. Rodder, I.-C. Chen: IEEE Trans. Electron. Dev. **45**, 1246 (1998)

27.18 I-C. Chen, W. Liu: In: *ULSI Devices*, ed. by C.Y. Chang, S.M. Sze (Wiley, New York 2000) Chap. 10

27.19 R. Dennard, F. Gaensslen, H.-N. Yu, V.L. Rideout, E. Bassous, A.R. LeBlanc: J. Solid State Circuits **SC-9**, 256 (1974)

27.20 R. Dennard, F. Gaensslen, E. Walker, P. Cook: J. Solid State Circuits **SC-14**, 247 (1979)

27.21 D. Frank, R.H. Dennard, E. Nowak, P.M. Solomon, Y. Taur, H.-S.P. Wong: Proc. IEEE **89**, 259 (2001)

27.22 S. Thompson, P. Packan, M. Bohr: Intel Technol. J. Q **3**, 223 (1998)

27.23 S.O. Kasap: *Principles of Electronic Materials and Devices*, 3rd edn. (McGraw–Hill, New York 2005)

27.24 R. Ramesh: *Thin Film Ferroelectric Materials and Devices* (Kluwer, Boston 1997)

27.25 D.-S. Yoon, J.S. Roh, H.K. Baik, S.-M. Lee: Crit. Rev. Solid State Mater. Sci. **27**, 143 (2002)

27.26 B. Bivari: IEEE Tech. Dig. Int. Electron. Dev. Meet. (1996) p. 555

27.27 S. Wolf: *Silicon Processing for the VLSI Era*, Vol. 4 (Lattice, Sunset Beach 2002)

27.28 S. Wolf: *Silicon Processing for the VLSI Era*, Vol. 1 (Lattice, Sunset Beach 1986)

27.29 S. Wolf: *Silicon Processing for the VLSI Era*, Vol. 2 (Lattice, Sunset Beach 1990)

27.30 J. Robertson, R.M. Wallace: Mater. Sci. Eng. R **88**, 1 (2015)

27.31 S. Banerjee, B. Streetman: In: *ULSI Devices*, ed. by C.Y. Chang, S.M. Sze (Wiley, New York 2000) Chap. 4

27.32 R. Rios, N. D. Arora: IEEE Tech. Dig. Int. Electron. Dev. Meet., San Francisco (1994) p. 613

27.33 Y.-C. King, C. Hu, H. Fujioka, S. Kamohara: Appl. Phys. Lett. **72**, 3476 (1998)

27.34 K. Yang, Y.-C. King, C. Hu: Symp. VLSI Tech. Tech. Dig. Papers, Kyoto (1999) p. 77
27.35 C.Y. Wong, J.Y. Sun, Y. Taur, C.S. Oh, R. Angelucci, B. Davari: IEEE Tech. Dig. Int. Electron. Dev. Meet., San Francisco (1988) p. 238
27.36 A.C. Diebold, D. Venables, Y. Chabal, D. Muller, M. Weldon, E. Garfunkel: Mater. Sci. Semicond. Proc. **2**, 104 (1999)
27.37 Z.H. Lu, J.P. McCaffrey, B. Brar, G.D. Wilk, R.M. Wallace, L.C. Feldman, S.P. Tay: Appl. Phys. Lett. **71**, 2764 (1997)
27.38 E.M. Vogel, K.Z. Ahmed, B. Hornung, W.K. Henson, P.K. McLarty, G. Lucovsky, J.R. Hauser, J.J. Wortman: IEEE Trans. Electron. Dev. **45**, 1350 (1998)
27.39 M.L. Green, E.P. Gusev, R. Degraeve, E.L. Garfunkel: J. Appl. Phys. **90**, 2057 (2001)
27.40 D.K. Schroder: *Semiconductor Material and Device Characterization*, 2nd edn. (Wiley, New York 1998)
27.41 C.R. Helms, E.H. Poindexter: Rep. Prog. Phys. **57**, 791 (1994)
27.42 R. Hummel: *Electronic Properties of Materials*, 2nd edn. (Springer, New York 1993)
27.43 W.R. Runyan, K.E. Bean: *Semiconductor Integrated Circuit Processing Technology* (Addison-Wesley, New York 1990)
27.44 F.J. Himpsel, F.R. McFeely, A. Taleb-Ibrahimi, J.A. Yarmoff, G. Hollinger: Phys. Rev. B **38**, 6084 (1988)
27.45 S.T. Pantilides: *The Physics of Si0 and its Interfaces* (Pergamon, New York 1978)
27.46 G. Lucovsky, S.T. Pantilides, F.L. Galeener: *The Physics of MOS Insulators* (Pergamon, New York 1980)
27.47 C.R. Helms, B.E. Deal: *The Physics and Chemistry of Si0 and the Si–Si0 Interface* (Plenum, New York 1988)
27.48 P. Balk (Ed.): *The Si–Si0 $_2$ System* , Material Science Monographs (Elsevier, New York 1988) p. 32
27.49 C.R. Helms, B.E. Deal: *The Physics and Chemistry of Si0 and the Si–Si0 Interface*, Vol. 2 (Plenum, New York 1993)
27.50 H.Z. Massoud, E.H. Poindexter, C.R. Helms: *The Physics of Si0 and its Interfaces – 3*, Vol. 96-1 (Electrochemical Society, Pennington 1996)
27.51 R.A.B. Devine: *The Physics and Technology of Amorphous Si0* (Plenum, New York 1988) p. 2
27.52 E. Garfunkel, E. Gusev, A. Vul: *Fundamental Aspects of Ultrathin Dielectrics on Si-based Devices*, NATO Science Series, Vol. 3/47 (Kluwer, Nowell 1998)
27.53 R.A.B. Devine, J.-P. Duraud, E. Dooryhee: *Structure and Imperfections in Amorphous and Crystalline Silicon Dioxide* (Wiley, New York 2000)
27.54 H.Z. Massoud, I.J.R. Baumvol, M. Hirose, E.H. Poindexter: *The Physics of Si0 and its Interfaces – 4*, Vol. PV2000-2 (Electrochemical Society, Pennington 2000)
27.55 A. Bongiorno, A. Pasquarello: Mater. Sci. Eng. B **96**, 102 (2002)
27.56 J.H. Oh, H.W. Yeom, Y. Hagimoto, K. Ono, M. Oshima, N. Hirashita, M. Nywa, A. Toriumi, A. Kakizaki: Phys. Rev. B **63**, 205310 (2001)
27.57 G. Dolino: In: *Structure and Imperfections in Amorphous and Crystalline Silicon Dioxide*, ed. by R.A.B. Devine, J.-P. Duraud, E. Dooryhee (Wiley, New York 2000) Chap. 2
27.58 L.W. Hobbs, C.E. Jesurum, B. Berger: In: *Structure and Imperfections in Amorphous and Crystalline Silicon Dioxide*, ed. by R.A.B. Devine, J.-P. Duraud, E. Dooryhee (Wiley, New York 2000) Chap. 1
27.59 A. Bongiorno, A. Pasquarello: Appl. Phys. Lett. **83**, 1417 (2003)
27.60 F. Mauri, A. Pasquarello, B.G. Pfrommer, Y.-G. Yoon, S.G. Louie: Phys. Rev. B **62**, R4786 (2000)
27.61 W.H. Zachariasen: J. Am. Chem. Soc. **54**, 3841 (1932)
27.62 P. Balk: J. Nanocryst. Sol. **187**, 1 (1995)
27.63 F.J. Grunthaner, P.J. Grunthaner: Mater. Sci. Rep. **1**, 65 (1986)
27.64 A. Stirling, A. Pasquerello: Phys. Rev. B **66**, 24521 (2002)
27.65 G. Hollinger, F.R. Himpsel: Phys. Rev. B **28**, 3651 (1983)
27.66 G. Hollinger, F.R. Himpsel: Appl. Phys. Lett. **44**, 93 (1984)
27.67 T. Hattori: Crit. Rev. Solid State Mater. Sci. **20**, 339 (1995)
27.68 F. Rochet, Ch. Poncey, G. Dufour, H. Roulet, C. Guillot, F. Sirotti: J. Non-Cryst. Solids **216**, 148 (1997)
27.69 D.A. Muller, T. Sorsch, S. Moccio, F.H. Baumann, K. Evans-Lutterodt, G. Timp: Nature **399**, 758 (1999)
27.70 I.J.R. Baumvol: Surf. Sci. Rep. **36**, 1 (1999)
27.71 J.J. Lander, J. Morrison: J. Appl. Phys. **33**, 2089 (1962)
27.72 F.W. Smith, G. Ghidini: J. Electrochem. Soc. **129**, 1300 (1982)
27.73 K. Wurm, R. Kliese, Y. Hong, B. Röttger, Y. Wei, H. Neddermeyer, I.S.T. Tsong: Phys. Rev. B **50**, 1567 (1994)
27.74 J. Seiple, J.P. Pelz: Phys. Rev. Lett. **73**, 999 (1994)
27.75 J. Seiple, J.P. Pelz: J. Vac. Sci. Technol. A **13**, 772 (1995)
27.76 Y. Wei, R.M. Wallace, A.C. Seabaugh: J. Appl. Phys. **81**, 6415 (1997)
27.77 E.P. Gusev, H.C. Lu, T. Gustafsson, E. Garfunkel: Phys. Rev. B **52**, 1759 (1995)
27.78 B.E. Deal, A.S. Grove: J. Appl. Phys. **36**, 3770 (1965)
27.79 J.D. Plummer: Silicon oxidation kinetics-from Deal-Grove to VLSI process models. In: *The Physics of Si0 and its Interfaces – 3*, Vol. 96-1, ed. by H.Z. Massoud, E.H. Poindexter, C.R. Helms (Electrochemical Society, Pennington 1996) p. 129
27.80 P. Balk: Trans. IEEE **53**, 2133 (1965)
27.81 G. Abowitz, E. Arnold, J. Ladell: Phys. Rev. Lett. **18**, 543 (1967)
27.82 B.E. Deal, M. Sklar, A.S. Grove, E.H. Snow: J. Electrochem. Soc. **114**, 266 (1967)
27.83 E. Arnold, J. Ladell, G. Abowitz: Appl. Phys. Lett. **13**, 413 (1968)
27.84 R.R. Razouk, B.E. Deal: J. Electrochem. Soc. **126**(9), 1573 (1979)

27.85 J.H. Weil, J.R. Bolton, J.E. Wertz: *Electron Paramagnetic Resonance: Elementary Theory and Practical Applications* (Wiley, New York 1994)
27.86 P.J. Caplan, J.N. Helbert, B.E. Wagner, E.H. Poindexter: Surf. Sci. **54**, 33 (1976)
27.87 E.H. Poindexter, P.J. Caplan: Prog. Surf. Sci. **14**, 201 (1983)
27.88 E.H. Poindexter, P.J. Caplan: J. Vac. Sci. Technol. A **6**, 390 (1988)
27.89 J.F. Conley: Mater. Res. Soc. Symp. Proc. **428**, 293 (1996)
27.90 J.F. Conley, P.M. Lenahan: A review of electron spin resonance spectroscopy of defects in thin film SiO_2 on Si. In: *The Physics of SiO_2 and its Interfaces – 3* , Vol. 96-1, ed. by H.Z. Massoud, E.H. Poindexter, C.R. Helms (Electrochemical Society, Pennington 1996) p. 214
27.91 P.M. Lenahan, J.F. Conley: J. Vac Sci. Technol. B **16**, 2134 (1998)
27.92 A.G. Revesz, B. Goldstein: Surf. Sci. **14**, 361 (1969)
27.93 Y. Nishi: J. Appl. Phys. **10**, 52 (1971)
27.94 I. Shiota, N. Miyamoto, J.-I. Nishizawa: J. Appl. Phys. **48**, 2556 (1977)
27.95 E.H. Poindexter, E.R. Ahlstrom, P.J. Caplan: In: *The Physics of SiO_2 and its Interfaces*, ed. by S.T. Pantilides (Pergamon, New York 1978) p. 227
27.96 P.J. Caplan, E.H. Poindexter, B.E. Deal, R.R. Razouk: J. Appl. Phys. **50**, 5847 (1979)
27.97 G.J. Gerardi, E.H. Poindexter, P.J. Caplan, N.M. Johnson: Appl. Phys. Lett. **49**, 348 (1986)
27.98 D. Sands, K.M. Brunson, M.H. Tayarani-Najaran: Semicond. Sci. Technol. **7**, 1091 (1992)
27.99 K.L. Brower: Phys. Rev. B **38**, 9657 (1988)
27.100 K.L. Brower: Phys. Rev. B **42**, 3444 (1990)
27.101 A. Stesmans, V.V. Afanas'ev: J. Appl. Phys. **83**, 2449 (1998)
27.102 A. Stesmans, V.V. Afanas'ev: J. Phys. Condens. Matter **10**, L19 (1998)
27.103 A. Stesmans, V.V. Afanas'ev: Micro. Eng. **48**, 116 (1999)
27.104 A. Stesmans, B. Nouwen, V. Afanas'ev: Phys. Rev. B **58**, 15801 (1998)
27.105 E.H. Poindexter, P.J. Caplan: In: *Insulating Films on Semiconductors*, ed. by M. Schulz, G. Pensl (Springer, Berlin, Heidelberg 1981) p. 150
27.106 M. Shulz, G. Pensl: *Insulating Films on Semiconductors* (Springer, New York 1981)
27.107 P.J. Caplan, E.H. Poindexter, B.E. Deal, R.R. Razouk: In: *The Physics of MOS Insulators*, ed. by G. Lucovsky, S.T. Pantilides, F.L. Galeener (Pergamon, New York 1980) p. 306
27.108 W.H. Brattain, J. Bardeen: Bell Syst. Tech. J. **13**, 1 (1953)
27.109 W.L. Brown: Phys. Rev. **91**, 518 (1953)
27.110 E.N. Clarke: Phys. Rev. **91**, 756 (1953)
27.111 H.R. Huff: J. Electrochem. Soc. **149**, S35 (2002)
27.112 J.T. Law: J. Phys. Chem. **59**, 67 (1955)
27.113 J.T. Law, E.E. Francois: J. Phys. Chem. **60**, 353 (1956)
27.114 J.T. Law: J. Phys. Chem. **61**, 1200 (1957)
27.115 J.T. Law: J. Appl. Phys. **32**, 600 (1961)
27.116 P. Balk: Microelectron. Eng. **48**, 3 (1999)
27.117 E. Kooi: Philips Res. Rep. **20**, 578 (1965)
27.118 P. Balk: Electrochem. Soc. Ext. Abstr. **14**(109), 237 (1965)
27.119 P. Balk: Electrochem. Soc. Ext. Abs. **14**(111), 29 (1965)
27.120 P. Balk: J. Electrochem. Soc. **112**, 185C (1965)
27.121 E. Kooi: Philips Res. Rep. **21**, 477 (1966)
27.122 A. Stesmans: Appl. Phys. Lett. **68**, 2076 (1996)
27.123 A. Stesmans, V. Afanas'ev: Micro. Eng. **36**, 201 (1997)
27.124 K.L. Brower, S.M. Myers: Appl. Phys. Lett. **57**, 162 (1999)
27.125 A. Stesmans: Phys. Rev. B **48**, 2418 (1993)
27.126 A.G. Revesz: J. Electrochem. Soc. **126**, 122 (1979)
27.127 G.J. Gerardi, E.H. Poindexter: J. Electronchem. Soc. **136**, 588 (1989)
27.128 T.R. Oldham, F.B. McLean, H.E. Boesch Jr., J.M. McGarrity: Semicond. Sci. Technol. **4**, 986 (1989)
27.129 A. Stesmans, V. Afanas'ev: Phys. Rev. B **54**, 11129 (1996)
27.130 E. Wu, B. Linder, J. Stathis, W. Lai: IEEE Tech. Dig. Int. Electron. Dev. Meet., Washington DC (2003) p. 919
27.131 D.A. Buchanan: IBM J. Res. Devel. **43**, 245 (1999)
27.132 Y. Wu, G. Lucovsky, Y.-M. Lee: IEEE Trans. Electron. Dev. **47**, 1361 (2000)
27.133 M. Hillert, S. Jonsson, B. Sundman: Z. Metallkd. **83**, 648 (1992)
27.134 D.M. Brown, P.V. Gray, F.K. Heumann, H.R. Philipp, E.A. Taft: J. Electrochem. Soc. **115**, 311 (1968)
27.135 E.H. Poindexter, W.L. Warren: J. Electrochem. Soc. **142**, 2508 (1995)
27.136 J.T. Yount, P.M. Lenahan, P.W. Wyatt: J. Appl. Phys. **74**, 5867 (1993)
27.137 K. Kushida-Abdelghafar, K. Watanabe, T. Kikawa, Y. Kamigaki, J. Ushio: J. Appl. Phys. **92**, 2475 (2002)
27.138 G. Lucovsky, T. Yasuda, Y. Ma, S. Hattangady, V. Misra, X.-L. Xu, B. Hornung, J.J. Wortman: J. Non-Cryst. Solids **179**, 354 (1994)
27.139 S.V. Hattangady, H. Niimi, G. Lucovsky: Appl. Phys. Lett. **66**, 3495 (1995)
27.140 S.V. Hattangady, R. Kraft, D.T. Grider, M.A. Douglas, G.A. Brown, P.A. Tiner, J.W. Kuehne, P.E. Nicollian, M.F. Pas: IEEE Tech. Dig. Int. Electron. Dev. Meet., San Francisco (1996) p. 495
27.141 H. Yang, G. Lucovsky: Tech. Dig. Int. Electron. Dev. Meet., Washington DC (1999) p. 245
27.142 J.P. Chang, M.L. Green, V.M. Donnelly, R.L. Opila, J. Eng Jr., J. Sapjeta, P.J. Silverman, B. Weir, H.C. Lu, T. Gustafsson, E. Garfunkel: J. Appl. Phys. **87**, 4449 (2000)
27.143 S. Guha, E. Cartier, N.A. Bojarczuk, J. Bruley, L. Gignac, J. Karasinski: J. Appl. Phys. **90**, 512 (2001)
27.144 S. Miyazaki: J. Vac. Sci. Technol. B **19**, 2212 (2001)
27.145 J. Robertson, C.W. Chen: Appl. Phys. Lett. **74**, 1168 (1999)
27.146 J. Robertson: J. Vac. Sci. Technol. B **18**, 1785 (2000)
27.147 J. Robertson: J. Non-Cryst. Solids **303**, 94 (2002)
27.148 H.Y. Yu, M.F. Li, B.J. Cho, C.C. Yeo, M.S. Joo, D.-L. Kwong, J.S. Pan, C.H. Ang, J.Z. Zheng, S. Ramanathan: Appl. Phys. Lett. **81**, 376 (2002)

27.149 D.G. Schlom, J.H. Haeni: Mater. Res. Soc. Bull. **27**(3), 198 (2002)
27.150 E. Zhu, T.P. Ma, T. Tamagawa, Y. Di, J. Kim, R. Carruthers, M. Gibson, T. Furukawa: IEEE Tech. Dig. Int. Electron. Dev. Meet. (IEEE, Washington DC 2001) p. 20.4.1.
27.151 G.D. Wilk, M.L. Green, M.-Y. Ho, B.W. Busch, T.W. Sorsch, F.P. Klemens, B. Brijs, R.B. van Dover, A. Kornblit, T. Gustafsson, E. Garfunkel, S. Hillenius, D. Monroe, P. Kalavade, J.M. Hergenrother: IEEE Tech. Dig. VLSI Symp., Honolulu (2002) p. 88
27.152 H. Nohira, W. Tsai, W. Besling, E. Young, J. Petry, T. Conard, W. Vandervorst, S. De Gendt, M. Heyns, J. Maes, M. Tuominen: J. Non-Cryst. Solids **303**, 83 (2002)
27.153 M.R. Visokay, J.J. Chambers, A.L.P. Rotondaro, A. Shanware, L. Colombo: Appl. Phys. Lett. **80**, 3183 (2002)
27.154 M.S. Akbar, S. Gopalan, H.-J. Cho, K. Onishi, R. Choi, R. Nieh, C.S. Kang, Y.H. Kim, J. Han, S. Krishnan, J.C. Lee: Appl. Phys. Lett. **82**, 1757 (2003)
27.155 K. Torii, T. Aoyama, S. Kamiyama, Y. Tamura, S. Miyazaki, H. Kitajima, T. Arikado: Tech. Dig. VLSI Symp., Honolulu (2004) p. 112
27.156 R.M. Wallace, G. Wilk: Mater. Res. Soc. Bull. **27**(3), 192 (2002), also see this focus issue for reviews of other aspects on gate dielectric issues
27.157 R.M. Wallace, G. Wilk: Critical Rev. Solid State Mater. Sci. **28**, 231 (2003)
27.158 G.D. Wilk, R.M. Wallace: Appl. Phys. Lett. **74**, 2854 (1999)
27.159 G.D. Wilk, R.M. Wallace: Appl. Phys. Lett. **76**, 112 (2000)
27.160 R.M. Wallace, R.A. Stolz, G.D. Wilk: Zirconium and/or hafnium silicon-oxynitride gate, US Patents 6013553, 6020243, 6291866, 6291867 (2000)
27.161 G.D. Wilk, R.M. Wallace, J.M. Anthony: J. Appl. Phys. **87**, 484 (2000)
27.162 A.L.P. Rotondaro, M.R. Visokay, J.J. Chambers, A. Shanware, R. Khamankar, H. Bu, R.T. Laaksonen, L. Tsung, M. Douglas, R. Kuan, M.J. Bevan, T. Grider, J. McPherson, L. Colombo: Symp. VLSI Technol. Tech. Dig. Papers, Honolulu (2002) p. 148
27.163 R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, M. Metz: IEEE Elec. Dev. Lett. **25**(6), 408 (2004)
27.164 T. Poletti: Biggest advances in chips in 40 years, EE-Times, January 26 2007
27.165 D. Scansen: 45nm: What Intel didn't tell you, EE Times, June 23 2008
27.166 Y.-C. Yeo: Thin Solids Films **462-3**, 34 (2004)
27.167 M.V. Fischetti, D.A. Nuemayer, E.A. Cartier: J. Appl. Phys. **90**, 4587 (2001)
27.168 R.M.C. de Almeida, I.J.R. Baumvol: Surf. Sci. Rep. **49**, 1 (2003)
27.169 M. Quevedo-Lopez, M. El-Bouanani, M.J. Kim, B.E. Gnade, M.R. Visokay, A. LiFatou, M.J. Bevan, L. Colombo, R.M. Wallace: Appl. Phys. Lett. **81**, 1609 (2002)
27.170 M. Quevedo-Lopez, M. El-Bouanani, M.J. Kim, B.E. Gnade, M.R. Visokay, A. LiFatou, M.J. Bevan, L. Colombo, R.M. Wallace: Appl. Phys. Lett. **82**, 4669 (2003)
27.171 M.-Y. Ho, H. Gong, G.D. Wilk, B.W. Busch, M.L. Green, W.H. Lin, A. See, S.K. Lahiri, M.E. Loomans, P.I. Räisänen, T. Gustafsson: Appl. Phys. Lett. **81**, 4218 (2002)
27.172 S. Guha, E.P. Gusev, H. Okorn-Schmidt, M. Copel, L.Å. Ragnarsson, N.A. Bojarczuk: Appl. Phys. Lett. **81**, 2956 (2002)
27.173 Y.-C. Yeo, T.-J. King, C. Hu: J. Appl. Phys. **92**, 7266 (2002)
27.174 I.S. Jeon, J. Lee, P. Zhao, P. Sivasubramani, T. Oh, H.J. Kim, D. Cha, J. Huang, M.J. Kim, B.E. Gnade, J. Kim, R.M. Wallace: IEEE Tech. Dig. Int. Electron. Dev. Meet. (IEEE, San Francisco 2004)
27.175 P.C. Fazan: Integr. Ferroelectr. **4**, 247 (1994)
27.176 H. Schichijo: In: *ULSI Devices*, ed. by C.Y. Chang, S.M. Sze (Wiley, New York 2000) Chap. 7
27.177 P.J. Harrop, D.S. Campbell: Thin Solid Films **2**, 273 (1968)
27.178 D.-S. Kil, H.-S. Song, K.-J. Lee, K. Hong, J.-H. Kim, K.-S. Park, S.-J. Yeom, J.-S. Roh, N.-J. Kwak, H.-C. Sohn, J.-W. Kim, S.-W. Park: Symp. VLSI Technology Digest of Technical Papers (2006) p. 38
27.179 For example, see Texas Instruments FRAM products at: http://www.ti.com/lsds/ti/microcontrollers_16-bit_32-bit/msp/ultra-low_power/msp430frxx_fram/what_is_fram.page?DCMP=FRAM&HQS=fram
27.180 G.F. Derbenwick, A.F. Isaacson: IEEE Circuits Dev. **20**, 20 (2001)
27.181 S. S. Chung, P.-Y. Chiang, G. Chou, C.-T. Huang, P. Chen, C.-H. Chu, C.C.-H. Hsu: Tech. Dig. Int. Electron. Dev. Meet. (IEEE, Washington DC 2003) p. 26.6.1.
27.182 B. De Salvo: Tech. Dig. Int. Electron. Dev. Meet. (IEEE, Washington DC 2003) p. 26.1.1.
27.183 A.R. Krauss, A. Dhote, O. Auciello, J. Im, R. Ramesh, A. Aggarwal: Integr. Ferroelectr. **27**, 147 (1999)
27.184 J. Im, O. Auciello, A.R. Krauss, D.M. Gruen, R.P.H. Chang, S.H. Kim, A.I. Kingon: Appl. Phys. Lett. **74**, 1162 (1999)
27.185 N. Poonawala, V.P. Dravid, O. Auciello, J. Im, A.R. Krauss: J. Appl. Phys. **87**, 2227 (2000)
27.186 T. Mikolajick, C. Dehm, W. Hartner, I. Kasko, M.J. Kastner, N. Nagel, M. Moert, C. Mazure: Microelect. Rel. **41**, 947 (2001)
27.187 H. Kohlstedt, Y. Mustafa, A. Gerber, A. Petraru, M. Fitsilis, R. Meyer, U. Böttger, R. Waser: Microelec. Eng. **80**, 296 (2005)
27.188 A. Chowdhury, J. Kim, R.M. Wallace: Proc. IEEE 8th Ann. Non-Volatile Mem. Tech. Symp. (2007) p. 48
27.189 Y. Shacham-Diamand, T. Osaka, M. Datta, T. Ohba (Eds.): *Advanced Nanoscale ULSI Interconnects: Fundamentals and Applications* (Springer Science+Business Media, New York 2009)

28. Thin Films

Robert D. Gould†, Safa Kasap, Asim K. Ray

This chapter provides an extended introduction to the basic principles of thin-film technology, including deposition processes, structure, and some optical and electrical properties relevant to this volume. The material is accessible to scientists and engineers with no previous experience in this field, and contains extensive references to both the primary literature and earlier review articles. Although it is impossible to provide full coverage of all areas or of the most recent developments in this survey, references are included to enable the reader to access the information elsewhere, while the coverage of fundamentals will allow this to be appreciated.

Deposition of thin films by the main physical deposition methods of vacuum evaporation, molecular-beam epitaxy and sputtering are described in some detail, as are those by the chemical deposition methods of electrodeposition, chemical vapour deposition and the Langmuir–Blodgett technique. Examples of structural features of some thin films are given, including their crystallography, larger-scale structure and film morphology. The dependence of these features on the deposition conditions are stressed, including those required for the growth of epitaxial films and the use of zone models in the classification of the morphological characteristics. The main optical properties of thin films are reviewed, including the use of Fresnel coefficients at media boundaries, reflectance and transmittance, matrix methods and the application of these techniques to the design of antireflection coatings, mirrors and filters. The dependence of electrical conductivity (or resistivity) and the temperature coefficient of resistivity in metallic thin films is discussed, in particular the models of Thomson, Fuchs–Sondheimer and the grain-boundary model of Mayadas–Shatzkes. For insulating and semiconducting thin films the origin and effects of several high-field conduction processes are examined, including space-charge-limited conductivity, the Poole–Frenkel effect, hopping, tunnelling and the Schottky effect. Finally, some speculations regarding future developments are made.

28.1 **Deposition Methods** 647
28.1.1 Physical Deposition Methods 647
28.1.2 Chemical Deposition Methods 662
28.1.3 Sol–Gel Method 667
28.2 **Structure** 669
28.2.1 Crystallography 669
28.2.2 Film Structure 670
28.2.3 Morphology 676
28.2.4 Raman Spectroscopy 678
28.3 **Properties** 680
28.3.1 Optical Properties 680
28.3.2 Optical Absorption 684
28.3.3 Electrical Properties 685
28.4 **Concluding Remarks** 699
References 702

The earliest use of thin films by mankind is probably in the application of glazes to brickwork and pottery. Tin glazing was discovered by the Assyrians and was used to coat decorative brickwork. Glazed brick panels have been recovered from various archeological sites, with a fragment from Nimrud being dated to about 890 B.C. Various alternative types of glaze have since been developed, and are used both to overcome the effects of porosity in pottery and for decorative purposes. The importance of thin films in optics was first recognised in the 17th century by Hooke, Newton and others. In 1675 Newton described observations of colours in thin films of transparent material and from geometrical considerations was able to calculate effective film thickness and correlate this with the periodicity of observed colours. The films in question were typically of air or water in the space between two glass surfaces having different radii of curvature. The phenomena described are seen most conveniently as *Newton's rings*, where an optically flat glass surface is in contact with

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_28

a convex glass surface having a large radius of curvature. Under monochromatic illumination in reflection, a series of concentric coloured rings were observed, the radii being proportional to $\sqrt{1}$, $\sqrt{2}$, $\sqrt{3}$, etc. Although the wave nature of light and interference phenomena were not universally accepted at that time (especially by Newton) these observations, and others where the radii of the rings were seen to be larger for red than for violet light, established one of the most important properties of thin films. Since these early observations of thin film properties, they have been found to display not only characteristic and distinctive optical behaviour, but also mechanical, electrical and magnetic properties. These properties underpin many of their present-day industrial applications, such as antireflection coatings and optical filters, surface acoustic wave devices, electronic components (both discrete and integrated) such as resistors, capacitors, thin-film transistors and other active devices, magnetic data storage and superconductors.

Notwithstanding this wide range of properties and applications, it should be stressed that there is no watertight definition of the thickness below which a film becomes thin. The reason for this is primarily because different properties scale differently with thickness. *Eckertová* [28.1] has pointed out that it is permissible to say in general that the physical limit is determined by the thickness below which certain described anomalies appear, but that this differs for different physical phenomena. For instance a film which is optically thin may be of the order of the wavelength of light ($\approx 500\,\mathrm{nm}$) although some electrical properties are determined by the mean free path of conduction electrons ($\approx 50\,\mathrm{nm}$ for metals) or by the thickness at which a given applied voltage produces an electric field which leads to high-field (nonlinear) behaviour. Generally speaking a thin film has a thickness of less than $1\,\mu\mathrm{m}$, and is deposited on the surface of a substrate by one of several distinct deposition methods [28.2]. In keeping with the remit of this Handbook, the emphasis in this chapter is on the electrical and optical properties of thin films, as these are clearly the most important in the field of electronic and optoelectronic materials. It is not the intention to dismiss thermal and mechanical properties as irrelevant in this sphere, but to stress that, although they may both also have an influence on the electrical and optical properties, they are both mainly dependent on the material itself. Mechanical properties of thin films have been reviewed by *Hoffman* [28.3] and also by *Campbell* [28.4]. However, the *structure* of thin films often has a crucial effect in determining other thin-film properties and, unlike the thermal and mechanical properties, is very largely determined by the film deposition method and conditions. For these reasons sections on both the structure and deposition methods of thin films are included in this review.

The basic structure of the chapter, following this brief introduction, covers the various aspects of thin films starting from their deposition methods, through their structures to their optical and electrical properties. There is no attempt to cover all of these aspects comprehensively; in a chapter such as this there is sufficient space to cover only the most important techniques and properties. In Sect. 28.1 the major deposition methods are described, such as physical processes like evaporation and sputtering, and chemical methods including chemical vapour deposition (CVD) and the Langmuir–Blodgett technique for molecular films. In Sect. 28.2 some examples of the crystalline structure of various types of films are examined. These can be relatively simple in the case of evaporated metal films, whereas compound films (such as compound semiconductors) can exhibit a variety of different structural forms depending on the deposition conditions and the thickness. Films prepared from larger molecules, such as the organic phthalocyanines having molecular weights of the order of 500, can also show several quite complex structures. In addition to the crystal structure per se, which is normally determined using X-ray, electron or neutron diffraction techniques, the morphologies of thin films often show interesting features, which are observed using electron microscopy or one of the newer scanning-probe imaging techniques such as scanning–tunnelling microscopy (STM) or atomic force microscopy (AFM). Section 28.3 covers the basic optical and electrical properties of thin films. The optical properties covered in Sect. 28.3.1 are largely the result of the electromagnetic wave nature of light, leading to the interference phenomena which were first observed in the 17th century. Films of suitable thickness and refractive index may be used in simple antireflection coatings, and quite complex multilayer film structures are used as optical filters. In Sect. 28.3.3 electrical properties are described. These include electrical conductivity (or resistivity) in relatively high-conductivity materials such as metals, where the conductivity depends not only on the carrier concentration and bulk mean free path, but is also modified by the effects of scattering at one or both of the film surfaces and internal grain boundaries. Lateral conduction in discontinuous (island) films where there are significant potential barriers between highly conducting regions has been investigated by *Neugebauer* and *Webb* [28.5], although this not discussed in the present work. This mechanism is reviewed elsewhere in the literature [28.1, 6]. Various nonlinear conduction processes arise as the result of the high electric fields that can be applied across thin films; for example a film of thickness 100 nm having a poten-

tial difference of only 1 V applied between its surfaces, would experience a field of 1×10^7 V/m, which approaches the dielectric breakdown strength of many materials. Finally in Sect. 28.4 the main points or the discussion are summarised and some speculations concerning future trends are given.

Much of the technology associated with thin-film deposition and growth is explicitly excluded from this discussion. The measurement of film thickness, both during deposition (monitoring) or post-deposition, is not covered, although good accounts are given in the literature [28.7, 8]. Furthermore the condensation and growth mechanisms of thin films are also omitted [28.9, 10], although the structure and other properties are largely influenced by these. Many of the up-to-date materials properties and device applications are covered elsewhere in this volume. The reader is referred to Chap. 14 for details of epitaxial growth techniques and to Chap. 17 for structural characterisation. Optical properties and characterisation are discussed in Chap. 3, while the corresponding materials and devices are addressed in Chap. 35. Sensors and transducers, which are frequently based on various thin-film techniques (both inorganic and organic) are described in Chap. 37 and 52. In the area of electronics, much current effort is in the areas of molecular electronics as described in Chap. 51. Indeed the present author has predicted that many of the high-field conduction effects commonly observed in thin films will, in due course, also be detected in nanostructures in a suitably modified form [28.11].

Many of the topics coved in this chapter are explained in more detail in various books. The *Handbook of Thin Film Technology* [28.12], although published over three decades ago, has an enormous amount of detail concerning basic thin-film phenomena, most of which is still of interest today. Another very useful resource is the series of volumes originally entitled *Physics of Thin Films* and later simply *Thin Films* (recent volumes retitled *Thin Films and Nanotechnology*) published by Academic Press. Volume 25 of this series [28.13] contains an index for all the earlier volumes covering the years 1963–1998. Detailed references to several articles from both these sources are given in the references. Also of particular interest in the materials field is the *Handbook of Thin Film Materials* [28.14]. The reader is also referred to several other books covering the general field of thin films and technology [28.1, 15–19]. Other books covering more restricted areas of thin-film technology (e.g., optical properties, electronics) are referred to in the appropriate sections.

The SI system of units is used throughout, and where original work was presented in non-SI units, these have normally been converted. Other customary units, such as the electronvolt, are used where appropriate. The main emphasis in this chapter is on the physics of the deposition processes and the optical and electrical properties, and on their relationships with film structure, while keeping mathematical details to a minimum. As mentioned earlier, only the most important methods and properties are covered, although references to techniques omitted and fundamental factors are included to aid the acquisition of a deeper understanding, where this is required, and to appreciate the applications of thin films in those areas not covered elsewhere in this volume.

28.1 Deposition Methods

The deposition methods described in this section are those most commonly used. The physical methods of vacuum evaporation and sputtering are perhaps the most controllable, giving the best-quality films. In particular, molecular-beam epitaxy (MBE) is capable of providing very high-quality films. Examples of chemical methods, which are somewhat less flexible and limited to particular types of films, are also given. Perhaps the most important of these is CVD, but as will be seen the particular route chosen depends on the chemical properties of the material to be deposited.

28.1.1 Physical Deposition Methods

These could perhaps be described as physical vapour deposition (PVD) methods, since they involve direct deposition from the vapour phase, normally under reduced pressure. There are two main classes which are considered. The first of these is vacuum evaporation, which takes place at a low background pressure, i. e., high vacuum (HV) of the order of 10^{-4} Pa, although MBE systems normally operate at pressures significantly below this level. Secondly we have sputtering, which relies on the ejection of atoms or molecules from the surface of a material target by a mechanism of momentum transfer involving an impact from (usually) ionised inert gas atoms such as argon; in order for this to occur a sufficient inert gas pressure is required to sustain a discharge, and thus the typical operating pressure is of the order of 1 Pa. Thus evaporated films are liable to have a higher intrinsic purity than sputtered films, which are prone to the incorporation of sputtering gas

atoms. However, as these are generally inert there is often no deleterious effect on the film properties.

Vacuum Evaporation

During evaporation the evaporant is heated in vacuum by, for example, employing a high current through a refractory boat in which the evaporant is placed, or by using a high-current electron beam focussed on the evaporant. The purpose of this heating is to convert the condensed phase of the evaporant into the vapour phase, which then condenses elsewhere in the system in the form of a thin film. At a given absolute temperature T, there will in general be some of the evaporant in a condensed phase and some in the vapour phase. In thermodynamic equilibrium the Clausius–Clapeyron equation, relating the equilibrium vapour pressure p^* to the temperature T may be applied

$$\frac{\mathrm{d}p^*}{\mathrm{d}T} = \frac{L}{T(v_\mathrm{g} - v_\mathrm{c})}\,, \tag{28.1}$$

where L is the latent heat of vaporisation, representing the amount of energy required to convert a quantity of the condensed phase (liquid or solid) into the vapour phase at the given temperature. If L is expressed in J/mol, then v_g and v_c represent the molar volumes of the gaseous and condensed phases respectively. Although evaporation is not an equilibrium process, the importance of the equilibrium vapour pressure to evaporation is that maximum evaporation rates are dependent on the vapour pressure. Expressions for p^* may be obtained from (28.1) provided the temperature dependence of L is known. For the simplest assumption, where L is independent of T, we assume that the vapour obeys the ideal gas law. For one mole of the vapour

$$p^* v_\mathrm{g} = RT\,, \tag{28.2}$$

where R is the universal gas constant. Further assuming that the volume of the vapour phase is very much greater than that of the condensed phase ($v_\mathrm{g} \gg v_\mathrm{c}$), combining the above two equations and integrating, an expression for p^* is obtained. This can be expressed in logarithmic form as

$$\log_{10} p^* = AT^{-1} + B\,, \tag{28.3}$$

where $A = -(L/R)\log_{10}\mathrm{e}$ and B is a constant of integration. The constants A and B are tabulated by *Dushman* and *Lafferty* [28.20] for various metals using non-SI units, although in a slightly different form from the definitions of (28.3). In many cases such an expression for determining equilibrium vapour pressure is inadequate. The main reason for this is that the assumption of a constant latent heat L is invalid, particularly over a wide temperature range. When temperature variations in L are taken into account, (28.3) is modified by the addition of several additional terms, which may be determined from standard thermodynamic enthalpy and entropy data [28.21]. Tabulated values of coefficients for a logarithmic p^*–T expression involving four coefficients are given by *Kubaschewski* et al. [28.22] for various elements and compounds over defined temperature ranges in both the liquid and solid phases. Even more helpful are direct tabulated and graphical data of the p^*–T relationship in the case of various solid and liquid elements [28.21, 23, 24]. Examples of the p^*–T dependence are given in Fig. 28.1 for several elements using the data of *Honig* and *Kramer* [28.24]. From this figure it is clear that p^* increases very rapidly with temperature, and that most metals require temperatures of up to about 1500 K to attain significant vapour pressures; however, the refractory metals (Pt, Ta, W, etc.) require far higher temperatures for evaporation and thus they are not normally deposited by this method, although they are very useful materials for evaporation boat materials, which require very low vapour pressures and high melting points.

It is a well-known result from the kinetic theory of gases, that the molecules of a gas at a pressure p impinge on surfaces at a rate

$$N_\mathrm{i} = \frac{p}{(2\pi m k_\mathrm{B} T)^{1/2}} \tag{28.4}$$

where N_i ($\mathrm{m^{-2}s^{-1}}$) is termed the impingement rate, m is the mass of a gas molecule, k_B is Boltzmann's con-

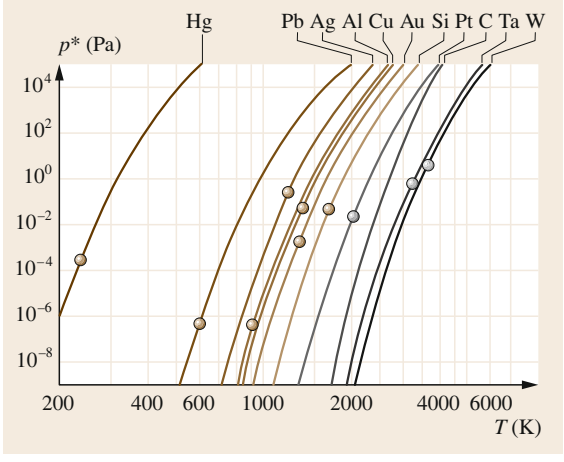


Fig. 28.1 Experimental dependence of the equilibrium vapour pressure p^* on absolute temperature T for several elements, derived from the compilation of *Honig* and *Kramer* [28.24]. The *point* (•) on each *curve* indicates the melting point, except for that of carbon, which sublimes

stant and T is the absolute temperature. In principle this equation applies equally well to residual gas molecules (N_2 and O_2) in an evaporation chamber, as well as to evaporant molecules in the vapour phase. It was noted previously that the equilibrium vapour pressure of a material largely determines the maximum evaporation rate. Hertz [28.25] observed for the case of mercury that evaporation rates were proportional to the difference between the equilibrium vapour pressure p^* and the reverse hydrostatic pressure p_h exerted at the surface of the evaporant. The *maximum* evaporation rate occurs when $p_h = 0$, and is equal to the impingement rate a vapour of the evaporant would exert at its equilibrium vapour pressure. This is given by an expression similar to (28.4), with pressure p^*, evaporation temperature T_e and mass of evaporant molecule m_e. This rate is rarely achieved, not only due to the hydrostatic pressure p_h, but also to a molecular reflection phenomenon proposed by *Knudsen* [28.26], which takes place at the evaporant surface. It was proposed that only a certain fraction α_v (the evaporation coefficient) of molecules make the transition from the condensed to the vapour phase. The net *molecular evaporation rate*, N_e, is then given by the *Hertz–Knudsen equation*

$$N_e = \frac{\alpha_v\,(p^* - p_h)}{(2\pi m_e k_B T_e)^{1/2}}. \tag{28.5}$$

Under equilibrium conditions $p_h = p^*$ and there is no net evaporation; however, under nonequilibrium conditions, especially when $p_h \ll p^*$, the evaporation rate can be considerable. Values of the coefficient α_v have been tabulated in the literature for a selection of different elements and compounds [28.27].

Equipped with vapour pressure and evaporation coefficient data, it is straightforward to estimate the evaporation rate using (28.5). The molecular deposition rate N_R represents the number of evaporated molecules deposited on the substrate per unit area per second. It is clearly proportional to the evaporation rate N_e, depending on the geometry of the evaporation/deposition system. N_R is simply related to the thickness deposition rate R (m/s) by

$$N_R = \frac{\rho R}{m_e}\,, \tag{28.6}$$

where ρ is the density of the depositing film. An indication of the film quality is given by the impingement ratio K, representing the ratio of the rate at which ambient gas molecules impinge on the substrate to the rate of evaporated molecules depositing on the substrate. Hence

$$K = \frac{N_i}{N_R} \tag{28.7}$$

From (28.4) and (28.6) it is clear that $K \propto p/R$ for a given material. In order to reduce contamination by gas incorporation into the film, it is therefore necessary to operate at a low background gas pressure with a relatively high deposition rate. A simple calculation for copper shows that for $R = 1\,\text{nm/s}$, $N_R \approx 8.5\times10^{19}\,\text{m}^{-2}\,\text{s}^{-1}$. If we require $K < 10^{-3}$ for impinging oxygen molecules then the maximum value of N_i allowed is $8.5\times10^{16}\,\text{m}^{-2}\,\text{s}^{-1}$, which from (28.4) corresponds at room temperature to a pressure of about $3\times10^{-6}\,\text{Pa}$. Such a pressure is in the range of ultrahigh vacuum (UHV), and is not attainable in normal vacuum deposition systems. Increasing the deposition rate is the normal method of lowering K without moving into UHV, as is used in MBE. This may in principle be achieved either by increasing the evaporation rate by increasing T_e to raise p^*, or by increasing the proportionality constant between N_e and N_R by shortening the distance between the evaporant and the substrate. A useful table of values of K for various combinations of chamber pressure and deposition rate is given by *Eckertová* [28.1].

There is in principle a second limitation on the level of the background pressure in a vacuum system which is also derived from kinetic theory. By using a Maxwell–Boltzmann distribution function for the molecular velocities in a gas, it can be shown that the mean free path between collisions in the gas λ is given by [28.21]

$$\lambda = \frac{k_B T}{p\pi\delta^2\sqrt{2}}\,, \tag{28.8}$$

where δ is the effective molecular diameter of a gas molecule. Values of δ are tabulated in the literature [28.20], and range from about 0.22 (He) to 0.49 nm (Xe). From (28.8) λ is inversely proportional to the pressure p, and this general relationship is shown for He, O_2 and Xe in Fig. 28.2. Since the values of δ do not vary greatly for different gases, the λ values for all common gases fall within a narrow range, as illustrated in the figure. The proportion of molecules $F(x)$ which have not suffered an intermolecular collision as a function of distance x since their last collision is given by

$$F(x) = \exp\left(-\frac{x}{\lambda}\right). \tag{28.9}$$

From this expression it is evident that for a deposition chamber size of about $\lambda/10$, less than 10% of molecules suffer a collision before hitting the walls of the chamber or the substrate, e.g., in a large vacuum chamber of dimensions about 1 m where the mean free path is about 10 m. From Fig. 28.2 a pressure of $10^{-4}\,\text{Pa}$ is sufficiently low to ensure that $\lambda > 10\,\text{m}$. In evaporation processes we require that the evaporant molecules

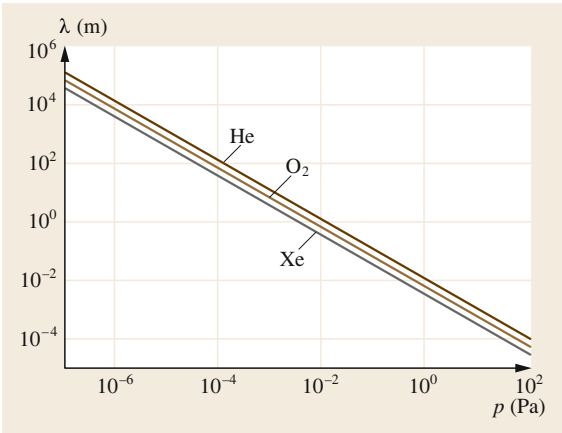


Fig. 28.2 The dependence of mean free path λ on pressure p for helium, oxygen and xenon at 300 K. Mean free paths are calculated from (28.8) using values of molecular diameter δ from *Dushman* and *Lafferty* [28.20]

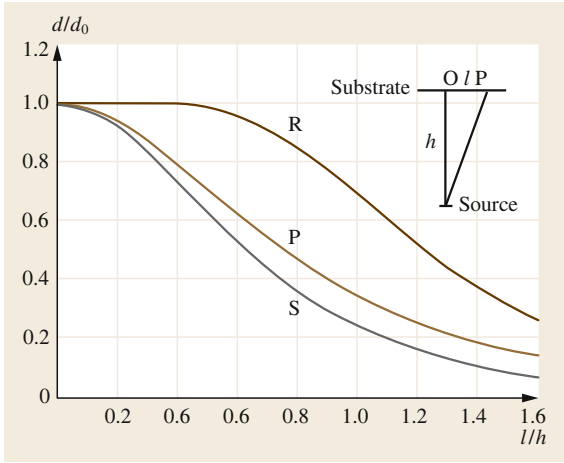


Fig. 28.3 Film thickness profiles calculated for evaporation from various sources onto a flat parallel substrate. Thickness ratio d/d_0 is given as a function of the ratio l/h, as shown in the *inset*. Sources include the small surface source (S), the point source (P) and a thin ring source (R) of radius b, where $b/h = 0.75$. Profiles are calculated, respectively, from (28.11), (28.10) and from an expression given by *Glang* [28.21]

are not significantly dispersed by collisions with gas molecules en route to the substrate, and since the mean free path is relatively insensitive to the molecular diameter, λ is a suitable criterion to determine this. Thus smaller deposition chambers will not require as low a residual gas pressure as larger ones.

The geometry of the evaporation source and its positioning with respect to the substrate is an important consideration in determining the uniformity of a thin-film deposit. Expressions for the film thickness variation over the substrate for various types of source have been derived by *Holland* and *Steckelmacher* [28.28]. Traditionally the *point source* and the small *surface source* are used. It is assumed that the evaporation source, of either type, is located a distance h below the centre of a flat substrate, as shown in the inset to Fig. 28.3. It is obvious that the film thickness deposited directly above the source at the point O will be thicker than elsewhere, since this is the closest point of the source to the substrate. For the case of the point source, molecules are emitted uniformly in all directions, and the ratio of the film thickness d deposited at the point P a distance l from O to the thickness d_0 deposited at O is given by

$$\frac{d}{d_0} = \frac{1}{[1 + (l/h)^2]^{3/2}} \,. \tag{28.10}$$

For a small surface source, the emission follows a cosine law, and the thickness distribution is given by

$$\frac{d}{d_0} = \frac{1}{[1 + (l/h)^2]^2} \,. \tag{28.11}$$

Plots of these two functions are shown on Fig. 28.3 for the point source P and the small surface source S. For both sources, there is a rapid diminution in the thickness ratio d/d_0, which falls off to 0.985 and 0.98 respectively for the point and surface sources at a value of $l/h = 0.1$. When the substrate is relatively small, the thickness uniformity is thus better than 2%. Uniformity can be enhanced, for these two small sources, by increasing the source–substrate distance h, but at the expense of consuming more evaporant material. It has also been shown that a substantially uniform deposit will be obtained for a point source if the source is placed at the centre of a sphere and the substrates placed on the surface. For a surface source a uniform thickness distribution is obtained if both the source and the substrates are placed on the surface of a sphere [28.28].

Other systems for obtaining uniform thickness are based on the use of either extended sources, or substrates rotating around a central axis with respect to an off-centre source, such that variations in the thickness average out. *Holland* and *Steckelmacher* [28.28] have addressed ring sources in addition to strip and cylindrical sources evaporating onto a plane-parallel substrate. The ring source has also been treated by *von Hippel* [28.29] and the small surface source evaporating onto a rotating plane parallel substrate has been discussed by *Macleod* [28.30]. Expressions for ring and large surface sources are given by *Glang* [28.21]. A ring source can be considered to be a narrow ring of radius b, which is usually placed parallel to a flat substrate. This introduces an additional controllable

variable which influences the thickness distribution. Thickness distributions may be obtained by calculating the distribution arising from an element of the ring source and then summing over all elements in the ring. For small b/h the thickness distribution reduces to that of the small surface source. For b/h larger than about unity the uniformity is also poor as the maximum in the distribution moves away from the central axis towards the source radius. However, for intermediate ranges, in particular b/h of about 0.7–0.8 there is considerable uniformity in thickness up to l/h of about 0.6. This is also illustrated in Fig. 28.3, where the curve R shows the thickness distribution for a narrow ring source with $b/h = 0.75$, calculated from the formula given by *Glang* [28.21]. Extended disc source thickness distributions may be obtained by integrating the thickness distribution derived for a narrow ring source from radius zero to radius b, but in general the thickness uniformity is not as good as for ring sources. Expressions for ring source of finite width can be determined by subtracting the disc source distribution for the inner radius from that of the outer radius. Diagrams illustrating the thickness distributions for several sets of ring and disc sources are given by *Glang* [28.21]. *Behrndt* [28.7] gives further details and references for such sources, and a discussion of the results of using combinations of several rod-like sources. Further attempts to minimise thickness variations have resulted in the evolution of various *planetary* deposition systems, so called because the rotating substrate plane is itself rotated around a vertical axis. A typical example of such a system and its performance is given in the literature [28.31].

Before describing the practical details of the evaporation process, it should be mentioned that, whatever type of evaporation process is used, some consideration needs to be given to the substrate. There are three basic types of substrate commonly used: glasses, single crystals and polycrystalline ceramics. Ceramics are particularly useful for high-temperature deposition processes. Although the substrate material may be predetermined (e.g., silicon wafers for microelectronics, optical glass for coatings on lenses, etc.), there are frequently factors which can be selected to increase the probability of obtaining good-quality films. The substrate planarity is important, in order to reduce the effect of defects. Electrical resistivity should in general be high in order to minimise surface leakage currents. Thermal properties include a consideration of the coefficient of expansion, which should be comparable to that of the deposited film to reduce the probability of thermally induced stresses in the films. Thermal conductivity should be high for films whose applications will entail significant power dissipation. A full description of thin-film substrates and the associated problems of substrate cleaning are discussed in detail by *Brown* [28.32].

Resistive evaporation is the most commonly used evaporation technique, although it is unfortunately not suitable for certain applications. It entails the resistive heating of an evaporation source, by passing a high current either through the source itself or through an adjacent heater. There are a large number of practical evaporation sources for various applications which are described in the literature [28.1, 21, 28]. The main requirement for any evaporation source is that the source itself will not emit its own vapour at the operating temperature. If metallic sources are used, the equilibrium vapour pressure of the metal must be negligible in comparison to that of the evaporant. A glance at Fig. 28.1 shows that the refractory metals (such as Pt, Ta and W) have equilibrium vapour pressures very low compared to about 1 Pa at temperatures up to about 2000 K, and are therefore suitable for evaporation of many materials. Carbon and oxides of aluminum or magnesium may also be used in evaporation, although owing to their low electrical conductivity refractory metal heaters are often used in conjunction with these. However, at elevated temperatures the oxides themselves may dissociate, with the liberated oxygen reacting with the evaporant to form its own oxide. For this reason the refractory metals cannot themselves be evaporated from oxide crucibles. Furthermore carbon may react with metals to form carbides well below 2000 K. Having chosen a suitable evaporation source, the vacuum deposition chamber is pumped down to the required background pressure and deposition may commence. The temperature of the source is increased such that the vapour pressure increases to a value consistent with a reasonable evaporation and hence deposition rate. The substrate is normally shielded using a metallic shutter until the required deposition rate is established, when the shutter is removed. The deposition rate is usually monitored by one of several methods [28.7, 21]. The quartz crystal method entails depositing the evaporant material on a vibrating quartz crystal, whose frequency varies linearly with the thickness of material deposited [28.33, 34]; the deposition rate is obtained by electronically differentiating the thickness signal with respect to time. Optical thickness monitoring methods are also used, particularly in applications where films are required for optical purposes, and require films having an optical thickness of multiples of $\lambda/4$, where λ is the wavelength for which the optical film is designed [28.7, 21]. These thickness multiples correspond to maxima and minima in reflected light intensity, which can easily be monitored during deposition. When the required film thickness is attained, a shutter is again used to

curtail further film deposition and the evaporation process is stopped. Sequences of several film layers may be deposited by using a suitable sequential masking system [28.35].

Apart from simple vacuum evaporation there are also several other types of evaporation, which entail using alternative methods to heat the evaporant, or cater for various problems encountered in the deposition of alloys and compounds. In electron-beam evaporation the evaporant material is placed in a ceramic hearth, which is bombarded by high-energy electrons The main advantage of this method is that considerably higher temperatures of above 3000 K may be obtained than with resistive evaporation methods. This allows evaporation of the refractory metals and also elements such as boron and carbon (which sublimes). Other advantages are the evaporation of reactive metals such as Al, which are prone to contamination due to alloying, and situations where high purity is required.

For non-elemental (multicomponent) films there are several methods which may be employed. The main problem with such films is that, because of differences between the vapour pressures of the evaporant components, the Hertz–Knudsen expression of (28.5) predicts that the compound will evaporate noncongruently, i. e., with different evaporation rates for each element, and the deposited film will not in general replicate the composition of the evaporant. The problem is compounded because the proportion of each component present will therefore vary throughout the evaporation process, and the overall composition of the deposited films will vary throughout its thickness. In the evaporation of alloys, each component establishes its own vapour pressure at a given temperature, which may be predicted by *Raoult's law* for an ideal solution. This states that the vapour pressure established by each individual component is proportional to the mole fraction of the component present. Under these circumstances it is relatively simple to adjust the proportions of the alloy components to achieve the required film composition. However, complications occur in nonideal solutions, where the interactions between different types of atoms vary and an empirical parameter known as the activity coefficient, a, is defined. Knowing this it is possible to determine the vapour pressure for the component and hence the evaporation and deposition rates. Activity coefficients for various important systems, such as nichrome (a Ni-Cr alloy), are given in the literature [28.36]. However, as noted previously the mole fractions present in the evaporant are time-dependent and the composition of the films will vary during deposition. A parameter K has been determined, which if maintained close to unity should result in congruent evaporation throughout the entire deposition process. Further details are given by *Zinsmeister* [28.37]. Additional comments on the evaporation of both nichrome and permalloy alloys are given by *Glang* [28.21].

On evaporation of compounds, a wealth of chemical reactions can occur, which leads to a variety of different evaporation mechanisms. Frequently more than one reaction may operate simultaneously. Four simple generic reactions have been listed by *Glang* [28.21], which include many of the materials which can be successfully evaporated. We consider only compounds AB composed of the elements A and B, and denote the phase of the material by s (solid), l (liquid) or g (gas). Some compounds (solids or liquids) do not dissociate on evaporation and the reaction can be expressed simply as

$$\mathrm{AB(s,l)} \rightarrow \mathrm{AB(g)}\,. \tag{28.12}$$

There is then only the single vapour phase AB(g) produced. Other compounds dissociate on evaporation into two different vapour species. One reaction which is typical of the evaporation of chalcogenides and some simple oxides is given by

$$\mathrm{AB(s)} \rightarrow \mathrm{A(g)} + \frac{1}{2}\mathrm{B_2(g)} \tag{28.13}$$

whereas the group IV dioxides tend to dissociate according to the reaction

$$\mathrm{AB_2(s)} \rightarrow \mathrm{AB(g)} + \frac{1}{2}\mathrm{B_2(g)}\,. \tag{28.14}$$

In compounds of elements whose volatilities are significantly different, frequently one element enters the vapour phase while the other remains in the solid (or possibly liquid) phase. The compound decomposes into its different elements, with the simplest reaction of this type being

$$\mathrm{AB(s)} \rightarrow \mathrm{A(s)} + \frac{1}{2}\mathrm{B_2(g)}\,. \tag{28.15}$$

Examples of materials evaporating without dissociation following (28.12) are the simple oxides GeO, SiO, SnO, the fluorides CaF_2 and MgF_2, and PbS. In general the evaporation of these materials can therefore be considered as analogous to those of the elements, and where available simple p^*–T data give the vapour pressure. Materials which follow (28.13) are primarily the II–VI chalcogenides, in particular the cadmium compounds CdS, CdSe and CdTe, and several simple oxides such as BaO, BeO, CaO, MgO, NiO and SrO. A full review of the electrical and structural properties of the cadmium chalcogenides is given by *Gould* [28.38].

Equation (28.14) is followed by many of the group IV dioxides, such as SiO_2, SnO_2, TiO_2 and ZrO_2, which dissociate into their lower oxides and molecular oxygen. Decomposition, as described by (28.15) is followed by most of the metallic borides, carbides and nitrides [28.21]. Other modes of decomposition, such as those for the technologically important III–V compounds, are also described in this reference.

Provided the relevant reaction for the evaporation process is known (e.g., as given by (28.12)–(28.15)) it is possible to determine the equilibrium vapour pressure using standard thermodynamic tables. The law of mass action may be used to determine the equilibrium state of the chemical reaction, and for evaporation, the pressures of the various phases present contribute to a temperature-dependent equilibrium constant K_p, which is given by

$$K_\mathrm{p} = \prod_i (P_i)^{\nu_\mathrm{i}} \qquad (28.16)$$

where P_i represents the vapour pressure for each of the various components of the vapour expressed in standard atmospheres and the ν_i are the molecular coefficients. K_p is related to the thermodynamic parameter ΔG^0, the standard change in Gibbs free energy per mole for the reaction, by the expression

$$\Delta G^0 = -RT \ln K_\mathrm{p} \,, \qquad (28.17)$$

where R is the universal gas constant. Free-energy data is tabulated for many materials in the literature [28.22, 39] and the use of (28.16) and (28.17) allows realistic estimates of the equilibrium vapour pressures to be determined and the evaporation rates estimated. Further discussions are given by *Glang* [28.21]. Nevertheless, although the theoretical framework is in place for the controlled evaporation of compounds, the exact type of dissociation and decomposition are not known in all cases, and for compounds consisting of three or more elements this problem is compounded. Some materials are therefore not amenable to direct evaporation from a single source.

For these cases various special evaporation techniques have been developed, all of which aim to replicate the composition of the evaporant in the deposited film, by ensuring that the different molecules arrive at the substrate in the desired proportions and at a constant rate. Reactive evaporation is a process whereby one component of the desired film is evaporated resistively in the normal manner, while the other component is present in the evaporation chamber in the form of a gas. This method is most commonly used where oxide films are required and oxygen gas present in the chamber combines with a metal at the substrate to form the oxide. The impingement ratio K, given by (28.7) is optimised so that evaporant and gas molecules reach the substrate in a predetermined ratio. Oxides of the common metals have been deposited by this method, including those of Al, Cr, Cu and Fe; additionally tantalum and titanium oxides have been prepared for dielectric applications. Reactive evaporation has also been used for compounds other than oxides. For instance, stoichiometric CdS films have been obtained by the evaporation of CdS in a sulphur vapour [28.40]. The sulphur vapour has the effect of discouraging the appearance of nonstoichiometric Cd-rich films which occur when CdS is evaporated alone. Nitrides may also be deposited in some cases by the use of a nitrogen atmosphere. The utility of this technique has been demonstrated in the reactive evaporation of silicon oxide, SiO_x, by *Timson* and *Hogarth* [28.41]. The evaporant is a mixture of the monoxide, the dioxide and possibly also free silicon. A sequence of films, varying in composition between SiO and SiO_2 as determined by electron spin resonance studies were obtained, depending on the ratio of the deposition rate to the oxygen pressure R/p. A useful nomogram for use in reactive evaporation is given by *Glang* [28.21].

Co-evaporation, sometimes known as the three-temperature technique, involves the simultaneous evaporation of two or more materials from separate evaporation sources. The essential feature of this process is that the rate of molecular arrival at the substrate may be controlled by individually controlling the temperatures of the evaporation sources. A major advantage of this method is that materials with widely different vapour pressures at a given temperature may be co-deposited by increasing the temperature of the less-volatile material. This method has found applications in the deposition of II–VI compounds such as CdS and CdSe, and also the III–V compounds such as AlSb, GaAs, InAs and InSb. It is also used for the deposition of alloy films.

Flash evaporation is another technique used for deposition of layers whose components have widely different vapour pressures. The evaporation is constrained so that only very small amounts of the compound completely vaporise almost instantaneously. The various components normally evaporate noncongruently, but because only very small quantities are evaporated rapidly, any inhomogeneity in the deposited films is limited to a few atomic layers. The process is repeated by evaporating further small quantities of evaporant. Several alternative arrangements have been developed to evaporate small particles, by continuously feeding powdered evaporant from a hopper into a heated evaporation source. The rate of delivery needs to be

controlled such that at any given time there will be a number of evaporant particles in the source at various stages of the evaporation process. This ensures that the vapour emerging from the source contains contributions from many individual particles, resulting in an overall vapour composition which approximates to that of the evaporant powder. The method has been used particularly for III–V compound semiconductors such as GaAs.

Finally, the method of molecular-beam epitaxy (MBE) should be included here as a particularly sophisticated example of evaporation (or co-evaporation). In MBE the background pressure is considerably lower than in the case of the simpler evaporation techniques, i. e., well into the UHV range. This effectively eliminates the effects of background gas impingement, drastically improving the purity of the films. Several sources (effusion ovens) may be used simultaneously, each of which is separately controlled. These permit the deposition of compounds and precisely controlled doping of semiconductors. Sophisticated monitoring and isolation techniques are employed, and deposition sequences are usually under computer control. MBE is used particularly for the preparation of III–V compounds [28.42, 43] and silicon [28.44]. Figure 28.4 illustrates a typical MBE system for use in the deposition of III–V compounds [28.42]. A full description of MBE, together with other epitaxial growth techniques is given elsewhere in this volume in Chap. 14.

Sputtering

The technique of sputtering in its various forms is of secondary importance only to evaporation for the deposition of thin films, offering certain advantages for the deposition of high-melting-point and dielectric materials. Sputtering is essentially the removal of particles from the surface of a *target* of the deposition material, by the action of incident energetic particles (normally positive ions). It is generally considered that the sputtering mechanism is the result of the transfer of momentum from the incident particles to the target atoms, some of which become dislodged. Energy acquired by other atoms interacting with the dislodged atoms can be sufficient to overcome the surface binding energy potential barrier. The net effect of a large number of atomic interactions is for a certain proportion of the target to be released (or sputtered) from the surface. Sputtered atoms typically have energies much higher than their evaporated counterparts. For low-energy sputtering with incident energy 1 keV, an ejected particle may have an energy of 10 eV, since typically 1% of the incident energy is transferred to each sputtered atom. In contrast, the thermal ($k_B T$) energy of an atom evaporated from a source at 2000 K is less than 0.2 eV. Such considerations have implications regarding the sticking coefficients for deposited materials (i. e., the probability that an atom reaching the substrate condenses on it), which is generally higher for higher-energy atoms.

Sputtering is limited by a threshold energy, below which it will not occur for a given ion/target combination. Threshold energies for a wide range of incident ions and target materials are of the order of 20–30 eV; experimental values for some common ion/target combinations are given by *Wehner* and *Anderson* [28.45]. These workers have pointed out that the threshold energy appears to be approximately four times the latent heat of sublimation per atom, equivalent to the surface

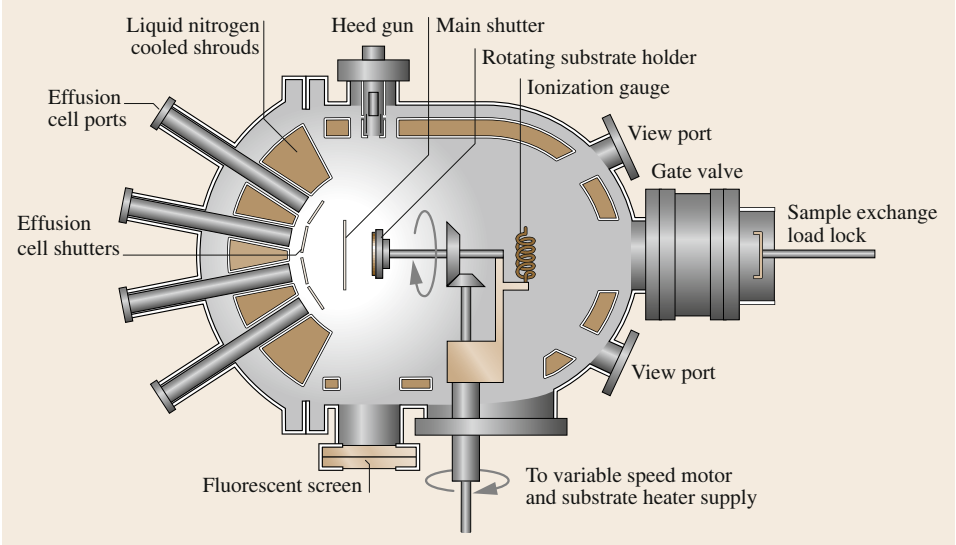


Fig. 28.4 Schematic diagram of a molecular-beam epitaxy (MBE) system for the deposition of III–V semiconductor compounds. (After [28.42] with permission from Elsevier Science)

binding energy. However, for practical sputtering considerations, incident energies are normally at least a few hundred electronvolts, well above the threshold energy. The efficiency of sputtering, measured by the sputtering yield, is also very low at low incident ion energies.

The sputtering yield is the most important sputtering parameter, and its variation with the target material and the incident ion energies is of major concern in both theories of sputtering and in practical sputtering deposition techniques. Sputtering yield $S(E)$ is defined as the ratio of the number of sputtered atoms to the number of incident ions. Hence

$$S(E) = \frac{N_s}{N_{ion}} , \tag{28.18}$$

where N_{ion} is the impingement rate of incident ions and N_s is the sputtering rate of the target. Its primary variation is with the incident ion energy E. However, it also depends on the atomic properties of the incident ions and the target atoms. In the following discussion Z_1 and m_1 refer to the atomic number and mass of the incident ions, and Z_2 and m_2 refer to the corresponding quantities for the target material. The dependence of sputtering yield on the atomic number of the incident ions is quite striking, as shown in Fig. 28.5 [28.46]. This shows the variations of $S(E)$ for 45 keV incident ions onto targets of silver, copper and tantalum. Maxima in $S(E)$ occur corresponding to the noble gases Ne, Ar, Kr, and Xe. This dependence is one reason why the noble gases are most commonly used as the incident ion species, the second being that their inclusion into deposited films is unlikely to cause significant problems, due to their chemically inactive nature.

A number of theories have been developed to predict the dependence of sputtering yield $S(E)$ on the incident beam energy E. *Keywell* [28.47] assumed that the incident ions interacted with the sputtering target as if they were hard spheres, with the incident ion energy decreasing exponentially with the number of collisions. A discussion of this and other early models is given in the literature [28.45, 48]. More recent models [28.49, 50] have taken account of the scattering of incident ions by the atomic nucleus, and screening effects due to the electron cloud are also included, based on the Thomas–Fermi potential. This predicts a characteristic screening length a, which depends on the atomic numbers of the incident and target atoms, Z_1 and Z_2, respectively. The screening length is given by

$$a = \frac{a_0}{\left(Z_1^{2/3} + Z_2^{2/3}\right)^{1/2}} , \tag{28.19}$$

where $a_0 = 5.29 \times 10^{-11}$ m is the Bohr radius of the ground state of the hydrogen atom. Another parameter which is often invoked in sputtering models is the energy transfer coefficient, γ. For an incident ion of

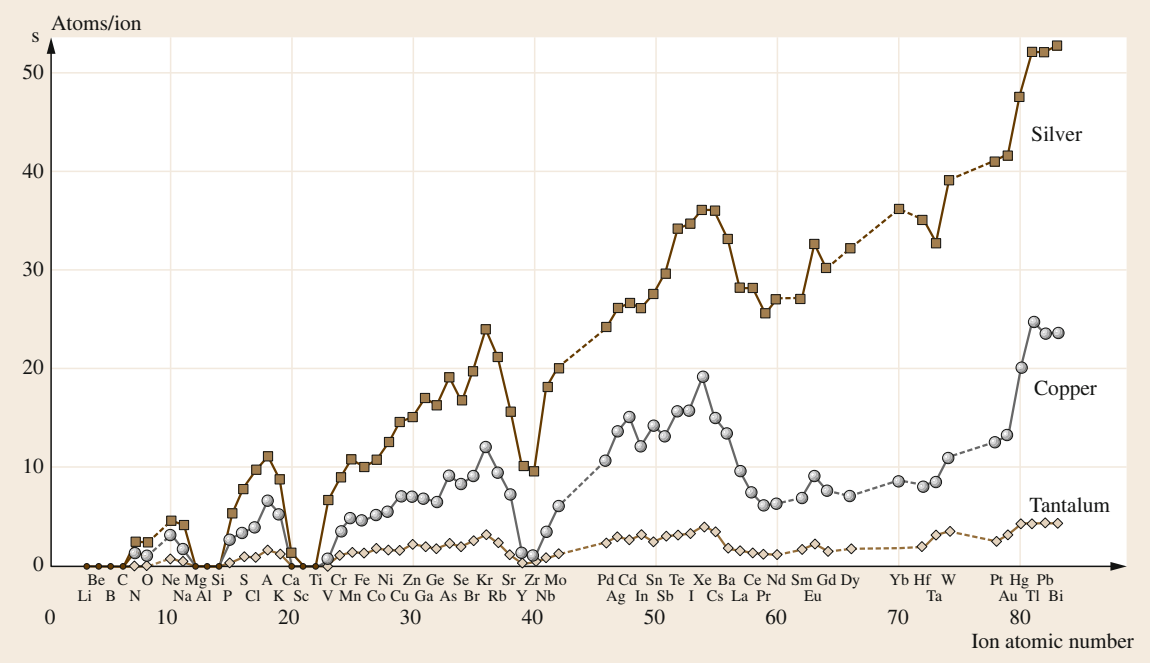


Fig. 28.5 Sputtering ratios of silver, copper and tantalum targets for different ions of energy 45 keV. (After [28.46] with permission from Elsevier Science)

mass m_1 and energy E, impacting on a stationary target atom of mass m_2, the maximum energy E_{max} the target atom will attain is given by

$$E_{max} = \gamma E \,, \tag{28.20}$$

where

$$\gamma = \frac{4m_1m_2}{(m_1+m_2)^2} \,. \tag{28.21}$$

In most collisions the incident ion will lose some energy to the target lattice, resulting in the acquisition of an energy less than E_{max} by the target atom. γ has a maximum value of unity for $m_1 = m_2$, but this is neither a sufficient nor a practical condition for obtaining high sputtering yields owing to the impracticability of having identical sputtering ions and target material as well as the striking periodic dependence of $S(E)$ on Z_1 noted earlier. Two well-known sputtering models are now outlined, both of which have been reasonably successful in predicting sputtering yields.

In the model of *Pease* [28.49], which is also outlined elsewhere in the literature [28.51], ions are considered to interact with the target atoms near to the surface, which if displaced from their lattice sites are termed *primary knock-ons*. The mean energy of primary knock-ons before further interactions is denoted by $\overline{E}$. The threshold energy for displacement of a target atom, either by an incident ion or by a primary knock-on is denoted by E_d. Collisions with stationary atoms result in slowing down of displaced atoms, and in some cases a reversal of momentum and escape from the target surface, providing they still retain energy greater than E_s, the surface binding energy per atom. The number of atomic layers which contribute to the sputtering process is established by considering the average number of collisions n that a primary knock-on of energy $\overline{E}$ will make before its kinetic energy is reduced to less than that required to leave the target surface, i.e E_s. This is given by $2^n = \overline{E}/E_s$ or

$$n = \frac{\ln(\overline{E}/E_s)}{\ln 2} \,. \tag{28.22}$$

While engaged in n collisions the primary knock-ons diffuse on average $n^{1/2}$ atomic layers toward the surface, and thus $(1+n^{1/2})$ layers are involved in sputtering of material liberated by primary knock-ons only. The total number of atomic layers involved in sputtering is also equal to this expression. We also have to consider the probability of interaction in a given atomic layer, and denote the number of atoms per unit volume in the target by N. If these atoms are considered to be arranged in uniform crystallographic layers in three dimensions, the atoms in each layer will have an area density of $N^{2/3}$. If σ_p denotes the cross section for interaction between incident ions and target atoms, the probability of interaction in a given atomic layer is given by the product of the number of atoms per unit area and their interaction cross sections, or $\sigma_p N^{2/3}$. The total probability of producing a primary knock-on which can contribute to the sputtering process is then given by the product of the probability of interaction in a given atomic layer and the number of layers involved in sputtering, i.e., $\sigma_p N^{2/3}(1+n^{1/2})$. An estimate of the number of atoms displaced per primary knock-on is $\overline{E}/(2E_d)$ [28.52] of which half, or $\overline{E}/(4E_d)$, will be directed towards the surface. The total sputtering yield $S(E)$ is then given by the product of the probability of producing primary knock-ons and the number of displaced atoms produced by each. Hence

$$S(E) = \sigma_p \overline{E} N^{2/3} \frac{1}{4E_d} \left[1 + \left(\frac{\ln(E/E_s)}{\ln 2}\right)^{2/3}\right] \tag{28.23}$$

where the expression for n in (28.22) has been substituted.

In (28.23) for the sputtering yield, N, E_d and E_s are constants depending on the target material and E_d is normally determined from radiation damage experiments. However, σ_p and $\overline{E}$ are energy-dependent variables, which depend on the atomic numbers Z_1 and Z_2, the Thomas–Fermi screening length a, and the maximum target-atom energy E_{max}. It is not appropriate to give these expressions here, but it is worth mentioning that different expressions are applicable in different incident ion energy ranges, defined by $E < L_A$, $L_A < E < L_B$, and $L_B < E$. Further details concerning the threshold energies L_A and L_B are given in [28.49, 51]. Calculations of sputtering yield predicted by this model are reasonably consistent with experiment, showing a sharp increase above a certain threshold energy which is determined by E_s, a broad maximum and a relatively slow decrease at higher energies [28.49].

The model of *Sigmund* [28.50] represents a comprehensive theory of sputtering, particularly for single-element solids, although this has led to a certain degree of complexity in its general form. The model has been revised and extended [28.48] and tested against a considerable amount of data produced by many different workers [28.53]. The emphasis in this section is to briefly outline the model, to provide a basis for the calculation of $S(E)$ at normal incidence for low- and high-energy sputtering, and to give useful references to the required data. Various approximations made and some of the conditions of validity are omitted. Being

a multiple-collision process, sputtering may be treated by using the formalism of transport theory, and in the Sigmund model the Boltzmann transport equation is solved. The sputtering yield at normal incidence is given by

$$S(E) = \Lambda\alpha N S_n(E) \,, \tag{28.24}$$

where Λ is termed the material factor, and depends only on the target material. α is a dimensionless factor, effectively depending only on the ratio of the target atom to incident ion masses m_2/m_1. The factor α is given for various different conditions in the literature [28.48, 50]. As in the model of *Pease* [28.49], N represents the number of atoms per unit volume in the target. $S_n(E)$ is the nuclear scattering cross section, depending on the interaction probability between the incident ions and the target atoms at a given energy.

The material factor Λ arises in the sputtering yield expression as a result of integrating the number of sputtered atoms emitted in all directions with all possible energies. According to Sigmund, the material factor

$$\Lambda = \frac{3}{4\pi^2}\frac{1}{NC_0E_s} \,, \tag{28.25}$$

where $C_0 = \frac{1}{2}\pi\lambda_0 a_{BM}^2$, λ_0 is a constant approximately equal to 24, and $a_{BM} = 2.19\times10^{-11}$ m is the Born–Meyer potential characteristic screening radius. E_s represents the surface barrier energy, as in the model of Pease. Making these substitutions into (28.24) gives a more useful formula for the sputtering yield [28.45] as

$$S(E) = \frac{1}{16}\frac{\alpha S_n(E)}{\pi^3 a_{BM}^2 E_s} \tag{28.26}$$

which applies both for low- and high-energy sputtering. An expression for E^*, the boundary between low-energy and high-energy sputtering, is given in the literature [28.50], and is typically a few hundred electronvolts, depending on the atomic masses and atomic numbers of the incident ions and the target atoms. For low-energy sputtering ($E < E^*$) the nuclear scattering cross section can be approximated by

$$S_n(E) = C_0E_{max} = \tfrac{1}{2}\pi\lambda_0 a_{BM}^2 E_{max} \,, \tag{28.27}$$

where $E_{max} = \gamma E$ from (28.20). Substituting into (28.26) yields

$$S(E) = \frac{3}{4}\frac{\alpha\gamma E}{\pi^2 E_s} \,. \tag{28.28}$$

Hence for low-energy sputtering $S(E)$ is directly proportional to the sputtering energy, where α and γ effectively depend only on the ratio m_2/m_1 and E_s is a constant for the target material.

For higher incident ion energies ($E > E^*$) the Thomas–Fermi, rather than the Born–Mayer interaction potential is applicable, and a simple approximation for $S_n(E)$ is not available. It is therefore necessary to determine its value from tables or graphical data, calculating $S(E)$ directly from (28.26). Incident ion energies E and nuclear scattering cross sections $S_n(E)$ may be expressed in terms of reduced or Thomas–Fermi variables ε and $s_n(\varepsilon)$ [28.48, 50]. $s_n(\varepsilon)$ is a universal function of ε, depending on the details of the screened Coulomb function used [28.48]. The detailed dependence of $s_n(\varepsilon)$ on ε utilised in the Sigmund model is that given by *Lindhard* et al. [28.54]. For ease of computation E_{TF}, the Thomas–Fermi energy unit has been calculated for various incident ion/target combinations [28.48]. E_{TF} represents the ratio E/ε, and thus ε and hence $s_n(\varepsilon)$ may be obtained. $S_n(E)$ may then be determined from the reduced value.

Agreement between predicted and measured sputtering yields as a function of incident ion energy are good. *Andersen* and *Bay* [28.53] have reviewed sputtering yield measurements at normal incidence for a considerable number of incident ion and target species, and generally concluded that sputtering yields for high-yield materials (e.g., Ag, Au, Cu, Zn) are predicted well by the theory, while for low-yield materials (e.g., Nb, Ta, Ti, W) it was over-estimated by a factor of up to three. A fuller discussion, to which the reader is referred, shows that the model breaks down for certain combinations of Z_1, Z_2 and E due to the existence of inelastic collisions, nonisotropic and nonlinear effects (spikes) in the collision cascade and unallowed-for surface effects. The regions of validity of the model are neatly presented in terms of a three-dimensional (Z_1, Z_2, E) space, and illustrate the fact that (28.26) is an amazingly good prediction for a large number of (Z_1, Z_2, E) combinations of practical interest.

Sputtering yields may be determined experimentally by measuring the decrease in mass of a sputtering target. Normally such experiments are performed using ion-beam sputtering, in a fairly high-vacuum environment. The target area bombarded by the beam is small, and fairly low sputtering rates are achieved, but it has the advantage that ion energies are accurately given by the accelerating potential. The methods of sputtering normally used for the preparation of thin films are based on the establishment of a population of positive ions in a low-pressure gaseous environment. Such techniques depend on the establishment of a glow discharge, and require a considerably higher operating pressure, whereas ion-beam sputtering does not require the presence of gas molecules in the deposition cham-

ber, and can therefore be used under UHV conditions, where gaseous contamination effects are minimised. A complicating factor, which often has to be taken into account, is that some incident ions may become embedded in the target, thus tending to increase its mass. The net loss of material from the target Δm is then the difference between the mass of material sputtered and the mass gain of the target. If N_{ion} and N_{s} are the incident ion impingement rate and the atomic sputtering rate of the target respectively (28.18), and m_1 and m_2 are the incident ion and target atom masses, it is simple to calculate the mass loss and increase of a given target with bombarded area A sputtered for a time t. The loss of mass is given by $N_{\text{s}} m_2 A t$, and the increase in mass is given by $N_{\text{ion}} \Gamma m_1 A t$, where Γ is the probability that an incident ion becomes embedded in the target. Hence

$$\Delta m = N_{\text{s}} m_2 A t - N_{\text{ion}} \Gamma m_1 A t \qquad (28.29)$$

or

$$S(E) = \frac{N_{\text{s}}}{N_{\text{ion}}} = \frac{\Delta m}{m_2 N_{\text{ion}} A t} + \Gamma \frac{m_1}{m_2} \,. \qquad (28.30)$$

Frequently the second term on the right-hand side of this equation is negligible, and the sputtering yield is then given by the first term only. This is particularly applicable for large ion doses, while for intermediate doses a detailed empirical knowledge of the value of Γ is required. For low ion doses the approximation $\Gamma = 1 - R_0$, where R_0 is an ion reflection coefficient, may be used [28.53]. It is often useful to express the sputtering yield in terms of the ion current at the cathode $I_{\text{c}} = e N_{\text{ion}} A$, where e is the electronic charge and the sputtering yield is then given by [28.1]

$$S(E) \approx \frac{e \Delta m}{m_2 I_{\text{c}} t} \,. \qquad (28.31)$$

Measurements of the ion current, target mass loss and the sputtering time are then all that are necessary to calculate $S(E)$. Although strictly speaking the above expressions are only applicable for sputtering performed with ions of a single energy, they may also be applied to a first approximation for practical sputtering methods using the glow discharge.

Although ion-beam sputtering is a very useful technique for measuring sputtering yields and comparing with theoretical predictions, it is not normally used for routine film deposition, owing to the need for UHV conditions, the low sputtering rate obtained and the associated expense. Virtually all sputtering methods used for the deposition of thin films utilise the glow discharge phenomenon. When a gas at reduced pressure is subjected to an applied voltage, usually in the range of a few hundred to several kilovolts, any free electrons will be accelerated in the field, acquiring energy and ionising the gas molecules to produce a plasma. Under the influence of the field the positively ionised gas molecules are accelerated towards the cathode and free electrons towards the anode. The gas pressure required for the establishment of such a discharge is of the order of a few pascals, which in turn is determined by the molecular concentration required to maintain the mean free path within suitable limits to initiate and sustain the discharge. The basic principles of the glow discharge are well known and documented [28.55, 56] and need not concern us here. However, it should be noted that the positive ions tend to accumulate in front of the cathode (the cathode-fall region), across which most of the applied voltage is dropped. Thus the electric field between the cathode and the anode is distinctly nonlinear, and most of the energy acquired by the ions in their journey to the cathode is obtained in the cathode-fall region.

It is relatively simple to set up a self-sustained glow discharge, although the behaviour of the discharge depends in a fairly complex manner on the pressure, the applied voltage and the geometry of the system. Inert gases such as argon are generally used, unless the sputtered film relies on a reaction between the sputtering gas and the sputtered material to produce the required compound film. Once the discharge is established, the cathode is continuously bombarded by energetic ions, and provided the energy of these ions is sufficient, sputtering of the cathode will occur. Thus if the target is sufficiently conductive, all that is necessary is to make the cathode of the target material. Substrates on which deposition will occur are normally placed parallel to the target at the anode so that the sputtered atoms need only cover a short distance before deposition. The energy of the sputtered atoms when they reach the substrate is considerably higher than in the case of evaporation, leading to better adhesion. Furthermore, sputtered compounds are normally deposited without dissociation, and therefore the stoichiometry of the target material is usually preserved in the deposited film. This basic form of sputtering is normally termed *diode sputtering*, since only two electrical connections are required.

Basic diode sputtering operates reasonably satisfactorily, but has a major disadvantage in that the sputtering rate N_{s} depends directly on the impingement rate of the incident ions N_{ion} and thus also on the discharge current. At low gas pressures the supply of ions is limited and thus the discharge current and the sputtering rate are low. Moreover, at sufficiently low pressures the electrons do not ionise a sufficient number of gas atoms to sustain the glow discharge. This low-pressure limit is in fact a major drawback in diode sputter-

ing, since enforced operation at higher pressures above about 3 Pa results in significant film contamination. For this reason sputtering was for many years rejected in favour of evaporation for thin-film deposition. A major improvement in sputtering technology, which enabled sputtering to be performed at lower pressures was the introduction of *triode sputtering*. In this technique, the discharge is maintained by using an additional source of electrons, so that it is not totally dependent on secondary electron emission from the cathode. Electrons are supplied by an additional thermionic emitter, consisting of a filament through which a current is passed. The electron concentration in the plasma can be readily controlled by varying the anode potential or the filament current, while the ion current is directly controlled by the target potential. A further innovation usually employed in triode sputtering systems is the use of a coil to set up a magnetic field, which forces electrons from the emitter to travel in spiral paths, thus greatly increasing the distance travelled from the emitter before they are collected by the anode. This enhances the ionisation probability α, and thus the supply of positive ions.

It should be emphasised that both diode and triode sputtering may not be used for the sputtering of insulators – a very important class of materials. Since positive ions are responsible for the sputtering process, these need to be neutralised by electrons from the external circuit. In the case of insulators a positive charge builds up at the surface of the target, and this charge cannot leak away through the insulator. Sputtering ceases when the positive charge causes the potential at the target surface to approach that of the plasma. Various methods, such as bombarding the target surface with electrons, have been attempted to solve this problem, but in general with little success. Glow discharge sputtering of insulators is generally therefore performed reactively, or with alternating radio-frequency (RF) fields, as described later.

As with evaporated films, there are a number of problems associated with the deposition of alloys and compounds. The best theories of sputtering yields are typically inaccurate by a factor of two or so. Empirical determinations are normally required for films of controlled composition to be deposited. Nevertheless, sputtering of multicomponent or compound targets is frequently all that is necessary to obtain films varying only marginally from the target composition. *Maissel* [28.57] suggests that during the initial sputtering run an *altered region* is formed on the surface of the target, which is deficient in the target component having the highest sputtering yield. Thus although stoichiometric films will not normally be deposited during the initial sputtering run, such films may generally be deposited during subsequent runs, since the deficiency in the higher sputtering rate material compensates for its greater removal rate [28.51, 57]. Minor differences between the compositions of the film and target do still occur, but are normally the result of oxidation or of the evaporation of one component at high temperatures. When compounds are sputtered the incident ion energies are often sufficient to break the chemical bonds, again leading to a deficiency in one component. *Reactive sputtering* and *co-sputtering* are specifically designed to give greater flexibility during the deposition of multicomponent films.

Reactive sputtering involves the sputtering of a target using a sputtering gas which is amenable to chemical combination with the target material. The gas used may be either solely the reactive species, or a mixture of the reactive species and an inert sputtering gas, such as argon. Clearly the process is suitable for preparation of certain insulators which cannot be prepared by the previously mentioned sputtering techniques. Deposition rate depends primarily on the target sputtering yield, through the energy of the incident sputtering ions, the gas pressure and the distance between the target and the substrate. For example, *Perny* et al. [28.58] have concluded that the deposition rate is primarily determined by a reduced electric field E^*, given by E/p, where E is the mean electric field between the cathode and the anode, and p is the gas pressure. When a copper target was sputtered in a mixture of oxygen and argon, either Cu_2O, CuO or Cu were deposited depending on the value of E^* and the oxygen concentration [28.59]. Reactive sputtering has been used mainly to deposit oxide films which have insulating or semiconducting properties. Among the common oxides that have been studied are those of aluminium, cadmium, copper, iron, silicon, tantalum, tin and titanium. Oxides of niobium, thorium, vanadium and zirconium have also been investigated, as have the rare-earth oxides of hafnium, lanthanum and yttrium. The other major non-oxide compounds which have been investigated are the nitrides, with silicon nitride, Si_3N_4, receiving most attention due to its utility both as an insulator and in variants of the silicon planar process for integrated-circuit fabrication. This material has been prepared using a Si_3N_4 target and a mixture of argon and nitrogen as the sputtering gas [28.60]. Carbides and sulphides have also been sputtered reactively. In all cases it is imperative that the stoichiometry of the deposited films is determined using the normal chemical techniques, and the deposition parameters adjusted until the required composition is attained. A very useful review of reactive sputtering, covering both the mechanisms and the primary reactive sputtering techniques, has been given by *Westwood* [28.61].

Co-sputtering has certain advantages from the point of view of the preparation of films whose components

have different sputtering yields. Segregation of components on cooling can lead to a nonstoichiometric composition of material in the target, and differences in sticking coefficients between components can lead to the deposited composition differing from that arriving at the substrate. Co-sputtering has the advantage that the proportion of each sputtered components may be controlled during deposition, and may readily be adjusted in response to varying conditions to ensure stoichiometry of the deposited film. Co-sputtering may be performed with two or more conventional sputtering targets operating simultaneously for each component, each directed towards a common substrate. Another method of co-sputtering is again to use conventional targets, but with the substrate alternately subjected to the sputtered material from each target. This may be achieved by mounting the substrate on a disc which rotates in front of the targets. Alternatively a specially designed target assembly may be used such that the components arrive at the substrate in their stoichiometric proportions; the potentials of the components may be independently controlled to compensate for differing sputtering yields and sticking coefficients. Generally a plane-parallel substrate arrangement, with the target large in comparison to the target–substrate distance is employed, which maximises the uniformity of the film thickness. *Sinclair* and *Peters* [28.62] have described some target-electrode assemblies which have proven suitable for co-sputtering, including a concentric ring and disc assembly and an interdigitated design. Each of these allows independent control of the voltage applied to the two parts of the electrode assemblies and thus control over the composition of the sputtered films. Although the spatial dependence of the electric field in such assemblies is complex, and the deposition rate of one component may be influenced by the potential applied to the electrode for the other component, it is relatively straightforward to calibrate such a source for different materials, and to compensate for such effects by varying the acceleration potentials accordingly.

It is clear that, in general, impurities present in the gaseous phase may be incorporated into the deposited films. The fraction f_i of the impurity species i trapped within a film is given by [28.57]

$$f_i = \frac{s_i N_\mathrm{i}}{s_i N_\mathrm{i} + N_\mathrm{R}} \qquad (28.32)$$

where s_i and N_i ($\mathrm{m}^{-2}\,\mathrm{s}^{-1}$) are respectively the sticking coefficient and impingement rate of the impurity species and N_R is the molecular deposition rate of the required film material; f_i can be reduced by increasing N_R, or by decreasing either or both of s_i or N_i. In *bias sputtering* the phenomenon of re-sputtering is employed, i. e., components of the sputtered film are re-sputtered by ions from the plasma. This is achieved by applying an additional small negative potential to the substrate. Preferential sputtering of the impurities takes place, providing the binding energy of the impurities is lower than that of the host film material. This leads to an effective reduction in f_i, below that predicted by (28.32) [28.63]. The value of s_i in the expression is effectively reduced by re-sputtering of impurities from the film surface. *Asymmetric alternating current (AC) sputtering* [28.64] achieves similar objectives to bias sputtering, i. e., bombardment of the cathode target by ions to cause sputtering and bombardment of the films by lower-energy ions to preferentially re-sputter the impurities. These are accomplished by establishing an alternating current between the substrate and cathode and the use of a simple diode-resistance network. During the negative half-cycle the full negative potential is applied to the cathode, allowing sputtering to take place. In the following positive half-cycle, a relatively small positive potential (limited by the resistance) is applied to the cathode. This allows low-voltage sputtering of the films, which is sufficient to remove only the gaseous impurities. In this technique, again the value of s_i is effectively reduced, leading to a decrease in f_i. In *getter sputtering*, the gettering action of a freshly deposited thin film is utilised to purge the chamber of unwanted active gases. The concentration of impurities in thin films has been observed to decrease as the film thickness increases, and is attributed to the trapping (or gettering) action of the film material. In getter sputtering systems this action is exploited so that the majority of impurities are trapped in a film region deposited between the gas entry point and the substrates, while the required film is deposited elsewhere in the chamber where the gas impurity concentration will already have been depleted by the gettering action. The design of a typical getter sputtering system is described by *Theuerer* and *Hauser* [28.65]. In the case of getter sputtering, N_i is effectively reduced in (28.32), lowering the value of f_i.

Trapping of inert gas molecules in the films is also a possibility, and their concentration should be given by (28.32). However, at thermal energies the sticking coefficient is effectively zero, and thus trapping ought not to be of great concern. Nevertheless, for sputtering systems using argon, concentrations of 0.1% have been observed [28.66], generally for ion energies greater than 100 eV. Although not a major problem for many applications the possibility of inert-gas contamination should not therefore be discounted.

It has been mentioned earlier that simple direct current (DC) discharges as used in the diode and triode sputtering techniques described above, are unable to

deposit insulating films. Although reactive sputtering may be used, the deposition of films by this method requires careful control of sputtering rates and gas pressure, and often calibrations of individual systems and materials. The use of *radio-frequency (RF) sputtering*, allows insulators to be sputtered directly [28.67]. When a high-frequency (typically several MHz) alternating signal is applied to a metal plate, on which an insulating (dielectric) target is mounted and is placed within an auxiliary low-pressure discharge, the positive charge which accumulates on the target is neutralised during half of each cycle. Owing to the much higher mobility of the electrons than the ions in the plasma, the plasma potential becomes positive with respect to the target. RF sputtering was first employed by *Davidse* and *Maissel* [28.68], who found that a self-sustained discharge could be started and maintained by the application of RF power to the electrode only. The mechanism by which sputtering takes place requires the frequency to be in the radio-frequency range [28.16]. The electron mobility is several orders of magnitude greater than that of the ions which have difficulty in keeping up with the field variations, and at high frequencies the ions accumulate in front of the target, in the same way that they do in a DC discharge. Because the number of electrons arriving at the target during the positive half-cycle are not matched by the number of ions arriving during the negative half cycle, the front of the target acquires a negative charge, which repels most of the electrons that would arrive at the surface during the positive half-cycle. The plasma itself is said to be positively *self-biassed* with respect to the target. It is therefore possible to both feed energy into the plasma, while also giving the ions sufficient energy to cause sputtering. The RF discharge is self-sustained, and Davidse and Maissel found that an externally generated plasma is not required. A schematic diagram of the RF sputtering equipment is shown in Fig. 28.6 [28.69] while more detailed discussion of the RF sputtering process is given in previous reviews [28.16, 57]. These authors point out that a major difference between diode and RF sputtering is a requirement for an impedance matching network or *match box* [28.70] in the RF case. The matching network is inserted between the power supply and the discharge chamber, its purpose being to maximise the power that can be delivered to the plasma. Typically a network containing an inductor and a capacitor is used. An example is shown in the literature [28.71]. A blocking capacitor is normally inserted in the network to ensure that the build up of the self-bias voltage proceeds satisfactorily. Although the target itself will normally perform this function in the case of insulator sputtering, the inclusion of a blocking capacitor allows RF sputtering to be used for the deposition of conductors as well as insulators, so semiconductors and metals are also easily deposited by this technique. Mainly for this reason RF sputtering has become a versatile technique which is used near-universally for semiconductor production as well as for insulators and metals [28.70]. A significant increase in the sputtering rate is also observed in RF systems.

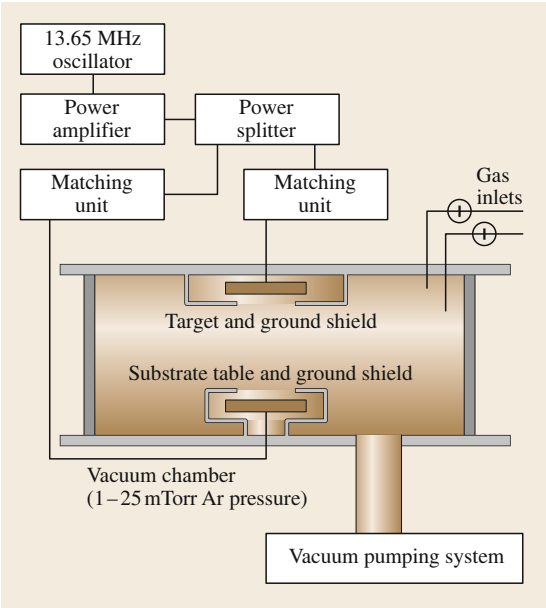


Fig. 28.6 A typical radio-frequency (RF) sputtering system. (After [28.69] with permission from Oxford Univ. Press)

Finally, a technique used to increase the sputtering rate still further is that of *magnetron sputtering*. In this technique, a magnetic field is applied between the target and the substrate. Electrons in the plasma then experience the Lorentz force from the magnetic field in addition to that derived from the electric field, so that the total force on the electron is given by

$$\boldsymbol{F} = m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = e\boldsymbol{E} + e\boldsymbol{v}\times\boldsymbol{B}\,, \tag{28.33}$$

where $\boldsymbol{F}$ is the total force experienced by the electron, m and e are the electron mass and charge respectively, $\boldsymbol{v}$ is the electron velocity, $\boldsymbol{E}$ is the electric field and $\boldsymbol{B}$ is the magnetic field [28.19, 72]. The effect of the two fields is to cause the electrons to move in spiral rather than linear paths, and therefore to greatly increase the distance travelled by the electrons in the plasma. This substantially increases the time the electrons are within the plasma, and thus the probability of their causing secondary ionisation is also significantly enhanced. In turn this leads to increased current flow and crucially increased deposition rates. The magnetron sputtering

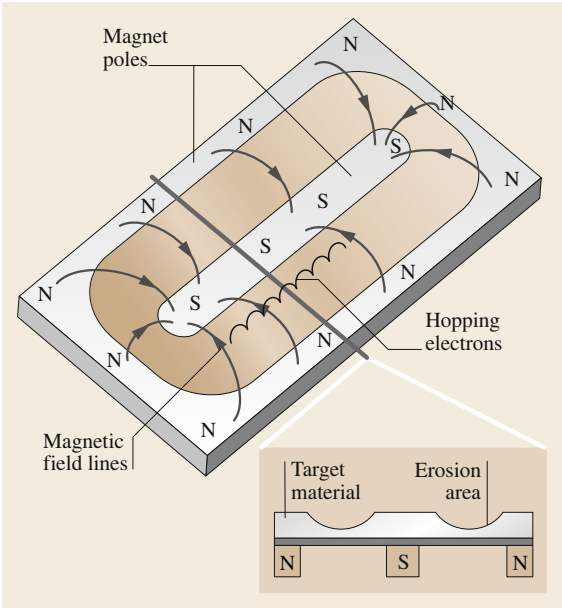


Fig. 28.7 The technique of magnetron sputtering, showing the hopping electrons of the target confined to a *racetrack* region by the magnetic fields from the magnets under the target. (After [28.69] with permission from Oxford Univ. Press)

source is constructed by mounting a powerful magnet below the target, with one pole concentric with the other as shown in Fig. 28.7 [28.69]. This establishes a magnetic field across the target surface; the direction of the magnetic field is perpendicular to the electric field at the furthest extent of the magnetic force lines from the target. The result of the two interacting fields is to confine the electrons for a significant period within a region termed the *racetrack* at the surface of the target, where the electrons effectively hop across the surface. Ionisation of the sputtering gas occurs mainly above the racetrack, and erosion of the target takes place preferentially in this region. In addition to the enhanced deposition rate, magnetron sputtering also has several other advantages. Since the ionisation region is effectively confined to the racetrack, unwanted sputtering from the substrate and the walls of the chamber are reduced, as is the substrate temperature. The gas pressure required to sustain the discharge is also reduced, owing to the greater ionisation efficiency.

A further improvement in the technique of magnetron sputtering may be obtained by the use of an unbalanced magnetron, as described by *Window* and *Savvides* [28.73]. In spite of its excellent high deposition rate, magnetron sputtering generally produces films which have a columnar structure. This leads to the presence of voids, both within grains and at grain boundaries. Low-energy ion bombardment has been shown to improve adatom mobility, allowing atoms to migrate on the depositing surface and filling some of the voids [28.72]. Various computer simulations described in this reference have illustrated the importance of ion beams in depositing dense, void-free films. It is possible to generate an ion current in the plasma by using an unbalanced magnetron where the inner and outer magnets are deliberately unbalanced from the ideal magnetron configuration. Further details concerning the field characteristics of unbalanced magnetrons and of the unbalanced magnetron technique are given in the literature [28.72, 73]. The method has been used to successfully deposit various types of film, including those intended for electrical and optical applications, corrosion protection, and wear and abrasion resistance. A fuller discussion of materials prepared using this method is given by *Rohde* [28.72].

Magnetron sputtering may also be used in conjunction with other sputtering techniques. A combination of RF and magnetron sputtering is frequently used, for example for Co:M alloys where M represents Al, Si, Ti, Cr, Fe, Ni, Cu, Zr, Nb, Mo, Ag, Ta or W [28.74], and also for the deposition of silicon nitride films [28.60].

Thus sputtering has developed from a relatively simple DC technique capable only of the deposition of conductors at a relatively low rate, to encompass both conductors and insulators (RF sputtering) and with sputtering rates suitable for high-throughput industrial processes (magnetron sputtering). Improvements in film morphology have been accomplished by the use of unbalanced magnetron sputtering.

28.1.2 Chemical Deposition Methods

There are many different chemical methods for preparing thin films, but by their nature many of these are restricted to one or perhaps a few different materials, related to a specific chemical reaction or series of reactions. It is therefore only possible to categorise some general techniques and to give examples for various particular materials. *Campbell* [28.75] divided chemical methods into two classes: those depending on an electrical source of ions, such as electroplating and anodisation, and those requiring a chemical reaction, usually, but not exclusively involving a chemical vapour as in the CVD method. A further technique, which has found favour in recent years, is the deposition of molecular films using the Langmuir–Blodgett technique. Brief examples of these are given below.

Electrodeposition

By electrodeposition, we include all methods which involve passing an electric current through a solution such

that a film of a material, usually a metal or an oxide, is deposited on one of the electrodes immersed in the solution. Electrolytic deposition, or *electroplating*, is the deposition of a metal by this method. According to *Campbell* [28.75, 76], of the 70 metallic elements, 33 can be electroplated successfully, although of these only 14 are deposited commercially. The metallic film is deposited on the cathode, and its thickness and the time in which this occurs are determined by the two laws of electrolysis:

1. The mass of the deposit is proportional to the quantity of electricity passed.
2. The mass of material deposited by the same quantity of electricity is proportional to the electrochemical equivalent, E.

This can be expressed as

$$\frac{W}{A} = JtE\alpha \,, \tag{28.34}$$

where W is the mass deposited on a cathode of surface area A, J is the current density, t is the time of deposition and α is the *current efficiency*, or ratio of the expected to the theoretical mass deposited. Its value is generally in the range 0.5−1.0. The previous expression can be rewritten in terms of the film deposition rate, R (or thickness d per unit time t), by noting that the film density $\rho = W/(Ad)$. Combining this with (28.34) gives

$$R = \frac{d}{t} = \frac{JE\alpha}{\rho} \,. \tag{28.35}$$

When a high current is passed through a suitable electrolyte, metallic ions move towards the cathode under the influence of the applied electric field. The mass deposited, and the deposition rate are given by (28.34) and (28.35), and may easily be estimated if E and ρ are known. These are tabulated for the 14 metals commonly deposited by this method [28.75]. Deposition rates can be very high, depending on the current drawn through the electrolyte. For example, silver will deposit at a rate of about 1 nm/s for a current density of 10 A/m^2, rising to 1 μm/s at 10 000 A/m^2 [28.75]. Further details of the deposition of alloys are given in a standard text [28.77], while the growth and structures of electrodeposits are described by *Lawless* [28.78].

A second technique which is used for the deposition of oxides (or less commonly for other compounds such as nitrides), is that of *anodisation*. As its name implies, films are deposited on the anode of the parent metal, following a chemical reaction. The chemical equations describing the anodisation process are given below [28.75]

$$\begin{aligned} &\mathrm{M} + n\mathrm{H_2O} \rightarrow \mathrm{MO}_n + 2n\mathrm{H}^+ + 2n\mathrm{e}^- \text{ (at the anode)} \,, \\ &2n\mathrm{e}^- + 2n\mathrm{H_2O} \rightarrow n\mathrm{H_2}\uparrow + 2n\mathrm{OH}^- \text{ (at the cathode)} \,, \end{aligned} \tag{28.36}$$

where M represents the metal and n is an integer specifying the oxide deposited. The oxide grows in an amorphous form at the metallic anode surface, while hydrogen is evolved at the cathode. The electrolyte is typically an aqueous solution of water containing a dilute acid.

In general a constant current is passed through the cell and film thickness is proportional to the time elapsed. Films grown under constant-voltage conditions, show a decreasing deposition rate with time, as the current falls to zero when the voltage dropped across the film increases to the total applied voltage. Thus the maximum thickness of the film is determined by the voltage used. Growth rates for anodic films under constant-current conditions are of the same order as those for the PVD methods of evaporation and sputtering. Aluminum and tantalum anodisation at a current density of 20 A/m^2 gives growth rates of the order of 1 nm/s [28.76]. Considerably more detailed discussions of anodic oxide films in particular are given in the literature [28.79, 80]. These cover considerably more details of the types of films that may be successfully prepared, their structures, characterisation methods and copious references to related work.

Chemical Vapour Deposition (CVD)

Chemical vapour deposition is a method whereby a volatile compound reacts, with other gas species, to produce the required compound on the substrate. CVD does not require operation under vacuum, as with the PVD processes of evaporation, MBE and sputtering. Moreover, it is capable of producing both silicon and gallium arsenide (GaAs) semiconductor films, and to grow these on substrates of the same material as high-quality epitaxial layers (homoepitaxy). For these reasons it has found wide acceptance in semiconductor processing. High deposition rates relative to PVD deposition methods are possible, and controlled doping performed during the deposition process is relatively straightforward. Objects having quite complex shapes can be coated relatively easily, since the film deposition does not depend on line-of-sight geometry as in the case of evaporation and to a lesser extent sputtering. There are however some formidable disadvantages with the technique. Generally much higher substrate temperatures are required in CVD than for the other methods. The reactive gases used in the deposition

process and/or the reaction products can be extremely toxic, explosive or corrosive, requiring considerable investment in safety features. This may well be justified in high-throughput semiconductor production processes, but uneconomic for smaller-scale development work, where PVD methods are preferable. Corrosion and effects of unwanted diffusion, alloying and chemical reactions at the substrate surface may occur under the high operating temperatures, and masking of substrates is particularly difficult [28.81].

In principle CVD may utilise any chemical reaction which produces the required material, and therefore the equipment used for the production of a given material is likely to differ from that required for another material. Nevertheless, there are several different classes of reaction which are used in CVD processes which are described in the literature [28.16, 18, 19, 75, 76]. Some of these are outlined below.

Pyrolysis or thermal decomposition of a gaseous compound entails passing a vapour over a heated substrate, which causes decomposition and the condensation of a stable solid. This has been used primarily in the production of silicon from SiH_4 and nickel from nickel carbonyl. *Oxidation* is frequently employed to provide SiO_2 in silicon processing. Halides of the required metal oxide are reacted with steam to give the oxide and a hydrogen halide. In *nitriding* the steam is replaced with ammonia (NH_3), and nitrides such as Si_3N_4 may be generated. *Reduction* reactions use hydrogen gas instead of steam or ammonia in the preparation of silicon and refractive metals such as tantalum or molybdenum from their halides. In *disproportionation*, a reversible reaction of the type

$$A + AB_2 \rightleftharpoons 2AB \qquad (28.37)$$

is usually employed [28.76], where A represents a metal, B is usually a halide and both AB and AB_2 are gases. The higher valence state is more stable at lower temperatures so, if the hot gas AB is passed into a colder region, AB_2 will be formed and A deposited. In *transfer reactions* for the preparation of compounds, two different strategies may be employed. Either a compound can be used as the source material, or the source comprises a volatile compound of one of the elements of the required compound. Gallium arsenide (GaAs) can be prepared by both methods [28.16, 83]. In Fig. 28.8 [28.82] a typical CVD system is shown which uses a transfer reactor of the second type to produce GaAs. Highly purified arsenous chloride ($AsCl_3$) and Ga are the source materials. Purified hydrogen is bubbled through the $AsCl_3$ and passed through the furnace. In zone 1 of the furnace, the following reaction takes place

$$2AsCl_3 + 3H_2 \rightarrow 6HCl + \frac{1}{2}As_4 \,. \qquad (28.38)$$

The arsenic vapour is completely absorbed by the Ga source in zone 2 until saturation occurs at 2.25 at. % As. No free arsenic condenses beyond the furnace, but gallium is transported as a lower chloride. Deposition of GaAs occurs in zone 3, when hydrogen is diverted to bypass the bubbler.

Tables of materials which can be deposited by CVD and the associated reactions are given in the literature [28.19, 75]. However, several modifications to the basic CVD process are also in current use. These seek to overcome some of the deficiencies associated with CVD, such as the operation at atmospheric pressure and the use of very high temperatures often involved. Systems employing organic precursors also have certain advantages for the deposition of various materials.

We first discuss low-pressure CVD or LPCVD. Typical pressures of the order of 100 Pa are used, approximately a thousandfold reduction from atmospheric pressure. Contamination is normally lower than for atmospheric CVD and film stoichiometry is often improved. To compensate for the reduced pressure, the

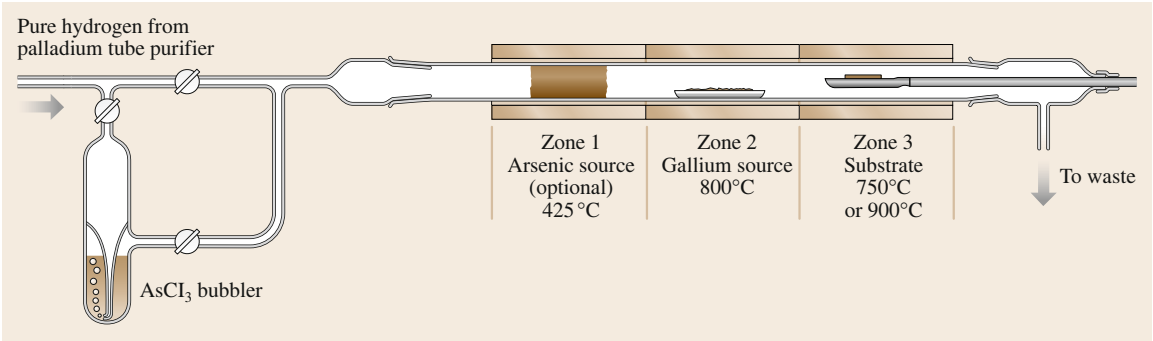


Fig. 28.8 Epitaxial growth apparatus for the production of GaAs using chemical vapour deposition (CVD). (After [28.82] with permission from Elsevier Science)

input reactant gas concentrations are increased, the gas flow velocity is increased, but the gas density is reduced. Because the mean free path of the gas molecules increases due to the reduced pressure, substrates can be placed closer together, giving a higher throughput in commercial applications. Silicon and dielectric films are generally prepared by this method. The operating temperatures required for LPCVD remain high, however.

In plasma-enhanced CVD (PECVD) a glow discharge is initiated inside the chamber where the reaction occurs. A comprehensive review of the PECVD technique is given by *Ojha* [28.84]. Dissociation of the gas in the plasma takes place, owing to the nonequilibrium nature of the plasma, which has a high electron temperature but a low gas temperature. The main advantage of this technique is that the plasma decomposition allows film deposition to take place at much lower temperatures than those employed in atmospheric CVD using decomposition reactions and also in LPCVD. In PECVD, as in LPCVD, the pressure is reduced considerably below atmospheric pressure, although the range of pressures employed varies over a wider range, and can be as high as a few hundred Pa. The discharge is usually excited by an RF field, although both DC and microwave frequency fields have also been used. Average electron energies are of the order of several eV, which is sufficient to cause ionisation and decomposition of the gas molecules. A major application of PECVD is in the deposition of silicon nitride films, an important dielectric used in semiconductor processing for microelectronic passivation. Here the temperatures may not exceed about 300 °C, otherwise damage to the circuitry can result, and so atmospheric CVD and indeed LPCVD are not suitable techniques. For details of the chemical reactions involved, and a comparison of the properties of silicon nitride prepared by LPCVD and by PECVD the reader is referred to the literature [28.19, 85]. The former reference also includes a table giving details for the deposition of many other thin-film materials, including silicon, germanium and other semiconductors, as well as various oxides, nitrides and carbides. PECVD systems operating at microwave frequencies utilise the phenomenon of electron cyclotron resonance (ECR); this deposition method is known as electron cyclotron PECVD (or ECRPECVD). Frequencies are of the order of 1−10 GHz and the ECR plasma may be generated at pressures of about 10^{-3}−10^{-1} Pa, leading to a degree of ionisation up to about 1000 times that achieved in an RF plasma [28.19]. High deposition rates, low-pressure operation and absence of contamination have made ECRPECVD a very attractive technique for film deposition. A full review of this technique has been given by *Popov* [28.86].

Finally we briefly mention the technique of metalorganic CVD (MOCVD), which is becoming an increasingly important method of film deposition. The technique is essentially the same as CVD, except that at least one of the precursors is a volatile metalorganic compound, such as a metal alkyl. The formation of the required compound takes place by pyrolysis of the metalorganic compound with another (inorganic or organic) gaseous precursor. The main advantage of MOCVD is the relatively high volatility of the organic compounds at moderate temperatures, and the fact that, since all these are in the vapour phase, precise control of gas flow rates and partial pressures are possible [28.19]. The compound semiconductors such as GaAs, GaP and InP may be prepared using, for example, trimethyl-Ga ($(CH_3)_3Ga$) or trimethyl-In ($(CH_3)_3In$) as precursors. The basic reaction is with a Group V hydride, e.g., AsH_3 or PH_3 for the compounds listed above. A short table of some film materials which have been successfully generated by MOCVD is given in the literature [28.19]. In addition to its applications in preparing compound semiconductor films MOCVD has also been applied to the deposition of high-temperature superconducting thin films, with quite complex structures [28.87].

Langmuir–Blodgett Film Deposition

The Langmuir–Blodgett technique of film deposition entails the coating of solid substrates with molecular layers. *Langmuir* [28.88] first described the transfer of fatty-acid organic monolayers from the surface of water to a rigid substrate, while *Blodgett* [28.89] extended the work to show that the process could be repeated so that monolayers were sequentially transferred to form multilayer films. The early history of Langmuir–Blodgett films has been briefly described in the preface to the Proceedings of the First International Conference on Langmuir–Blodgett Films [28.90]. In the basic deposition process, a fatty-acid film is spread over the surface of a subphase (usually purified water). Most molecular-film materials have a molecular structure where one end of the molecule is hydrophilic, which seeks out the water subphase, while the other end of the molecule is hydrophobic and attempts to point away from the water subphase. We therefore have the situation where the hydrophilic end of the molecule is oriented downwards towards the water, with the hydrophobic end oriented upwards. In Langmuir–Blodgett deposition, the required molecules are dispersed in a volatile solution, which eventually evaporates, leaving the molecules with their hydrophilic ends immersed in the water, but with their axes randomly oriented [28.91]. A compressive surface-barrier system is employed to compress the molecules so that they are close-packed and oriented perpendicu-

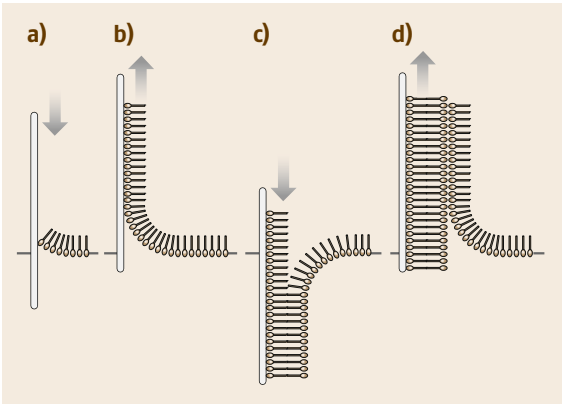


Fig. 28.9a–d Langmuir film deposition: **(a)** immersion of a hydrophilic substrate without film deposition; **(b)** withdrawal of the hydrophilic substrate; **(c)** repeated immersion with film deposition; **(d)** continuation of the process. (After [28.91] with permission from Elsevier Science)

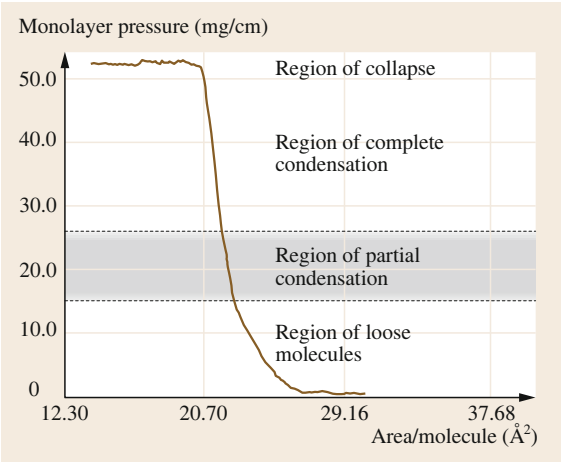


Fig. 28.10 The pressure–area relationship for a cadmium stearate molecular layer obtained at a temperature of 22 °C at pH 5.6. Film deposition should be carried out in the linear region of the curve, i. e., the region of complete condensation without collapse. (After [28.91] with permission from Elsevier Science)

lar to the surface of the subphase, with their hydrophilic ends pointing vertically downwards. The molecules can then be transferred to either a hydrophilic or hydrophobic substrate by slowly dipping into the subphase. The basis process is shown in Fig. 28.9 [28.91], for the case of a hydrophilic substrate. On the initial introduction of the substrate into the subphase no coating takes place (a), but on removing it a film attaches to the substrate by polar-bond formation (b). The substrate is then effective hydrophobic, and on dipping for a second time, the hydrophobic ends of the molecules on the subphase form van der Waals bonds with the hydrophobic substrate surface, thus again making the surface hydrophilic (c). On withdrawing the substrate for a second time a further molecular layer is deposited, with the hydrophobic ends of the molecules at the surface (d). Repeated dipping cycles then establish multilayer films with an odd number of layers. The use of an initially hydrophobic substrate gives an even number of molecular layers for each completed dipping cycle.

It is evident that we have a mechanism whereby a predicted number of molecular layers may be deposited, and knowing the molecule dimensions (i. e., the length of the molecule between its hydrophilic and hydrophobic ends) together with the number of dipping cycles, it is easy to accurately calculate the film thickness. It is important, however, that the surface pressure is maintained within a fixed range of values during deposition. If it is too high the surface film collapses in localised regions, giving a transferred layer thickness greater than one monolayer; conversely, if the pressure is too low, there may be voids in the film surface, which transfer on to the substrate. This is illustrated in Fig. 28.10 [28.91], which shows a pressure–area isotherm for a cadmium stearate molecular layer. It is normal to use a feedback system during deposition to control a moveable barrier on the film surface to maintain the surface pressure within a narrow range while the dipping process takes place. Here the molecular area remains effectively constant for a large change in pressure (region of complete condensation in the figure), ensuring that small changes in pressure do not cause the deposited film to fold over or break up.

There are, however, some limitations to the use of the Langmuir–Blodgett deposition method. The material must satisfy the following requirements [28.91]. The molecular chain structure needs to be hydrophilic at one end and hydrophobic at the other. Furthermore, it should be insoluble in water, but soluble in a suitable solvent in order to give a spreading solution. Some materials which are insoluble can, however, be modified by chemical substitution of molecular end groups. For example, metal-free phthalocyanine is normally insoluble, but the chemical addition of $C(CH_3)_3$ end groups to the molecule renders it soluble in some aromatic solutions such as toluene. This type of chemical modification allows the deposition of materials which cannot normally be deposited to be deposited in a substituted form [28.92]. Many other substituted organic compounds may also be deposited, including dyes, porphyrins, fullerenes, charge-transfer complexes and polymers [28.93]. It is also possible to prepare Langmuir–Blodgett film layers with alternating monolayer structures, by dividing the Langmuir–Blodgett trough into two compartments, one for each type of

monolayer, separated by a surface barrier [28.94, 95]. Deposition then takes place by alternately removing the substrate from the surfaces of the two compartments, by passing the substrate through the subphase under the surface barrier between deposition of each of the materials.

The advantages of the Langmuir–Blodgett technique over other thin-film deposition methods lie primarily in the low temperature at which deposition takes place (close to room temperature), and on the feasibility of depositing very thin mono-moleculer layers of known thickness. Several types of electronic devices incorporating Langmuir–Blodgett films have been discussed [28.96] as has the writing of very fine lines of width less than 10 nm separated by 20 nm in manganese stearate films [28.97]. Further work on the deposition of other materials by this method may well lead to novel device structures and to improved processing methods. An excellent discussion on the preparation and major properties of Langmuir–Blodgett films is given by *Petty* [28.93].

28.1.3 Sol–Gel Method

The sol–gel process is regarded as being a cost-effective and simple technique to synthesize primarily oxide ceramics. Among several significant advantages, products with a high degree of purity can be obtained by purifying precursors via either distillation, crystallization, or electrolysis [28.98]. In this process, the solution (sol) undergoes a sol–gel transition in which the gel becomes a relatively rigid, porous mass. This involves the use of metal alkoxides as precursors for the preparation of a colloidal solution [28.99]. Metal alkoxides have organic ligands attached to metal atoms and react readily with water. The most common precursor is silicon tetraethylorthosilicate (TEOS). In a two-phase system solid silica SiO_2 and solvent-filled pores are formed from the polymerization of $Si(OC_2H_5)_4$. Hydrolysis of TEOS takes place in the presence of water mixed with ethanol (C_2H_5OH) solvent according to the following chemical reaction

$$\mathrm{Si(OC_2H_5)_4 + xH_2O \xrightarrow{C_2H_5OH} Si(OC_2H_5)_{4-x}(OH)_x + xC_2H_5OH\,.} \qquad (28.39)$$

It is to be noted that no additional catalyst is required for the hydrolysis of TEOS in C_2H_5OH.

Depending upon amounts of water and catalyst present, complete hydrolysis may take place with conversion from $Si(OC_2H_5)_4$ to $Si(OH)_4$ producing silicic acid.

For partial hydrolysis, a siloxane Si–O–Si group is formed through possible condensation of the two silanol groups with elimination of water and ethanol

$$\mathrm{Si(OC_2H_5)_{4-x}(OH)_x + Si(OC_2H_5)_{4-x}(OH)_x \rightarrow Si(OC_2H_5)_{4-x}(OH)_{x-1}OSi(OH)_{x-1}(OC_2H_5)_{4-x} + H_2O} \qquad (28.40)$$

Intermediary Si–OH silanol groups may be formed as a result of partial hydrolysis [28.100].

Equations (28.39) and (28.40) describe the chemical reactions during the sol–gel process

The condensation process is dynamic, and may be manoeuvered in the desired direction by adjusting parameters like types of precursor, catalyst and solvent, reactant concentrations, and ratio of water to alkoxide. When the ratio of water to alkoxide = 1, hydrolysis becomes complete without condensation. A ratio greater than unity increases condensation. A precise control over the stoichiometry of the deposited film is achieved by exploiting polycondensation reactions of molecular precursor metal alkoxides in a solution over a very large pH range. Slow hydrolysis produces cross-linked linear molecules in acidic, low pH conditions. These molecules entangle and form additional branches resulting in gelation. The minimal reaction rate for hydrolysis is at pH 7, and for condensation at around pH 4.5. However, basic conditions at high pH levels promote the formation of more highly branched silica clusters as discrete species. Gelation occurs by linking these clusters. As the sol aggregates the viscosity will increase until a gel is formed. The pH where the particle is neutral is called the point of zero charge (PZC). The size of the surface potential depends on the difference between pH and PZC. The sol–gel transition (gel point) is reached when a continuous network is formed. The gel time is determined as the time at which it is possible to turn the container upside down [28.101]. All fluid is kept in the gel, and the volume is maintained. If the solvent from the pores is removed by drying, capillary forces will result in shrinkage and the gel network will collapse, forming a xerogel. The network structure may be retained with supercritical evaporation which yields low density ($0.085\,g/cm^3$) aerogels with a specific surface area of $700\,m^2/g$ and a large pore volume ($3.29\,cm^3/g$) [28.102].

Sol–gel processing allows materials to have any oxide composition. Uniform thin films of a large number of metal oxides can be deposited on a variety of substrates at room temperature in a hydro-alcoholic medium in the presence of a basic or acidic catalyst. Thin films of titania TiO_2 produced using sol–gel methods have been extensively investigated for various device applications due to the low fabrication cost. Sol–gel coatings may result from dipping, spin coating, or spray-

ing a sol precursor onto a substrate. The formation of a thin film through spin coating consists of four stages:

1. Deposition, which is the dispersion of the solution on the surface of the substrate
2. Spin-up, in which the solution flows toward the edges of the substrate
3. Spin-off, where the excess amount of solution is spread as droplets and leaves the surface of the substrate, and lastly
4. The evaporation of the solvent.

The alkoxide used during the current work is titanium isopropoxide ($C_{12}H_{28}O_4Ti$) with a purity of 99.99%, which is mixed with absolute ethanol and acetic acid as a catalyst in the molar ratios of 1 : 9 : 0.1. Thin films of TiO_2 were deposited by the method of spin coating using a standard photoresist spinner. In general the films were predominantly polycrystalline with an estimated grain size of a few nanometers. The grain size decreases with an increase in spin speed. The room temperature conductivity is an order of magnitude higher than that of TiO_2 films prepared from a similar solution by dip coating [28.103].

One promising transparent oxide that can be used as an insulating, semiconducting or metallic layer, is the zeolite-like calcium aluminate $12CaO \cdot 7Al_2O_3$ (C12A7) binary compound. This novel oxide exhibits physically interesting characteristics by exploiting its high-density intrinsic defect nanostructure. The pulsed laser deposition method has been extensively used for preparation of C12A7 thin films. The technique is relatively costly and requires a C12A7 target as the starting material for the deposition of the film. For the sol–gel process, C12A7 sol is synthesized through hydrolysis of two separate solutions. Aluminum oxide hydroxide (AlOOH) sol is prepared using aluminum sec-butoxide $Al(OBu)_3$ as the precursor. In order to avoid precipitation this precursor is modified using ethyl acetoacetate (etac). $Al(OBu)_3$ is hydrolytically unstable and its hydrolysis leads to immediate precipitation. However, modifying $Al(OBu)_3$ using ethyl acetoacetate (etac) mediates the hydrolysis of Al^{3+} ions, through substitution of monodentate butoxy ligands of $Al(OBu)_3$ with etac groups. Etac groups will be bonded to Al atoms in an enolic form (Fig. 28.11).

$Al(OBu)_3$ is mixed with etac at a molar ratio of $Al(OBu)_2$: etac = 1 : 1. The mixture is magnetically stirred for 30 min followed by 1 h of ultrasonic mixing. Ultrasonic mixing provides better dispersion compared to our previous work. Through this stage the $Al(OBu)_2$(etac) complex is formed. This new precursor is diluted with ethanol ($Al(OBu)_2$: ethanol = 1 : 4) prior to hydrolysis. In the final step ethanol is diluted with distilled water followed by pH regulation using hydrogen chloride (HCl) acid ($Al(OBu)_3$: H_2O : HCl = 1 : 3 : 0.075). Diluted distilled water is then added to the $Al(OBu)_2$(etac) solution initiating the hydrolysis process. The hydrolysis is performed under vigorous stirring using an ultrasonic bath, followed by gentle stirring at room temperature for 24 h using the magnetic stirrer. During hydrolysis of the modified $Al(OBu)_2$(etac) compound, etac groups will be separated from the compound in both enolic and ketonic forms. Following the separation of butoxy ligands, AlO bonding will be exposed to OH bonding of water, resulting in the formation of boehmite (γ-AlO(OH)). The reaction to this water content may be derived as

$$\begin{aligned} &Al(OC_4H_9)_2(\text{etac}) + 3H_2O \rightarrow \\ &AlOOH(H_2O) + \text{etac} + 2C_4H_9OH\,. \end{aligned} \qquad (28.41)$$

The molar ratio $Al(OBu)_3$: HCl = 1 : 0.15 was chosen for acid concentration used in the synthesis of AlOOH particles, thereby reducing the pH of the solution from 10.4 to 8, well below the isoelectric point of boehmite particles. Such acid regulation will reduce the chance of agglomeration due to columbic interactions between particles. The final product was AlOOH particles dissolved in ethanol.

A second solution was prepared using $Ca(NO_3)_2 \cdot 4H_2O$ dissolved in ethanol. Additional water was added to the solution ($Ca(NO_3)_2 \cdot 4H_2O$: H_2O = 3 : 4) to initiate partial hydrolysis. Addition of water expels some salt ions to form HNO_3, decreasing the pH of the solution to 4. The dispersion of calcium nitrate in the solvent was facilitated using an ultrasonic bath followed by 24 h of magnetic stirring. The solution formed required no further acid addition since HNO_3 was formed inside the solution.

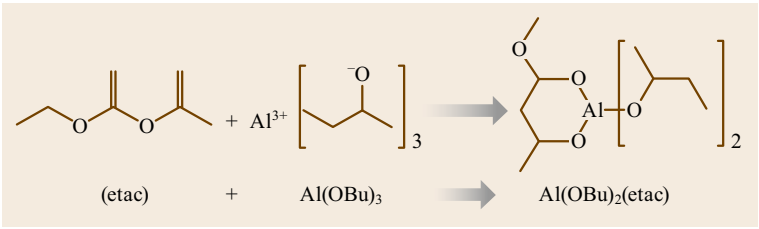


Fig. 28.11 Enolic bonding of etac groups to Al atoms

The sol–gel process starts with a homogenous solution of AlOOH and $Ca(NO_3)_2$ in ethanol solvent with the concentrations meeting the nominal composition of C12A7 compound. The AlOOH solution and $Ca(NO_3)_2$ were mixed allowing the formation of homogenous C12A7 amorphous solution. C12A7 nanoparticles were prepared by dispensing a thin layer of sol in a Petri dish followed by a 72 h drying process at room temperature. During drying, evaporation of ethanol progressively facilitates the formation of the amorphous phase from the solution [28.104].

Sol–gel processing has been widely employed for the synthesis of inorganic–organic hybrid materials.

The high-temperature synthesis route to ceramic materials is not suitable for the incorporation of thermally labile organic moieties. Two different routes are used for incorporation of organic groups: (1) embedding organic compounds into gel without chemical bonding and (2) covalent bonding of organic compound to gels [28.105]. The sol–gel process is commonly identified as an ecofriendly process compared to traditional routes to ceramics and glasses because the first steps are usually carried out at low temperature. Unlike the high-temperature classical methods, the possibility of chemical interaction between the container and the materials is very small. As the processing takes place in a liquid medium, there is no pollution due to dispersion of dust. Industrial and consumer efforts have been made to recycle various forms of silica gel [28.106].

28.2 Structure

In this section some examples of the structures of various types of thin films are described. By the *structure*, we mean mainly the crystallographic form in which the film is deposited, and also factors associated with this, such as the mean size of crystallites and their orientations. However, the gross morphology of films is also included, and this covers features generally larger than the grain size. The basic type of film, i. e., epitaxial, polycrystalline or amorphous, is usually deduced by diffraction methods (X-ray, electron or neutron), although we will mainly be concerned with the first of these only. Thus a brief description of the method of X-ray crystallography is first given, followed by examples of different types of films that have been observed. Finally some examples of the larger-scale morphology are described.

Several reviews of film structure have been given previously [28.107–109], as well as more generalised discussions in most of the texts covering thin films, to which the reader is referred. However, it should be noted that the large number of variables encountered in thin-film deposition processes (e.g., substrate temperature, deposition rate, angle between vapour stream and substrate) permit films of differing structures to be deposited under nominally similar conditions. It is always wise, therefore, before interpreting measurements of any thin-film properties (such as the electrical or optical properties) to first fully determine the structure of the films, and to take this into account in any interpretation.

28.2.1 Crystallography

In general films tend to deposit from the vapour phase in one or more of their documented bulk-crystal structures. In crystals, the spatial arrangement of the atoms exhibits a symmetry, which can be replicated throughout three-dimensional space by repetition of a basic unit, called the primitive unit cell. The unit cell of any crystal structure may be specified by the magnitudes of three vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ representing the unit cell lengths, and the angles α, β and γ between the three vectors. Symmetry considerations allow all crystals to be classified into seven crystallographic structures, namely the cubic, tetragonal, orthorhombic, hexagonal, trigonal, monoclinic and triclinic forms, of which only the first three are orthogonal. Conditions on the equality or otherwise of the lengths a, b and c and the angles between them determine into which type the structure of a given material falls. In three dimensions the atoms arrange themselves in planes, with gaps between the planes known as the planar spacing, d_{hkl}. In any perfect crystal there are many sets of crystal planes that may be identified. The possible planar spacings are given by different expressions, depending on the crystal structures. For example, in the cubic structure ($a = b = c$, $\alpha = \beta = \gamma = 90°$)

$$\frac{1}{d_{hkl}^2} = \frac{h^2 + k^2 + l^2}{a^2} \,. \tag{28.42}$$

In the tetragonal structure ($a = b \neq c$, $\alpha = \beta = \gamma = 90°$)

$$\frac{1}{d_{hkl}^2} = \frac{h^2 + k^2}{a^2} + \frac{l^2}{c^2} \tag{28.43}$$

and in the hexagonal structure ($a = b \neq c$, $c \perp a$, $\gamma = 120°$)

$$\frac{1}{d_{hkl}^2} = \frac{4}{3}\left(\frac{h^2 + hk + k^2}{a^2}\right) + \frac{l^2}{c^2} \,. \tag{28.44}$$

In these equations a, b and c represent the lengths of the sides of the primitive unit cell, and h, k and l are integers. Similar, but more complex, equations exist for the other crystal structures which have less symmetry. For bulk crystals having unit cell sides of the order of nanometres, the crystal planes act as three-dimensional diffraction gratings, which reflect an incident beam of radiation according to Bragg's law,

$$\lambda = 2d_{hkl} \sin\theta \,, \tag{28.45}$$

where λ is the wavelength of the radiation, and θ is the angle between the incident beam and the reflecting planes. The values of λ for X-rays, electrons and neutrons may be chosen such that they are of the same order as d_{hkl}, and therefore diffraction will occur at suitable angles. These angles are those which satisfy Bragg's law, and are therefore determined by the radiation wavelength λ and by the integer values (h, k and l) applying to the sets of reflecting planes present. (hkl) then specifies the set of reflecting planes, and also the orientation.

For perfect crystals, reflections only occur at certain well-defined angles. If the crystal is rotated with respect to the incident beam of radiation, the deviation angle between the incident and reflected beam is 2θ. If the material, instead of being perfect, is made up of a large number of differently oriented microcrystallites, i. e., it is polycrystalline, the diffraction method will pick up reflections from a large number of individual crystallites, and therefore provide data on all the reflecting planes present. A particular set of planes will only reflect at a specific value of 2θ, and by observing the reflecting angles the planar spacings d_{hkl} can be determined. By reference to standard tabulated data (such as the *International Tables of Crystallography*), which give details of the reflections and intensities expected for most crystalline materials, a considerable amount of information can be determined. This includes the nature of the phases present, their preferred orientations and a measure of the microcrystallite size.

X-ray diffraction is the most commonly used technique. The penetration depth of X-rays can be quite significant, and if the film is too thin they can penetrate down to the substrate. Thus X-ray diffraction often gives information on the substrate as well as the film. If this is to be avoided, the film thickness needs to be increased or the X-rays directed at a low glancing angle with respect to the substrate surface. Electrons have a considerably lower penetration ability, and therefore give information only on the immediate surface region. Neutron diffraction is used particularly in the investigation of magnetic films.

28.2.2 Film Structure

Apart from the detailed type of crystal structures observed, it is useful to divide the structures of thin films into three main categories, according to their crystallographic perfection (or lack of it). Epitaxial films are essentially perfect films grown on a substrate surface with all the atoms arranged predictably in their correct crystallographic positions. There may be certain defects, such as vacancies, interstitial atoms or impurities, but the deviation from perfection should be relatively small. Furthermore, there may be some slight deviations in the unit cell dimensions from those of the corresponding bulk material, owing to idiosyncrasies of the deposition method or the substrate surface, and in particular to stresses built up in the film during deposition. Various methods have been devised to deposit epitaxial films, such as molecular-beam epitaxy (MBE) and vapour-phase epitaxy, both of which are used in the semiconductor industry for preparation of high-quality films. Non-epitaxial films are frequently polycrystalline, having a large number of grains, oriented in different directions, and separated by grain boundaries. The grain size is heavily influenced by the deposition conditions, and can have a profound effect on secondary properties, such as the electrical conductivity, in which grain boundaries influence the electron mean free path. Finally, amorphous films have no long-range order, being noncrystalline, with only short-range order apparent. Substrate temperature during deposition and subsequent thermal history after deposition can have a profound effect in determining the initial structure, orientation and subsequent phase changes. Some examples of these effects are included in the following discussion.

In the case of epitaxial films, a large number of factors determine whether the film is indeed epitaxial, and if so the crystalline orientation that will result. The main concepts that underlay the growth of epitaxial films are the equilibrium thermodynamics of the growing nuclei, the evaporation process as described by the Hertz–Knudsen expression (28.5), surface kinetics and mobility on the substrate and the growth of defects [28.110]. The interdependence and interaction of these various processes, which are generally all temperature-dependent, lead in many cases to a particular value of temperature (usually called the epitaxial temperature) only above which epitaxy can take place. Depending on the relative difference between the unit cell dimensions of the substrate and the films (often quantified as the *misfit*) there is frequently a considerable stress between them, which is released by a network of imperfections (dislocations). Details of the epitaxial growth are strongly related to various theories of nucleation. There are several of these, based

on the characteristics of microscopic nuclei on the substrate. These grow during the deposition process, and if the conditions are suitable, can result in the deposition of epitaxial films. In the capillarity model [28.9, 27] the nuclei become stable when they reach a particular size (the critical nucleus). Above this size the nucleus is stable and continues to grow. However, in this case the size of the critical nucleus is of the order of 100 atoms or more. An alternative theory of nucleation has been proposed for critical nuclei consisting of only a few atoms [28.111, 112]. In this model certain geometries have been postulated for the critical nucleus under different conditions of supersaturation and substrate temperature. The rate of nucleation is given by

$$I = Ra_0^2 N_0 \left(\frac{R}{\gamma N_0}\right)^{n^*} \times \exp\left(\frac{(n^*+1)Q_{\mathrm{ad}} - Q_{\mathrm{D}} + E_{n^*}}{k_{\mathrm{B}}T}\right). \quad (28.46)$$

In this expression the symbols have the following meanings: R is the impingement rate of atoms on the substrate, a_0 is a characteristic jump distance of the atoms across the substrate surface, N_0 is the concentration of adsorption sites, $\gamma = h/(k_{\mathrm{B}}T)$ is the vibrational frequency, where T is the substrate temperature, and n^* represents the number of atoms in the critical nucleus. The remaining terms in the numerator of the exponential represent energies: Q_{ad} is the binding energy of a single atom to the substrate surface, Q_{D} is the activation energy for surface diffusion of adsorbed atoms and E_{n^*} is the dissociation energy of a critical nucleus containing n^* atoms. By substituting $n^* = 1, 2, 3$, etc., into (28.46), expressions can be obtained for critical nuclei of size 1, 2 or 3 atoms. These correspond to stable nuclei of size $(n^*+1) = 2, 3$ or 4. By simultaneously solving pairs of these equations, it is possible to calculate the temperature at which the size of the critical nucleus changes from one value of n^* to another. For example, the temperature at which the critical nucleus size changes from one to two atoms, T_{12}, is given by

$$T_{12} = -\frac{(Q_{\mathrm{ad}} + E_2)}{k\ln(R/(\gamma N_0))}, \quad (28.47)$$

where E_2 is the dissociation energy of a two-atom nucleus into single atoms. The model has been tested, notably in the case of a face-centred cubic metal (silver) deposited on a high-purity (100)NaCl substrate [28.112]. At low substrate temperatures and high supersaturations, it was found that the nucleation rate I is proportional to R^2, which from (28.46) implies that $n^* = 1$ (the critical nucleus is just one atom). At higher temperatures, critical nuclei were found to have $n^* = 2$ or 3, with stable nuclei adopting the (111) or (100) orientations. Further discussion is given in several later reviews [28.1, 10, 16, 19]. More-sophisticated models of nucleation are based on the kinetic behaviour of nuclei, and involve a set of coupled differential equations, relating growth and dissociation of nuclei. More details are given in the literature [28.113, 114].

Another example of where the occurrence of epitaxy is defined by well-defined temperature ranges, is given in the work of *Kalinkin* et al. [28.115] *Muravjeva* et al. [28.116]. This important work has also been discussed in other reviews [28.117, 118]. These workers made extensive measurements on the evaporation of both cadmium and zinc chalcogenides (sulphides, selenides and tellurides), which were evaporated from Knudsen cells at a pressure of about 10^{-2}–10^{-3} Pa. The substrates used were mica in all cases. It was found that for all three cadmium compounds, and all three zinc compounds, the conditions for epitaxial growth were dependent on both the temperature of the substrate (the epitaxial temperature T_{Ep}) and the evaporator temperature T_{ev}. The latter influenced the evaporation rate N_{e} (28.5) and also the supersaturation ratio of the growing films. For each of the six compounds, different combinations of T_{Ep} and T_{ev} were investigated. We discuss in detail only the cadmium compounds, although the behaviour of the zinc compounds is entirely similar, as can be seen from Fig. 28.12 [28.116]. Below about $T_{\mathrm{Ep}} \approx 310$–$320\,^{\circ}\mathrm{C}$, the evaporator temperature for epitaxy T_{ev} increased approximately linearly with T_{Ep} according to $T_{\mathrm{ev}} = A_1 + T_{\mathrm{Ep}}$, where A_1 is a constant. Above $T_{\mathrm{Ep}} \approx 310$–$320\,^{\circ}\mathrm{C}$, T_{ev} decreased linearly with T_{Ep} according to $T_{\mathrm{ev}} = A_2 - 2T_{\mathrm{Ep}}$, where A_2 is a second constant. These relationships between the epitaxial temperature and the evaporator temperature are illustrated in Fig. 28.12 [28.116]. For the cadmium chalcogenides, values of the range of the constants A_1 and A_2 for which epitaxial behaviour was observed are $A_1 = 800$–$950\,^{\circ}\mathrm{C}$ and $A_2 = 1750$–$1900\,^{\circ}\mathrm{C}$ for CdS, $A_1 = 600$–$750\,^{\circ}\mathrm{C}$ and $A_2 = 1550$–$1700\,^{\circ}\mathrm{C}$ for CdSe, and $A_1 = 400$–$550\,^{\circ}\mathrm{C}$ and $A_2 = 1350$–$1500\,^{\circ}\mathrm{C}$ for CdTe [28.116]. It is evident from these values that for each substance the evaporator temperature T_{ev} lies within a range of about $150\,^{\circ}\mathrm{C}$ for values of T_{Ep} at which epitaxy is possible. In addition, for a given substrate temperature the evaporator temperature required for epitaxial CdS is highest, becoming progressively lower for CdSe and for CdTe. For example for $T_{\mathrm{Ep}} = 320\,^{\circ}\mathrm{C}$, the evaporator temperatures for epitaxy are 1120–$1270\,^{\circ}\mathrm{C}$ for CdS, 920–$1070\,^{\circ}\mathrm{C}$ for CdSe and 720–$870\,^{\circ}\mathrm{C}$ for CdTe. The same sequence is also followed for the zinc chalcogenides. At combinations of temperatures outside the limits given by the values of A_1 and A_2, the films deposited were not epitaxial, being polycrystalline and only partially ordered.

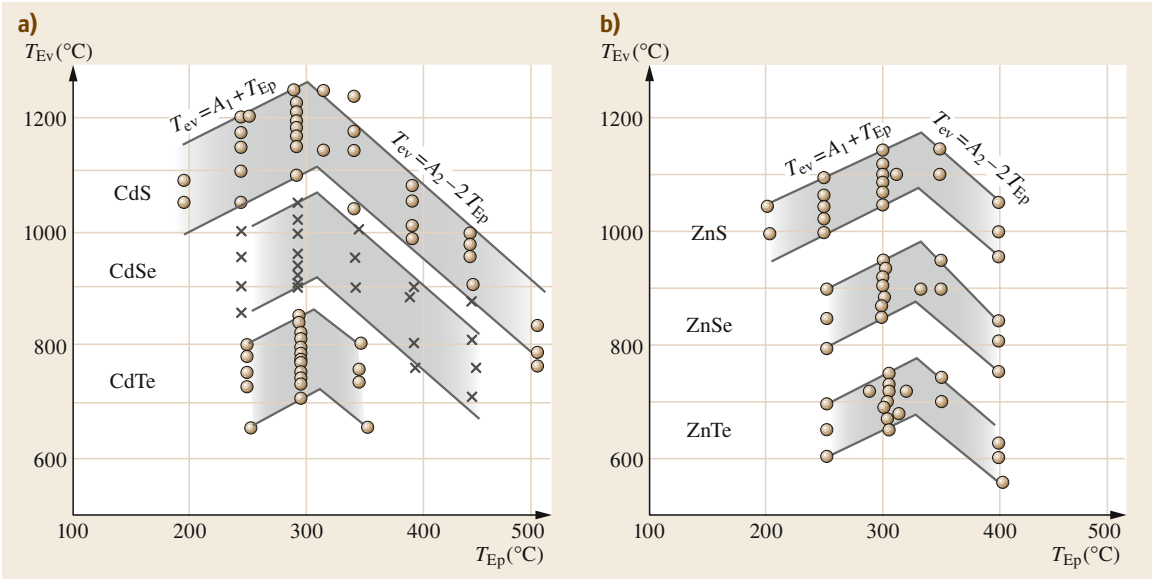


Fig. 28.12 Relationship between the evaporation temperature T_{ev} and the epitaxial temperature T_{Ep} in the growth of cadmium and zinc chalcogenide films on mica. (After [28.116] with permission from Elsevier Science)

We now consider some examples of polycrystalline thin films, and the factors which account for their various crystallographic characteristics. X-ray diffractometry is particularly useful in determining the basic crystal structure adopted by a particular film, particularly when two or more phases are possible. The cadmium chalcogenides have, on average, four valence electrons per atom, which leads to tetrahedral bonding. However, there are two different ways in which tetrahedra formed from each of the constituent atoms can penetrate into each other. This leads to two allowed crystal structures: the hexagonal wurtzite structure and the cubic zinc-blende structure. The planar spacings for these are described by (28.44) and (28.42), respectively. In the evaporant material, which is usually composed of a polycrystalline powder, a large number of diffraction peaks normally occur, which can be indexed using the appropriate planar spacing expression. Figure 28.13a [28.119] shows an example of a diffraction trace obtained from CdS evaporant powder. The diffraction planes are indexed in terms of the hexagonal structure, and rule out the possibility of the material being of the cubic form. More interestingly, Fig. 28.13b shows a trace obtained from a thin film prepared from this material. Although the two peaks shown correspond to those observed for the evaporant powder, it is clear that reflections from most of the crystal planes are missing. If the intensities of the peaks are different from those for a random powder (or more specifically from those tabulated in the *International Tables of Crystallography*), the crystallites are said to be preferentially oriented. In this case it can be seen that virtually all the crystallites must be aligned in the same direction, with the *c*-axis directed perpendicular to the substrate plane, along a preferential orientation in the [002] direction.

Apart from the simple determination of film crystal structures and preferential orientations, it is also possible to monitor changes from one crystal structure to another. For example, the phthalocyanines are organic crystals, which are known to possess several basic crystal structures, of which the α-phase and the β-phase are the most common. There is some controversy over whether the α-phase is orthorhombic or monoclinic, and it was originally identified as tetragonal [28.120]. The β-phase is monoclinic. However, X-ray diffraction has identified most films prepared at lower substrate temperatures as being of the α-structure, while those deposited at higher temperatures are of the β-structure. Both copper phthalocyanine (CuPc) and cobalt phthalocyanine (CoPc) films deposited on to substrates held at room temperature were of the α-structure, but on annealing at 240 °C for an extended period, the CuPc films underwent a phase transition to the β-structure [28.121]; the CoPc films also showed a transition to the β-phase on annealing at temperatures up to 325 °C [28.122]. The α-form films of both materials were preferentially aligned in the [001] direction assuming a tetragonal structure, or the [002] direction assuming an orthorhombic or monoclinic structure. The annealed β-form films were also preferentially oriented, in the [20$\bar{1}$] direction for CuPc

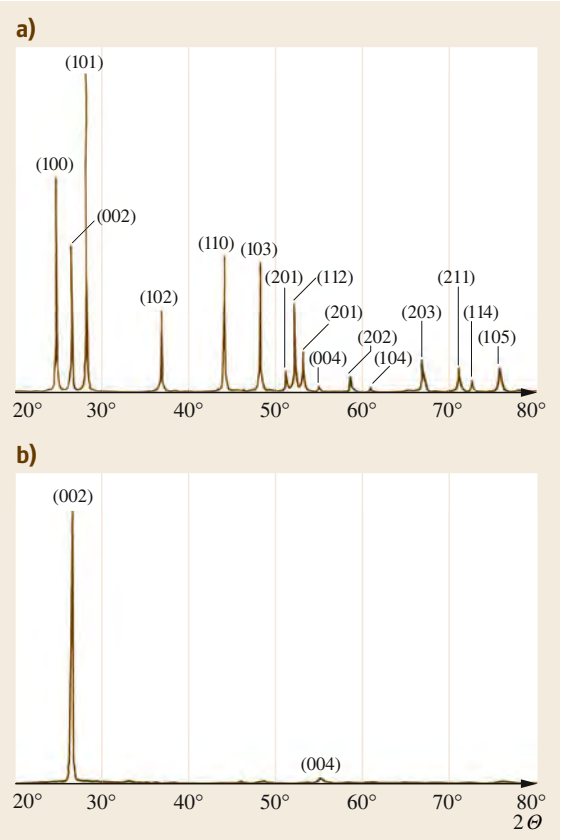


Fig. 28.13a,b X-ray diffraction traces obtained from **(a)** CdS powder, and **(b)** a CdS thin film. (After [28.119] with permission from Wiley-VCH)

and the [001] direction for CoPc. In the latter case it was possible to monitor the dominant peak over the temperature range 200–325 °C where the phase change occurred. d_{hkl} increased slightly with annealing temperature up to 200 °C, then decreased significantly at 250–300 °C. The value of 2θ decreased slightly with T_{ann} up to 200 °C, again followed by a significant *increase* at approximately 250–300 °C. For T_{ann} up to 250 °C d_{hkl} lay in the range 1.271–1.277 nm and 2θ in the range 6.92–6.95°, which is close to those expected for α form CoPc; at higher annealing temperatures d_{hkl} was in the range 1.241–1.246 nm and 2θ in the range 7.09–7.12°, which are in better agreement with those expected for β form CoPc.

X-ray diffraction also allows an estimate of the mean grain size L in polycrystalline films to be obtained, by detailed observation of the peak profiles. The well-known Scherrer equation [28.123]

$$L = \frac{K\lambda}{\eta \cos\theta} \qquad (28.48)$$

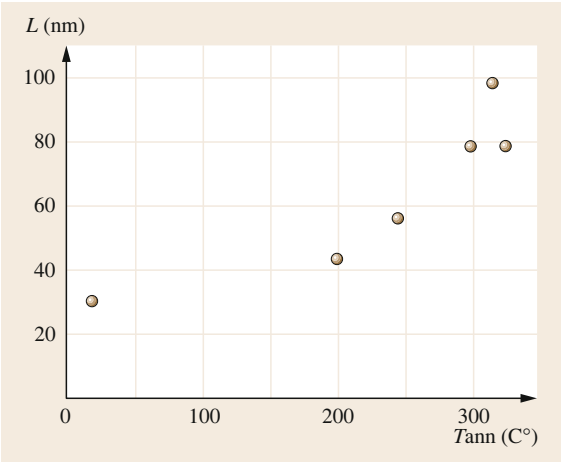


Fig. 28.14 Mean crystallite grain size L as a function of maximum cumulative annealing temperature T_{ann} for a cobalt phthalocyanine film of thickness 0.1 μm. (After [28.122] with permission from Wiley-VCH)

allows an estimate of L to be made from simple details of the peak profiles. In (28.48), λ represents the X-ray wavelength, η the width of a strong peak in radians at half-maximum intensity, and θ is the Bragg angle, where K is a constant which is approximately unity. An example of the use of this method is in the determination of L from the first Bragg peak of α-form CoPc films before annealing [28.122]. The films were subsequently heated to a higher temperature T_{ann} in oxygen-free nitrogen for two hours and the value of L again measured. Measurements were repeated at even higher temperatures until the dependence of L on T_{ann} over the entire temperature range of the α-to-β phase change was determined. These results are illustrated in Fig. 28.14 [28.122]. It is clear that at temperatures up to 200 °C there was an increase in the mean microcrystallite size from about 30 to over 40 nm, although there was no indication of a phase change. *Ashida* et al. [28.124] ascribed this effect in CuPc to preliminary crystallite growth at the early stages of the phase transformation process. Above annealing temperatures of about 200 °C there was a rapid increase in crystallite size to about 100 nm at 315 °C; similar behaviour has also been observed in which α-form CuPc films were heated to 300 °C when the diffraction patterns became sharper and the crystallites larger [28.124].

Some general conclusions concerning the dependence of grain size in vapour-deposited polycrystalline films on various film attributes have been summarised by *Chopra* [28.18]. The major feature determining mean grain size is the surface mobility of adsorbed atoms and clusters during the deposition process. A high surface mobility will allow adsorbed atoms to

move large distances before being incorporated into growing nuclei, resulting in a relatively low concentration of nuclei and thus larger crystallites. The substrate temperature is the most obvious factor contributing to high mobility, and therefore grain size tends to increase with increasing substrate temperature. We have already seen that epitaxy requires a minimum substrate temperature; this too is a consequence of the surface mobility requirement. The total film thickness obviously sets an upper limit on the crystallite size, and therefore the grain size increases in general with film thickness. Subsequent annealing at temperatures greater than the substrate temperature during deposition can also cause an increase in the mean crystallite size as was shown in Fig. 28.14, as surface atoms augment the crystallites and grains coalesce. The dependence of grain size on deposition rate is not consistent between different materials and substrates. Frequently there is an increase with increasing deposition rate [28.18], although a decrease has been observed in some cases, such as in CdTe films where the establishment of a higher concentration of adsorbed nuclei led to the growth of finer grains [28.38].

The interplay between the various deposition parameters can have a profound effect on the structure of the films. This is illustrated particularly well in Fig. 28.15 [28.125], for cadmium arsenide (Cd_3As_2) films evaporated onto mica substrates. Both the substrate temperature T_s and the growth rate v were varied, with the latter covering a wide range. Diffraction patterns indicated that the crystalline films were of the lower temperature tetragonal α-phase of Cd_3As_2; the structure of this phase is documented in the literature [28.126] and the planar spacings are given by (28.43). Amorphous films recrystallised following interaction with the electron beam in the electron microscope and also reverted to a tetragonal structure. The figure shows that, over the substrate temperature and deposition rate ranges investigated, the films could have an amorphous, polycrystalline or well-oriented crystal structure. Below approximately 125–145 °C the films were amorphous, with polycrystalline films resulting when this temperature was exceeded. The amorphous-to-crystalline transition was only weakly dependent on the deposition rate. However, the epitaxial temperature, which delineates polycrystalline and epitaxial films, increased with increasing deposition rate. For deposition rates less than about 4 nm/s (40 Å/s) there was a direct transition from the amorphous to the well-oriented crystalline structure, whereas at higher rates there was a transitional polycrystalline region. The preferred orientation of the films was [001] with the c-axis of the tetragonal structure perpendicular to the plane of the substrate. Also shown on the figure are examples of X-ray diffraction patterns for the crystalline, polycrystalline and amorphous regions. These are characterised, respectively, by discrete reflection points, *spotty* rings and diffuse rings.

Finally, a brief discussion on amorphous films is appropriate. These are films which show no evidence of a regularly repeated crystalline structure, i. e., they do not exhibit long-range order. We have already seen that

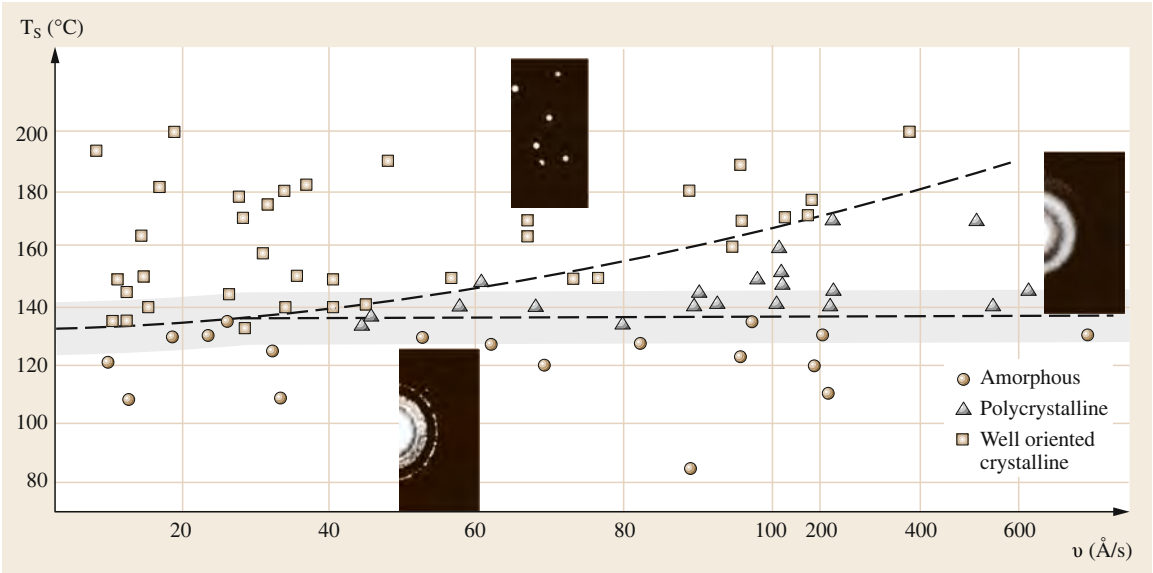


Fig. 28.15 Effects of film growth rate υ and substrate temperature T_s on the structure of Cd_3As_2 films vacuum-evaporated onto mica substrates. (After [28.125] with permission from Elsevier Science)

in Cd_3As_2 films (Fig. 28.15), a relatively low substrate temperature during deposition leads to the growth of amorphous films. This is a general rule, applicable to most materials, and is basically the opposite to the condition for epitaxy – the deposition of amorphous films is encouraged if the surface mobility of atoms on the substrate is low. This causes the adsorbed atoms to be incorporated into the films almost at the point of impact, rather than after diffusion across the substrate surface and joining a growing nucleus. High deposition rates also predispose atoms to remain at their impact position within the growing film – they have only a very limited time to diffuse before they are locked into the growing film structure. These considerations are pertinent to the deposition of amorphous metallic films, which require substrate temperatures to be lowered to liquid-nitrogen temperature or below, and also to elemental chalcogenide and semiconductor films (Ge and Si) and carbon, all of which have a sufficiently low surface mobility at room temperature. A table giving examples of amorphous thin film materials, their preparation methods, and the transition from the amorphous to a crystalline phase is given by *Maissel* and *Francombe* [28.16]. The amorphous state is metastable, requiring some thermal energy to transfer it to a crystalline state.

Another method of producing amorphous films is the incorporation of residual gases during deposition. A relatively high pressure of oxygen ($\approx 10^{-3}$ Pa) can cause oxidation of the evaporant. Large-scale crystallites of metals are then unable to form. The admission of nitrogen during sputtering of refractory metals is also sufficient to result in the deposition of an amorphous phase [28.1].

Since the structure of amorphous materials do not possess any long-range order, it is necessary to describe the short-range order using some function which can be determined directly from diffraction experiments. The radial distribution function $J(r)$ is defined as the number of atoms lying at a distance between r and $r + \mathrm{d}r$ from the centre of an arbitrary atom at the origin [28.127]. It is given by

$$J(r) = 4\pi r^2 \rho(r) \,. \tag{28.49}$$

Here, $\rho(r)$ is an atomic pair correlation function, which is equal to zero at values of $r < r_1$, where r_1 is the average nearest neighbour interatomic separation, $\rho(r)$ is equal to the average value of density ρ^0 at large values of r where the material becomes homogeneous. Between these values $\rho(r)$ is oscillatory, with the peaks in the distribution corresponding to the average interatomic separations [28.127]. $J(r)$ also shows oscillations superimposed on the average density parabola $4\pi r^2 \rho^0$, and the area under a particular peak gives

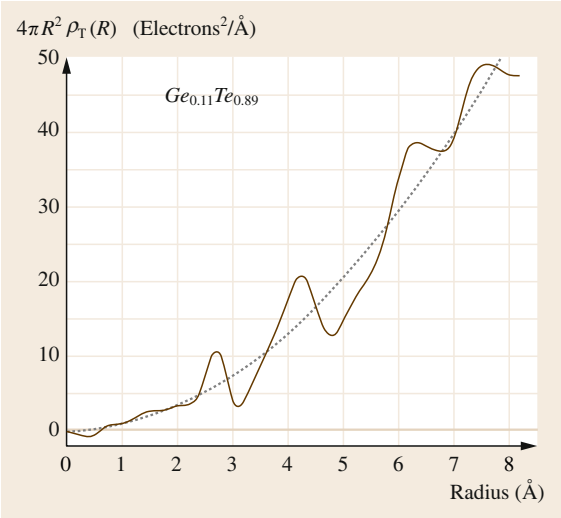


Fig. 28.16 Calculated radial distribution function for a sample of composition $Ge_{0.11}Te_{0.89}$. The *dashed curve* represents the contribution of the average electron density. (After [28.128] with permission from Elsevier Science)

the coordination number for the given shell of atoms. The coordination number represents the number of nearest neighbours in the structure. An example of a radial distribution function, derived from X-ray data, for thin films of the compound Ge_xTe_{1-x} where $x = 0.11$, is shown in Fig. 28.16 [28.128]. For materials having covalent bonds the coordination number is four, whereas for amorphous metals, which generally crystallise in close-packed structures, the coordination number is higher. For covalently bonded materials, the atoms tend to arrange themselves in tetrahedra, which in the amorphous state interconnect at their apexes. However, the apexes themselves interconnect within the material in a random manner, whereas the tetrahedra dimensions are effectively determined by the size of the atoms. Thus there is no long-range order because of the random interconnections, but the short-range order survives, and is similar to that for the corresponding crystalline material. *Bosnell* [28.129] has described the important amorphous semiconducting thin films, and classified them into three major categories: covalent noncrystalline solids, semiconducting oxide glasses and dielectric films. The importance of these materials, in particular their electrical properties, cannot be overestimated. The standard text on this subject [28.130] gives copious information on their electrical properties, as well as their structures and radial distribution functions. Local atomic order in amorphous films is discussed in the literature by *Dove* [28.131], with the primary emphasis being on data derived from electron diffraction. This work also gives the theory behind the radial dis-

tribution function, and provides a wide review of radial distribution functions measured in many systems performed prior to the date of the review.

From the foregoing discussion and examples, it should be clear that the crystallographic structure of thin films, whether epitaxial, polycrystalline or amorphous, depends heavily on the deposition conditions. In turn the film structure strongly influences the mechanical, thermal, optical and electronic properties.

28.2.3 Morphology

Perhaps the most striking morphological characteristic observed in thin films grown by the physical vapour deposition techniques is that of the columnar grain structure. Electron microscopy frequently reveals that high-density regions tend to grow in parallel columns. They tend to develop when the surface mobility of adsorbed atoms is relatively low, so that the area of the base of the columns is small. Further growth of the grains is away from the substrate. The direction of column growth is not necessarily perpendicular to the substrate surface, although it frequently is for depositing material arriving at normal incidence. For oblique incidence it has been observed that the columns are oriented away from the substrate normal and towards the vapour source, although not necessarily pointing directly towards it. If α is the angle between the substrate normal and the vapour stream, and β is the angle between the substrate normal and the direction of the columns, it has been shown experimentally that β is usually less than α. The tangent rule [28.133] relates these two angles and is given by

$$\tan\alpha = 2\tan\beta \qquad (28.50)$$

and was found to be followed for the case of directly evaporated aluminium. In experiments on CdS films, *Hussain* [28.134] deposited films of thickness up to 100 μm with angle of incidence α varying between 0 and 60°. Up to a certain film thickness the *c*-axis of the grains was perpendicular to the direction of vapour incidence, but above this thickness the *c*-axis started to shift towards the substrate normal, eventually becoming aligned with it. Etching of a thin film deposited at oblique incidence confirmed that the top layers of the film had their *c*-axis normal to the substrate, whereas in the bottom layers it was parallel to the direction of vapour incidence. In other work on this material [28.135] similar thickness-dependent orientation effects have been reported. In this case, films of thickness a few tens of nanometres had randomly orientated crystallites, while those of thickness of about 20 μm were aligned with the *c*-axis normal to the substrate.

Evaporated CdS films also show an interesting morphology on a somewhat larger scale, which is illustrated in Fig. 28.17 [28.132]. This illustrates features which are evident in many different types of thin films. The top surface of as-deposited films of thickness 25–30 μm were observed by scanning electron microscopy to consist of partially faceted hills with a height of 5 μm and a width of about 15 μm at the base. These workers also determined the grain size and crystal defect structure using transmission electron microscopy. The grain size was about 2 μm at the top and at the centre of the films, but was only of the order of 10–50 nm at the back surface adjacent to the substrate. Typical dislocation densities were of the order of 10^{13} m^{-2} (10^{9} cm^{-2}). Ultrafine-grain regions of 2–15 μm in diameter were observed. These were thought to be spits of material rapidly ejected from the source. Similar features have also been observed in this laboratory [28.119]. The partially faceted protrusions were reported, and cracks in the films were also present, as a result of differential thermal contraction between the CdS films and the glass substrates during cooling. Larger regions of CdS were seen, which were considered to have been ejected from the evaporation boat during deposition. This phenomenon has been referred to as *spattering* or *splattering* by *Stanley* [28.136], and its elimination is an important goal for deposition processes operating at higher temperatures.

In early work on CdTe films [28.137] photovoltages were measured when there was an angle between the vapour stream and the normal to the substrate. This led

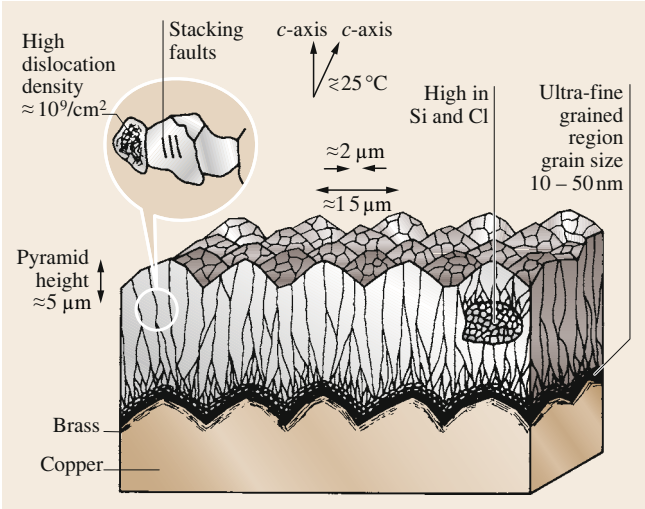


Fig. 28.17 Schematic representation of evaporated CdS film microstructure inferred from scanning electron microscopy studies. (After [28.132] with permission from Elsevier Science)

to the suggestion that in this case the film growth was directionally anisotropic where the nucleating material depositing on a heated substrate has sufficient surface mobility to form islands rather than a uniform thin film. Additional material arrives at an angle, shadowing the far side of the island from the deposition source, and the shadowed side grows relatively slowly and is well ordered, while the exposed face grows towards the vapour source. When islands begin to touch each other there is expected to be a mismatch, leading to a disordered state in the region between the crystallites. Each interface will therefore have an ordered and a disordered side and at the interface between grains a photovoltage develops. Photomicrographs of epitaxial CdTe layers grown by the close-space technique [28.138] revealed smooth epitaxial layers when the deposition was on to (100), (110) and (111)Cd oriented substrates, but large hillocks grew on (111)Te substrate faces. When grown on mica substrates [28.139] it was observed that the morphology was significantly affected by the composition of the incident vapour phase. This was characterised by the ratios γ_{Cd} and γ_{Te}, which depended on the ratio of the actual concentration to the stoichiometric concentration of the Cd atoms or the Te_2 molecules in the gas phase. When $\gamma_{Cd} = 53$ when CdTe and Cd were co-evaporated the films were of high quality consisting of large crystallites with degenerated (111) faces. When CdTe and Te were co-evaporated and $\gamma_{Te} = 12$ formations resembling dendrites were observed.

CdSe film morphologies have been observed by several workers. *Tanaka* [28.140] prepared films using Ar and H_2 sputtering gases; some films were subsequently annealed in Ar. Microscopic examination showed that sputtering in H_2 gas led to uniform grain growth of the films over the whole substrate surface, while films sputtered and annealed using Ar gave rise to geometric pillars and a textured structure. An investigation of the optimum conditions for obtaining CdSe epitaxial layers on (0001) oriented sapphire substrates was performed by *Ratcheva-Stambolieva* et al. [28.141]. For substrate temperatures up to 460 °C the layers were polycrystalline, from 460 to 530 °C a small degree of preferred orientation was observed, with a more significant textured structure for temperatures of 530–570 °C. At higher temperatures of 570–620 °C a monocrystalline structure was obtained, varying in its degree of perfection depending on the actual temperature and the growth rate. Hexagonal pyramids with peaks, regular or irregular facets and flat tops were also identified. RF-sputtered CdSe films [28.142] were porous when deposited at a substrate temperature of 550 °C onto vitreous silica. When deposited onto substrates of single-crystal sapphire the surface appeared shiny, and showed features of hexagonal flat tops and pyramids with edge lengths in the range 1–5 μm.

We finally look at the morphological characteristics of another cadmium compound, Cd_3As_2. Evaporated films of this material have been investigated by *Jurusik* and *Żdanowicz* [28.143] using electron microscopy. It was found that the microstructure of the films was strongly dependent on the substrate temperature T_s. This type of behaviour is frequently addressed in terms of a zone model, and examples of some of these for other materials are discussed by *Ohring* [28.19]. In Cd_3As_2 the microstructures were interpreted in terms of a simple three-zone model in which the structure was correlated with the ratio T_s/T_m, where $T_m = 994$ K is the melting point of Cd_3As_2 [28.143]. They found that for $T_s/T_m \leq 0.4$ the films were amorphous, consisting of spherical or elongated fibrous-like clusters, or were uniform with a large number of pores or cavities. For $0.4 < T_s/T_m < 0.5$ columnar growth was identified, while for $T_s/T_m \geq 0.5$ films with large crystallites and clear grain boundaries were observed. These structures are illustrated in Fig. 28.18 [28.143]. More recently a very detailed study was performed on amorphous Cd_3As_2 films evaporated onto substrates maintained at 293 K [28.144, 145]. It was found that the size of post-nucleation islands was a function of the substrate temperature, type of substrate and the deposition rate. Furthermore the shape and type of growth of the islands varied with these parameters. At a low deposition rate (1 nm/s) the initial nuclei grew until a certain size was reached, and new nuclei started to grow around the original islands. For a moderate deposition rate (3.5 nm/s) the initial nuclei grew into islands which in turn increased in size until the islands had only narrow interfaces between them. Some islands formed polygons. Films deposited at a high rate (50 nm/s) consisted of small islands connected by relatively large low-density areas. Films deposited at low rates were of approximately stoichiometric proportions (i. e., 40 at. % As) while those deposited at higher rates had excess arsenic (typically 51.8 at. % As). It was suggested that the growth process could be considered as a superposition of two growth stages, as shown in Fig. 28.19 [28.144]. The vapour flux onto the substrate is denoted by S_1 and the adatom flux over the substrate surface by S_2. In the first stage of growth the initial nuclei enlarge owing to the adatom flux S_2. The second stage of growth incorporates atoms immediately from the incident flux S_1. This is usually very slow, but increases as the island surface area increases. At this stage branched complex islands or only compact islands grow as the result of a shadowing mechanism. When the islands touch each other the molecular flux S_2 over the substrate surface is eliminated, and the films grow in thickness

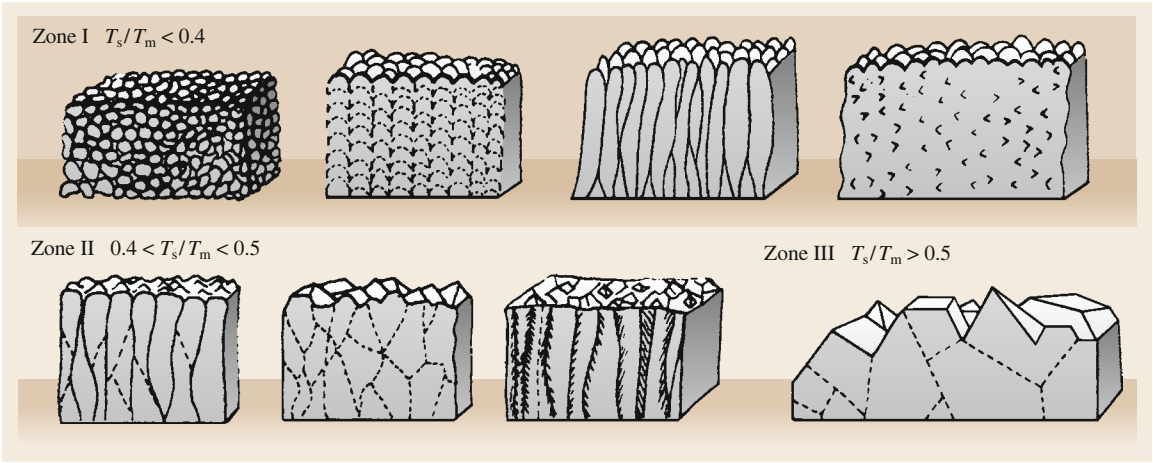


Fig. 28.18 Schematic three-zone model for the temperature dependence of the growth morphology in evaporated Cd_3As_2 films. T_s/T_m represents the ratio of the substrate temperature to the melting point of Cd_3As_2. (After [28.143] with permission from Elsevier Science)

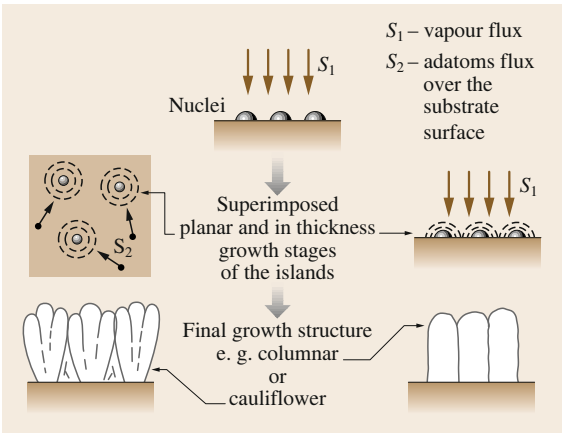


Fig. 28.19 Post-nucleation growth in evaporated Cd_3As_2 films considered as a consequence of two superimposed (planar and thickness) growth stages. (After [28.144] with permission from Elsevier Science)

as the result of the incident flux S_1 and the structure initiated in the earlier stage (i. e., large branched or smaller compact islands). The final growth structure gives rise to cauliflower-like or columnar structures. In later work [28.145] Cd_3As_2 films were evaporated from bulk crystallites of Cd_3As_2 or $CdAs_2$. For films grown from the Cd_3As_2 source a pronounced columnar structure was observed, whereas those evaporated from the $CdAs_2$ source grew uniformly, and the columnar structure was absent. In the latter case the initial island structure was not observed in the early stages of film growth. It was concluded that prerequisites for columnar growth were a limited surface mobility and an oblique component in the deposited flux, particularly where the films are deposited normally to the substrate surface.

In this section some of the morphological characteristics of some cadmium compound films prepared by PVD methods have been described. The emphasis in this discussion of the films was not primarily because of their usefulness as semiconductors in various applications, bur because they exhibit many features typical of a wide range of other film materials. Surface mobility of atomic species on the substrate during deposition is clearly of great importance, as this tends to control the grain size at the substrate interface, while the angle of vapour incidence relative to the substrate normal is an important factor in determining the major direction of crystallite growth and whether a columnar structure develops. Shadowing of the vapour stream by islands during film growth can lead to differential growth rates at different points on the film surface. Bulk diffusion and desorption also contribute to the film morphology variations [28.19]. The morphological characteristics of many materials may be explained using zone models in which the substrate temperature and other variables, such as sputtering gas pressure, conspire to determine the overall morphology via the mechanisms detailed above. The morphology of thin films is particularly rich, and an important factor in influencing film properties.

28.2.4 Raman Spectroscopy

Raman spectroscopy is a nondestructive spectroscopic technique used to study the vibrational, rotational, and other low-frequency modes of a system based on the inelastic scattering of a monochromatic light. A laser

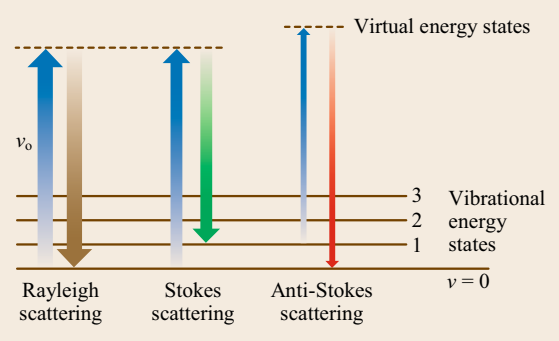


Fig. 28.20 Schematic illustration of the vibrational energy states involved in Raman scattering. The thickness of the *arrows* is a rough indication of the signal strength. (After [28.146])

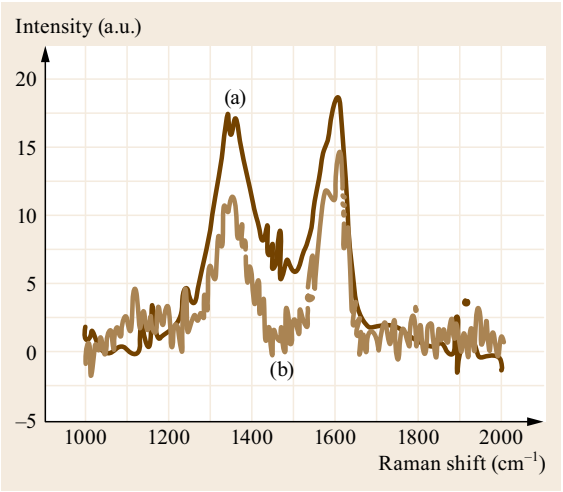


Fig. 28.21 Raman spectrum for the sample (a) freshly prepared before (b) RF hydrogen plasma treatment at 60 W for 30 min

light, usually in the visible, near-infrared or near-ultraviolet range, is used as the incident beam and interacts with molecular vibrations, phonons, and other excitations. The interaction between the laser beam and the electrons and bonds of a molecule results in an excitation of the molecule from the ground state to a virtual energy state. The relaxation of the molecule is accompanied by the emission of a photon and a change in rotational or vibrational state of the molecule. The energy difference between this new state and the original state causes a shift in the emitted photon frequency. A schematic illustration of vibrational energy states involved in Raman scattering is shown in Fig. 28.20.

Raman scattering is regarded as being inelastic because of an energy transfer during the photon–molecule interaction. When the incident photon with a frequency of ν_0 is absorbed by a Raman-active molecule in the ground vibrational state, some of the photon energy with a frequency of ν_m is transferred to the molecule and the scattered light is shifted to the lower frequency of $\nu_0 - \nu_m$. This shift in the frequency is also known as the Stokes shift. If the photon energy is absorbed by a molecule which is already in the excited state, the excess amount of energy is released from the molecule upon relaxation and the resulting frequency of the scattered light is increased to $\nu_0 + \nu_m$, or the anti-Stokes frequency. The Raman shift is typically expressed in wavenumbers. The following formula relates the wavenumbers of shift in the spectrum to the spectral wavelengths [28.106]

$$\Delta\omega = \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right) \times 10^7 \,, \tag{28.51}$$

where $\Delta\omega$ is the Raman shift in cm^{-1}, and λ_0 and λ_1 are the excitation and Raman spectrum wavelengths in nm, respectively. Since the vibrational information obtained from Raman spectroscopy is specific to the chemical bonds and the symmetry of the molecules, it can be useful in order to identify the chemical composition of the material.

Commercially available graphene oxide (GO) nanoparticles have been utilized to produce solution-processed thin films of GO on silicon substrates. The films were then reduced by 60 W RF hydrogen plasma at room temperature for 30 min in an ordinary laboratory environment. Figure 28.21 shows the Raman spectrum for the samples before and after plasma treatment. The structural changes occurring due to plasma treatment could be evidenced from the Raman spectrum. The characteristic D and G bands corresponding to 1350 and 1578 cm^{-1} were present in both samples.

The D band response is related to defects or any sort of translational disorder, whereas the G band corresponds to scattering from the E_{2g} mode of the sp^2 graphitic domain. It can be seen in the figure that plasma treatment leads to a shift in the D and G band from 1348.3 to 1357.7 cm^{-1} and 1594.7 to 1593.5 cm^{-1}, respectively. The width of the D band seems to change from 76.12 to 130 cm^{-1}, whilst the G band width also changes from 60.25 to 81.693 cm^{-1}. The present sample shows an increase in intensity ratio of both the D and G band from 0.77 to 0.93 with plasma treatment indicating a decrease of the in-plane sp^2 graphitic domain size. This may imply that plasma treatment has induced a decrease in the size of the in-plane sp^2 graphitic domain, an increase in edge planes, and also an increase in disorder in the samples [28.147].

The band at 437 cm^{-1} in the Raman spectra in Fig. 28.22 for spin-coated zinc oxide thin films on

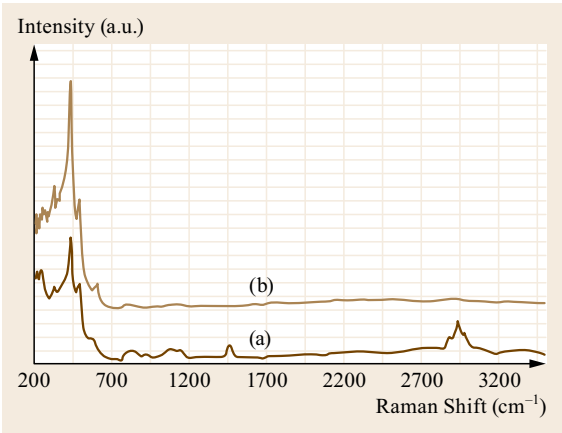


Fig. 28.22 Raman scattering for (a) as-deposited and (b) annealed ZnO coatings

fluorine-doped tin oxide (FTO) coated glass slides may be identified as the nonpolar E_2 mode of the scattering. This marker band became sharper on annealing indicating the dominance of hexagonal crystalline structure of wurtzite ZnO. The polar E_1 mode due to the interaction between electrons and transverse-optical phonons was observed as a shoulder at 549 cm^{-1} for the as-deposited spectrum. The polar A_1 mode bands at 748, 1048, and 1447 cm^{-1} arising from the superimposition of longitudinal-optical and transverse-optical phonons are found to be considerably diminished in intensity on annealing due to the elimination of impurities or defects in the crystals.

Bands in the range 2800 and 3000 cm^{-1} were observed indicating the presence of the hydroxyl groups in the hydrated as-deposited film. The peak at 2943 cm^{-1} disappeared upon annealing of the film [28.148].

28.3 Properties

As mentioned in the introduction, the properties covered in this section are restricted to the optical and electrical properties. Moreover, only the very fundamental properties, which are a direct consequence of the thin-film nature of the material, are covered. Thus the more advanced features, such as those exhibited by specific classes of materials, individual elements or compounds, and electronic, optical and optoelectronic devices are not normally discussed, although many of these are considered elsewhere in this Handbook. The basic thin-film properties, which have been exploited in many of these developments are stressed. Examples include the use of thin films in layers or stacks in antireflection coatings and mirrors [28.30], the contribution of grain boundaries in conducting thin films to the resistivity of tracks and connections in integrated circuits, the importance of high-field conduction in some electronic devices [28.17], and the possible influence of high-field conduction in nanostructures [28.11].

28.3.1 Optical Properties

This section comprises an introduction to the main optical properties of thin films, starting with a discussion of the properties of nonabsorbing single films at normal and non-normal incidence, and then considering modifications due to absorption. Finally a brief discussion of the use of multilayer film stacks is given. Thin-film optics is a particularly wide field, and readers interested in further details and a more in-depth analysis are referred to one of the standard texts [28.30, 149–151], or indeed to most general textbooks on optics, nearly all of which cover thin films in some detail. The general description of optical constants focused on electronic materials and various techniques for their determination in the case of bulk and thin film materials is also given in this handbook in Chap. 3. As mentioned earlier, optical effects in thin films were first recorded in the 17th century, which predates by two centuries the emergence of Maxwell's equations, which encapsulate the behaviour of all electromagnetic radiation, including visible light. The boundary conditions at the interface between two different optical media are derivable directly from Maxwell's equations, and allow us to calculate the reflection and transmission coefficients for the amplitude of the electromagnetic wave at the interface. These are termed the Fresnel coefficients.

Figure 28.23a shows the simplest situation, where a beam of light in a medium having refractive index n_0 is incident at an angle φ_0 to the normal on a film of refractive index n_1. The following expressions for the Fresnel coefficients apply to all individual interfaces in a structure, and for all values of the refractive index. At this stage in the discussion we assume that the materials are isotropic and nonabsorbing. At the interface the electromagnetic radiation is refracted into the medium of index n_1 at an angle φ_1, with a reflected component directed back into the incident medium at an angle φ_0. Conventionally, the coefficients are quoted for components of the light beam resolved parallel (denoted by p) and perpendicular (denoted by s) to the plane of incidence. Therefore there are two reflection coefficients r_{1p} and r_{1s}, and two transmission coefficients, t_{1p} and t_{1s}. The 1 in the subscripts in this notation refers to the first interface, i. e., between the materials with refractive indices n_0 and n_1 in this case. More complex systems will

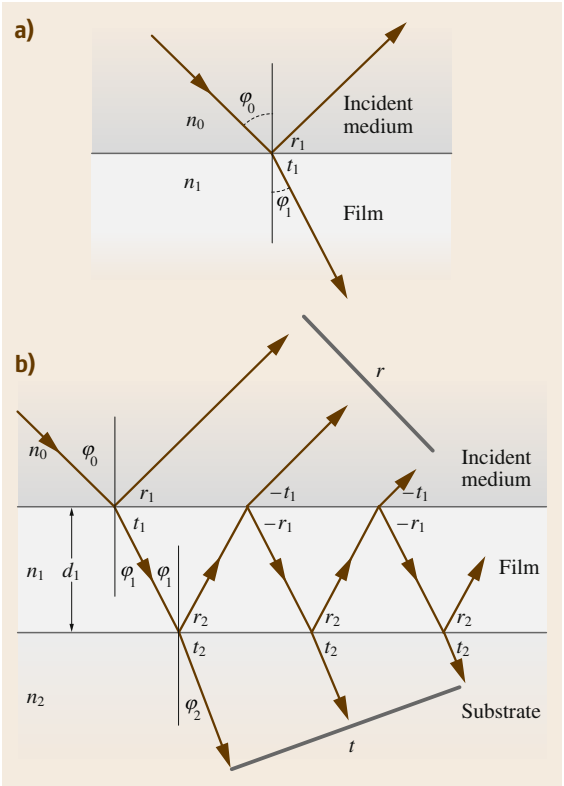


Fig. 28.23 **(a)** Reflection and transmission at a single interface between two media with refractive indices n_0 and n_1. r_1 and t_1 represent the amplitude reflection and transmission coefficients at the interface (the Fresnel coefficients). **(b)** Total reflection and transmission as the result of multiple reflection and transmission through a nonabsorbing thin film of thickness d_1 and refractive index n_1 deposited on a substrate of refractive index n_2. The Fresnel coefficients are indicated at each transit of a boundary, and change sign if the direction of incidence is reversed. r and t represent the total amplitude of reflection and transmission, respectively, each obtained by summing the amplitudes of an infinite number of components, although only three are shown ►

have more than one interface, where 2, 3, etc., signify the second or third interfaces. The Fresnel coefficients for reflection are [28.1, 15]

$$r_{1p} = \frac{n_1 \cos\varphi_0 - n_0 \cos\varphi_1}{n_1 \cos\varphi_0 + n_0 \cos\varphi_1} ,$$

$$r_{1s} = \frac{n_1 \cos\varphi_1 - n_0 \cos\varphi_0}{n_1 \cos\varphi_1 + n_0 \cos\varphi_0} , \qquad (28.52)$$

and those for transmission are

$$t_{1p} = \frac{2n_0 \cos\varphi_0}{n_1 \cos\varphi_0 + n_0 \cos\varphi_1} ,$$

$$t_{1s} = \frac{2n_0 \cos\varphi_0}{n_1 \cos\varphi_1 + n_0 \cos\varphi_0} . \qquad (28.53)$$

Clearly these coefficients reduce to

$$r = \frac{n_1 - n_0}{n_1 + n_0} , \quad t = \frac{2n_0}{n_1 + n_0} \qquad (28.54)$$

for the case of normal incidence where all the cosine terms become unity. There is no difference in this case between the coefficients for the p and s components. An expression for the irradiance (or energy) reflectance at the interface R can be obtained by evaluating rr^*, where r^* is the complex conjugate of r. For nonabsorbing films r is real and this reduces to r^2, and gives for normal incidence

$$R = \left(\frac{n_1 - n_0}{n_1 + n_0}\right)^2 . \qquad (28.55)$$

A similar expression for the irradiance (energy) transmittance T at the boundary can also be derived from t, but in this case the incident and transmitted beams are in different media, and the expression is not simply t^2, but includes the ratio n_1/n_0. Full details are given elsewhere [28.18, 150].

We consider now the slightly more complex situation of a thin film of thickness d_1 and refractive index n_1 deposited on a substrate of refractive index n_2. In this case the value of n_0 for the incident medium will usually be the free-space value of unity. There will be multiple reflections as shown in Fig. 28.23b. It is clear that the total amplitude of the reflected beam is given by summing the amplitudes of the individual reflections into the incident medium, while the total amplitude of the transmitted beam is obtained by summing the amplitudes of the transmitted components into the substrate. In traversing the film of thickness d_1, the radiation undergoes a change in phase, which is accounted for by a *phase thickness* δ_1, given by

$$\delta_1 = \frac{2\pi}{\lambda} n_1 d_1 \cos\varphi_1 , \qquad (28.56)$$

where λ is the wavelength in vacuum. In each case the summations lead to infinite series. These involve $\exp(-i2\delta_1)$ for reflection, where the beam traverses the film an even number of times, and $\exp(-i\delta_1)$ for transmission, where the beam traverses an odd number of times. The results of these summations give the reflection coefficient (sometimes called the reflectance amplitude) r and transmission coefficient t, respectively, as

$$r = \frac{r_1 + r_2 \exp(-2i\delta_1)}{1 + r_1 r_2 \exp(-2i\delta_1)} \qquad (28.57)$$

and

$$t = \frac{t_1 t_2 \exp(-\mathrm{i}\delta_1)}{1 + r_1 r_2 \exp(-2\mathrm{i}\delta_1)} \,. \tag{28.58}$$

(These correspond to (3.86) and (3.87) in Chap. 3. Note also that the three refractive indices were n_1, n_2 and n_3 whereas n_0, n_1 and n_2 are used here.) In this case r_1 and t_1 represent the Fresnel coefficients between n_0 and n_1, and r_2 and t_2 the coefficients between the media n_1 and n_2. Depending on whether we are considering parallel or perpendicular polarisation (p or s) the relevant Fresnel coefficients from (28.52) and (28.53) should be used in these expressions. Equation (28.57) and (28.58) refer to the ratios of the *amplitudes* of the reflected and transmitted beams to that of the incident beam. In terms of the beam irradiances (or beam energies), the reflectance R and the transmittance T may be defined as the ratios of the reflected and transmitted beam energies to that of the incident energy. The energies of the reflected and transmitted beams are given by $n_0 RR^*$ and $n_2 tt^*$ respectively, where r^* and t^* represent the complex conjugates of r and t. We then obtain

$$R = \frac{n_0}{n_0} rr^* = \frac{r_1^2 + r_2^2 + 2r_1 r_2 \cos 2\delta_1}{1 + r_1^2 r_2^2 + 2r_1 r_2 \cos 2\delta_1} \tag{28.59}$$

for the reflectance, and

$$T = \frac{n_2}{n_0} tt^* = \frac{n_2}{n_0} \frac{t_1^2 t_2^2}{\left(1 + 2r_1 r_2 \cos 2\delta_1 + r_1^2 r_2^2\right)} \tag{28.60}$$

for the transmittance. Although these latter two expressions are reasonably compact, when the four Fresnel coefficients are substituted they become somewhat unwieldy. However, for the important case of normal incidence, where (28.54) applies, and where (28.56) simplifies to $\delta_1 = (2\pi/\lambda) n_1 d_1$, they reduce to still fairly long expressions, but involving only the three refractive indices and δ_1, i. e.,

$$\begin{aligned} R = \Big[& \left(n_0^2 + n_1^2\right)\left(n_1^2 + n_2^2\right) - 4n_0 n_1^2 n_2 \\ & + \left(n_0^2 - n_1^2\right)\left(n_1^2 - n_2^2\right)\cos 2\delta_1 \Big] \Big/ \\ & \Big[\left(n_0^2 + n_1^2\right)\left(n_1^2 + n_2^2\right) \\ & + 4n_0 n_1^2 n_2 + \left(n_0^2 - n_1^2\right)\left(n_1^2 - n_2^2\right)\cos 2\delta_1 \Big] \end{aligned} \tag{28.61}$$

and

$$\begin{aligned} T = 8n_0 n_1^2 n_2 \Big/ \Big[& \left(n_0^2 + n_1^2\right)\left(n_1^2 + n_2^2\right) + 4n_0 n_1^2 n_2 \\ & + \left(n_0^2 - n_1^2\right)\left(n_1^2 - n_2^2\right)\cos 2\delta_1 \Big] \,. \end{aligned} \tag{28.62}$$

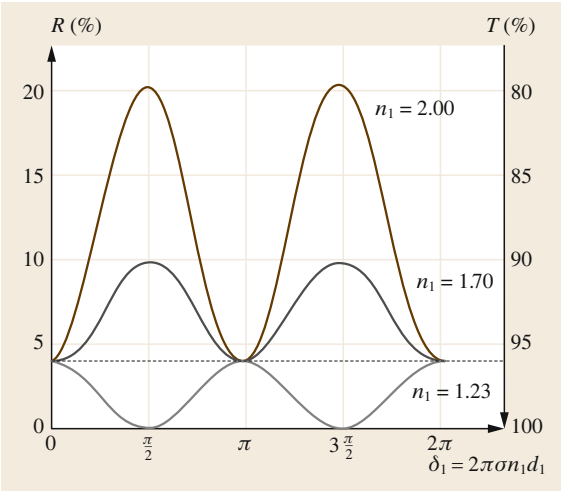


Fig. 28.24 Variation of the energy reflectance R and transmittance T with phase thickness δ_1 for a nonabsorbing film at normal incidence for various film indices n_1, with $n_0 = 1$ and $n_2 = 1.5$. The wavenumber σ is the reciprocal of the wavelength in vacuum λ. (After [28.152] with permission from Elsevier Science)

The variations of R and T with δ_1 are shown in Fig. 28.24 [28.152] for normal incidence, where the refractive index of the incident medium $n_0 = 1$ and that of the substrate $n_2 = 1.5$. These correspond to air and glass respectively. Similar plots of these variations as a function of optical thickness $n_1 d_1$ expressed as multiples of the wavelength λ are also common in the literature [28.1, 16, 150], where an optical thickness equal to the wavelength corresponds to $\delta_1 = 2\pi$. The figure shows curves for three values of the film refractive index of 2.00, 1.70 and 1.23. For the first two of these $n_0 < n_1 > n_2$ and maxima in reflectance occur at $\pi/2$ and $3\pi/2$ with minima at 0, π and 2π, while for the third $n_0 < n_1 < n_2$ and the positions of the maxima and minima are reversed. In general the maximum and minimum reflectance values $R_{\max}$ and $R_{\min}$, depend on the relative values of n_0, n_1 and n_2. For the film index greater or less than those at both its boundaries, i. e., $n_0 < n_1 > n_2$ or $n_0 > n_1 < n_2$, we obtain

$$R_{\max} = \left(\frac{n_1^2 - n_0 n_2}{n_1^2 + n_0 n_2}\right)^2 \,, \quad R_{\min} = \left(\frac{n_2 - n_0}{n_2 + n_0}\right)^2 \tag{28.63}$$

with maxima occurring when δ_1 is an odd multiple of $\pi/2$ (optical thickness $n_1 d_1$ an odd number of quarter-wavelengths) and minima when δ_1 is an even multiple of $\pi/2$ ($n_1 d_1$ an even number of quarter wavelengths). For the film index intermediate between those at its

boundaries, i. e., $n_0 < n_1 < n_2$ or $n_0 > n_1 > n_2$ these results are reversed, and we obtain

$$R_{\max} = \left(\frac{n_2 - n_0}{n_2 + n_0}\right)^2 , \quad R_{\min} = \left(\frac{n_1^2 - n_0 n_2}{n_1^2 + n_0 n_2}\right)^2 \tag{28.64}$$

with maxima occurring when δ_1 is an even multiple of $\pi/2$, and minima at odd multiples (or maxima for $n_1 d_1$ an even number of quarter wavelengths and minima an odd number). Equation (28.64) predicts that zero reflectance occurs when $n_1 = \sqrt{n_0 n_2}$, and for the refractive index values of Fig. 28.24 give $n_1 = 1.23$. Thus a simple *anti-reflection coating* can be made by using a film of this index with optical thickness $\lambda/4$. This is clearly evident from the figure at the points where δ_1 has values of $\pi/2$ and $3\pi/2$. Another point which emerges from the figure is that if the refractive index of the film is the same as that of the glass, the system behaves as if the glass is uncoated. This situation is shown by the dotted line in the figure. Here the reflectance is 4%, exactly as predicted by (28.55) for an uncoated substrate (i. e., $n_0 = 1$ and $n_1 = 1.5$).

If the electromagnetic light wave is attenuated by a film layer, that layer is said to be *absorbing*, in contrast to the *transparent* layers considered previously. Providing the incident light beam is incident *normally* on the surface of the absorbing medium, the theory described above may be applied, with the proviso that the refractive index n is replaced by a *complex refractive index* $\boldsymbol{n} = n - \mathrm{i}k$. The quantity k is called the *extinction coefficient*, and characterises absorption of energy within the film. Substitution of the complex refractive index in the appropriate expressions for the Fresnel coefficient of (28.54), and in the expression for the phase thickness of (28.56) at normal incidence allows the reflection coefficient r to be obtained. The irradiance (energy) reflectance R can be obtained by substituting the complex expressions for r and δ directly into (28.57) and evaluating rr^*. For more than two or three layers the results become intensely unwieldy, and matrix methods as mentioned below are considerably more useful. A simple example of the use of the complex refractive index will suffice. If we take the case of an incident beam in a nonabsorbing medium such as air with refractive index n_0, incident upon an absorbing medium with complex refractive index $\boldsymbol{n}_1 = n_1 - \mathrm{i}k_1$, the reflectance can be determined by replacing n_1 by $n_1 - \mathrm{i}k_1$ in (28.54) and evaluating rr^*. In this case the energy reflectance at the boundary becomes

$$R = \frac{(n_0 - n_1)^2 + k_1^2}{(n_0 + n_1)^2 + k_1^2} \tag{28.65}$$

in contrast to the result of (28.55) for a nonabsorbing medium.

In general, in absorbing films, the planes of constant phase are not parallel to the planes of constant amplitude for the electromagnetic wave disturbances. Only for the special case of normal incidence is this true, and thus in absorbing films it is usual to consider normal incidence only. The general case of oblique incidence is discussed in the standard texts on the optics of thin films [28.149, 150]. The behaviour of the energy reflectance and transmittance depend on the optical constants n and k. We have seen that in transparent films, R and T oscillate with the film thickness due to interference effects. In absorbing films with small k values, the oscillatory behaviour remains, but the amplitudes of the maxima decrease with increasing films thickness. For higher absorption values, particularly in the case of metals, intensity maxima are not observed, and an exponential decrease in transmittance with film thickness is seen.

Of particular use in the case of thin film stacks containing several layers is the matrix formalism, which allows the reflection and transmission properties to be determined. A 2×2 matrix characterises each layer of the stack, and allows the electric and magnetic field strengths to be expressed in terms of those of the adjacent layer. For transparent films this is given for the r-th layer in the form [28.153]

$$\mathbf{M}_r = \begin{pmatrix} \cos\delta_r & \frac{\mathrm{i}\sin\delta_r}{n_r} \\ \mathrm{i}n_r \sin\delta_r & \cos\delta_r \end{pmatrix} , \tag{28.66}$$

where δ_r is the phase thickness and n_r is the refractive index of the r-th layer. The electric and magnetic field intensities in the adjacent $(r-1)$-th layer, E_{r-1} and H_{r-1}, are related to those in the r-th layer E_r and H_r by

$$\begin{pmatrix} E_{r-1} \\ H_{r-1} \end{pmatrix} = \mathbf{M}_r \begin{pmatrix} E_r \\ H_r \end{pmatrix} . \tag{28.67}$$

Repeated application of (28.67) to a stack of layers enables E_0 and H_0 in the incident medium (n_0) to be expressed in terms of the field intensities in any other layer of the stack. Knowing the n values for each of the various layers of the stack, a *characteristic matrix* (also 2×2) for the entire stack is obtained by repeated matrix multiplication of the individual layer matrices, and allows E_0 and H_0 to be obtained in terms of the transmitted field intensities at the opposite side of the stack. The energy reflectance is then easily obtained. Similar matrices to those in (28.66) may also be defined for the case of absorbing films, although it is then usual to work in terms of the layer admittance η, defined as the ratio of the transverse components of the magnetic

to the electric vectors [28.150]. The matrix formalism can be extended even for optically anisotropic thin films [28.154] or magneto-optical multilayers where optical anisotropy is induced by film magnetization or external magnetic field [28.155].

A relatively simple example of an interesting system is a stack of layers of equal optical thickness, but with alternating high (H) and low (L) refractive indices [28.153]. If the optical thicknesses correspond to $\lambda/4$ (a quarter-wave stack), there is a highly reflective region at the wavelength λ, with other highly reflecting regions at 2λ, 3λ, etc. The reflectance value becomes higher the more layers there are in the stack, and approaches unity for a few layers. If the refractive index of the incident medium is n_0, that of the substrate n_S, and those of the high- and low-index regions n_H and n_L, respectively, the energy reflectance for an even number of layers $2s$ (n_L next to substrate) is given by

$$R_{2s} = \left(\frac{n_S f - n_0}{n_S f + n_0}\right)^2 \qquad (28.68)$$

where $f = (n_H/n_L)^{2s}$. For an odd number of layers $(2s+1)$, with n_H next to the substrate, the energy reflectance becomes

$$R_{2s+1} = \left(\frac{n_H^2 f - n_0 n_S}{n_H^2 f + n_0 n_S}\right)^2 . \qquad (28.69)$$

The bandwidth of the highly reflecting region depends on $\sin^{-1}[(n_H - n_L)/(n_L + n_H)]$, and thus is limited by the refractive indices available. Bandwidth can be increased by using a stack with layer thicknesses increasing in geometrical progression.

In addition to antireflection and highly reflecting coatings, various types of optical filters can be designed. Perhaps the simplest is the Fabry–Perot interference filter, which consists of two highly reflecting stacks, separated by a layer of optical thickness $\lambda/2$. This results in a very narrow transmission band at wavelength λ, with a rejection region on either side. Filter designs are usually based on computer evaluations of the optical matrices. These include band-pass, band-stop and edge filters. These are well beyond the scope of the present article and the reader is referred to the literature [28.30, 150, 156, 157]. The general trend in today's design and production of optical interference coatings is towards complex structures with refractive index profiles varying continuously along the structure. Theoretical approaches for the calculation of optical response in such inhomogeneous film stacks is reviewed for example in [28.158].

In this section some of the basic optical properties have been described, with the primary emphasis on non-absorbing films at normal incidence. Several interesting film arrangements have been described, again for non-absorbing films. Implicit to any system there is some absorption, but this can be accounted for if both the optical constants are known. Usually the design process involves the manipulation of several layer matrices, but this is handled effectively by modern computer technology. The optical properties of highly absorbing films are described by *Abelès* [28.159], who also considers inhomogeneous films. Other useful references on the theory and detailed optical properties of various aspects of thin films may be found in the literature: theory and calculations [28.160], antireflection coatings [28.161], filters [28.157, 162, 163] and mirrors [28.164].

28.3.2 Optical Absorption

The interpretation of the optical absorption behavior of thin films is one of the most beneficial methods for gaining a better understanding of the electronic structure of a material and to evaluate the energy band diagram. In 1972, *Tauc* stated a mathematical relation between the absorption tail and band gap of a material, written based on the absorption coefficient [28.165]

$$\alpha_{\text{direct}} = \frac{C}{h\nu}\left(h\nu - E_g\right)^{1/2} \quad \text{Direct band gap},$$
$$\alpha_{\text{indirect}} = \frac{C}{h\nu}\left(h\nu - E_g\right)^{2} \quad \text{Indirect band gap}, \qquad (28.70)$$

where, C is a constant, $h\nu$ is the photon energy, and E_g is the band gap energy. (The constant C is different in the above two equations.)

Direct band gap refers to the excitation in which the momentum energy of electrons is unaffected by excitation from the valence to conduction band, while in the case of indirect band gap some parts of the phonon energy are consumed to change the momentum energy of the electrons. Tauc's law is widely used to calculate the band gap of materials, ranging from bulk to quantum dot size. Further discussions on optical absorption, in particular modifications to the Tauc absorption rules above, can be found in Chap. 3.

Commercially available zinc oxide (ZnO) nanoparticles were used to produce spin-coated thin films on fluorine-doped tin oxide (FTO) coated glass slides. The UV/Vis absorbance (denoted as A and defined by $0.434\alpha d$, where d is the sample thickness) spectrum in Fig. 28.25 shows two distinct regions: (i) a sharp absorption edge due to interband transition in the incident photon energy $h\nu$ range between 3.15–3.25 eV and (ii) an absorption tail at low photon energies in the range 2.20–2.32 eV representing the localized states tailing into the band gap. The Tauc plot of $(Ah\nu)^2$ against

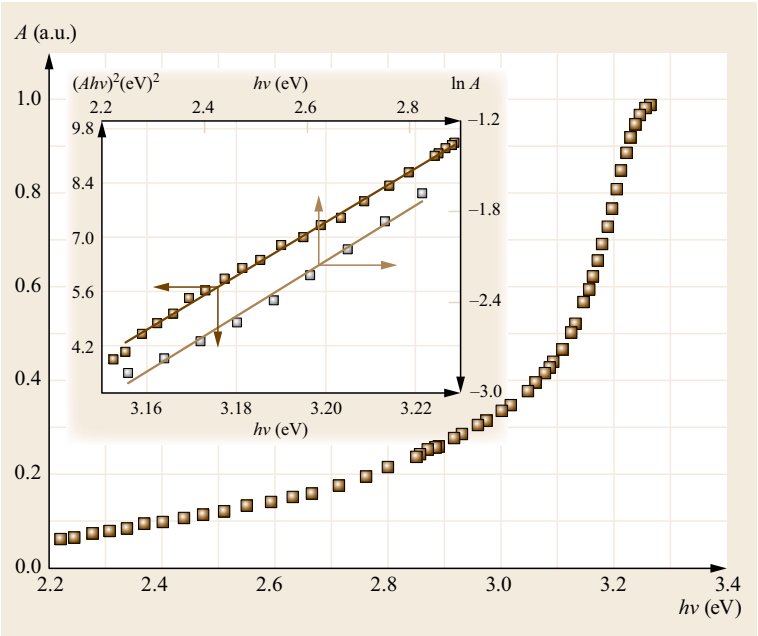


Fig. 28.25 UV-visible transmission spectra of as-deposited ZnO films on silica slides. Tauc and Urbach plots are included in the *inset*. *A* is the measured absorbance, and is plotted along the *y*-axis in arbitrary units

the incident energy $h\nu$ in the inset is linear, indicating the transition is direct for the sample. The intercept on the ordinate extrapolated to $h\nu = 0$ produced values of 3.15 eV for the direct optical band gap of the films. The linear plot of A against $h\nu$ between 2.20 and 2.32 eV in the second inset on a logarithmic linear scale represents the presence of the Urbach law in the following form

$$\alpha(h\nu) \propto \exp\left(\frac{h\nu - E_{\mathrm{CE}}}{E_{\mathrm{U}}}\right). \qquad (28.71)$$

The converging energy E_{CE} is taken to be higher than the optical band gap energy E_{g}. The Urbach energy E_{U}, which is related to the width of band tails of localized states extending into the gap, was found to be 0.53 meV from the inverse of the slope of the linear fit.

28.3.3 Electrical Properties

Chapter 2 in this Handbook provides an introduction to the electrical properties of materials and includes a section on the resistivity of thin films.

The electrical properties of thin films are one of their most important physical features, and differ from those of the corresponding bulk material for various reasons. Electrical properties cover a particularly wide field, and in this section only two important examples are covered. The first of these is the lateral conductivity (or corresponding resistivity) of thin metallic films. In metals we are usually concerned only with one type of charge carrier, electrons, and thus variations in conductivity σ due to the presence of both carrier types (such as compensation effects) are not considered. The main cause of deviations from the bulk material conductivity values is curtailment of the bulk electron mean free path λ_0 by the dimensions of the film. Clearly, if the film thickness $d < \lambda_0$, many instances of electron scattering occur at the film surfaces, thus reducing the effective electron mean free path and decreasing the conductivity. Modern theories also take into account electron scattering by individual grain boundaries as well as by the film surfaces. The second example of electrical properties in thin films is that of high-field conduction in insulators and semiconductors. Electric fields in thin films may approach the dielectric breakdown strength, and under these conditions the usual ohmic behaviour is not followed, but a range of different limiting conduction processes may take place. Conduction in semiconductors at somewhat lower electric fields is not discussed here, but is covered in detail elsewhere in the literature [28.1, 15, 18, 166]. Semiconductor conductivity is a broad subject, underlying the operation of many different electronic devices. We might note that the conductivity of semiconductors in their surface regions (and therefore also in semiconductor thin films) differs from that of the bulk material owing to depletion or accumulation regions at the surfaces, and that the application of a surface electric field forms the basis of several semiconductor devices, such as metal–oxide–semiconductor (MOS) transistors. Furthermore, the mobility of carriers in semiconductors frequently follows similar variations to those of the

conductivity in metallic films, for essentially the same regions, in that the surface and/or grain boundary scattering significantly influence the mobility and thus the conductivity.

Electrical Conductivity in Metallic Films

Chapter 2 included a basic introduction to the electrical resistivity of metallic thin films. In this chapter, we consider electrical conduction in thin films in more detail, and provide further useful references to both metallic and semiconducting thin films. The electrical conductivity in bulk metals σ_0 is given generally by an expression of the form

$$\sigma_0 = \frac{ne^2\lambda_0}{mv}\,, \tag{28.72}$$

where n is the concentration of free electrons, e is the electronic charge, λ_0 the bulk electron mean free path, m the electron effective mass and v the electron Fermi velocity. This expression results from the free-electron gas theory of metals, derived from the work of Drude, Lorentz and Sommerfield. The main feature of this result is that the electrical conductivity is proportional to the mean free path, and that if the mean free path is curtailed by any means, then the conductivity will be reduced, all other considerations remaining unaltered. A further consideration is the well-known result termed Matthiessen's rule,

$$\rho_0 = \rho_\mathrm{d} + \rho_\mathrm{ph}(T)\,, \tag{28.73}$$

where the bulk resistivity $\rho_0\,(=1/\sigma_0)$ is given by the sum of contributions from static defects and impurities ρ_d and from phonons $\rho_\mathrm{ph}(T)$. The latter contribution effectively isolates the main temperature-dependent effect from other scattering mechanisms, unless they too are temperature-dependent. Temperature-dependent phonon interactions increase with increasing temperature, giving rise to the positive temperature coefficient of resistivity (TCR) in metals. At very low temperatures $\rho_\mathrm{ph}(T) \to 0$, and only the residual resistivity ρ_d remains. For the case of a thin metal film, scattering at the surfaces also comes into play, and Matthiesen's rule is modified by the addition of an extra surface-scattering resistivity term ρ_s to give

$$\rho_\mathrm{f} = \rho_\mathrm{d} + \rho_\mathrm{ph}(T) + \rho_\mathrm{s} \tag{28.74}$$

for the thin-film resistivity ρ_f. For very thick films $\rho_\mathrm{s} \approx 0$, since the mean free path is not then limited by the surfaces, and the film resistivity is effectively the same as the bulk value. When ρ_s is significant for thinner films, $\rho_\mathrm{f} > \rho_0$, and the film conductivity $\sigma_\mathrm{f} < \sigma_0$. The above discussions follow naturally from the treatise in Chap. 2.

Over a century ago, *Thomson* [28.167] was the first to suggest that observed decreases in the conductivity of thin films as compared to the bulk values are the result of the limitation of the electron mean free path by the specimen size. He made the following assumption in obtaining an expression for the film conductivity: the bulk mean free path in the metal is a constant λ_0, greater than the film thickness d, and that when electrons collide with the internal film surfaces the scattering is independent of the angle at which they strike, i. e., the scattering is diffuse. Collisions occurring in the film were split into two regions, in the first the mean free path is determined by the surfaces (top and bottom) and in the second the mean free path is not limited by scattering at the surfaces. In terms of the parameter

$$k = \frac{d}{\lambda_0}\,, \tag{28.75}$$

the ratio of the film thickness to the bulk mean free path, the ratio of the film to bulk conductivity was determined as

$$\frac{\sigma_\mathrm{f}}{\sigma_0} = \frac{\rho_0}{\rho_\mathrm{f}} = \frac{k}{2}\left[\ln\left(\frac{1}{k}\right) + \frac{3}{2}\right]. \tag{28.76}$$

Fuchs [28.168] pointed out that this derivation neglected mean free paths starting at the film surfaces, and also the statistical distribution of the mean free paths about λ_0 in the bulk metal. A corrected version of Thomson's model for very thin films, i. e., $d \ll \lambda_0$ $(k \ll 1)$, where effectively all the mean free paths start from one of the surfaces, yields an alternative result [28.169]

$$\frac{\sigma_\mathrm{f}}{\sigma_0} = \frac{\rho_0}{\rho_\mathrm{f}} = k\left[\ln\left(\frac{1}{k}\right) + 1\right]. \tag{28.77}$$

Although both expressions predict a decrease in conductivity below the bulk value for films, neither gives a sensible prediction for very thick films, i. e., $d \gg \lambda_0$ $(k \gg 1)$.

A more useful general model to account for variations of conductivity with film thickness was introduced by *Fuchs* [28.168], who formulated the problem in terms of the Boltzmann transport equation, which describes the way in which the electron energy changes due to applied fields and collisions. Readers are referred to the original paper, or to one of the reviews mentioned at the end of this section, for details of the mathematical

arguments. The final result of this model is

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} = 1 + \frac{3}{4}\left(k - \frac{k^3}{12}\right) B(k) - \frac{3}{8k}\left(1 - e^{-k}\right) - \left(\frac{5}{8} + \frac{k}{16} + \frac{k^2}{16}\right) e^{-k}, \tag{28.78}$$

where $B(k)$ is a representative of the set of exponential integral functions, given by

$$B(x) = -\mathrm{Ei}(-x) = \int_x^\infty \frac{e^{-y}}{y}\mathrm{d}y\,. \tag{28.79}$$

Figure 28.26 shows the dependence of ρ_f/ρ_0 $(= \sigma_0/\sigma_f)$ on k over the range 0.1–10 for the curve labelled $p = 0$ (diffuse scattering). Here it is clear that the resistivity drops to the bulk value for thick films. Various approximations for (28.78) are frequently quoted to aid computation in specific thickness ranges. For thick films ($k \gg 1$)

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx 1 - \frac{3}{8k}\,. \tag{28.80}$$

This clearly shows that as $k \to \infty$, $\sigma_f \to \sigma_0$. For thinner films ($k \ll 1$), *Campbell* [28.170] quotes

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4}\left[\ln\left(\frac{1}{k}\right) + 0.4228\right] \tag{28.81}$$

and for *very* thin films *Sondheimer* [28.171] gives

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4}\left[\ln\left(\frac{1}{k}\right)\right]. \tag{28.82}$$

Equations (28.76), (28.77), (28.81) and (28.82) give some confidence in the general approach, since they are all of comparable form, with the right-hand sides each consisting of a constant $\times\, k \times$ [ln(1/k) + another constant].

Fuchs also considered the case where the electron scattering at the surfaces was not totally randomised, and that a constant proportion of electrons p are scattered specularly. p is termed the *specularity parameter*, where $p = 0$ for the case of totally diffuse scattering and $p = 1$ for totally specular scattering. Clearly in the latter case the film surfaces do not affect the conductivity, and σ_f reduces to the bulk value. Figure 28.26 also shows the variation of ρ_f/ρ_0 for $p = 0.5$ (intermediate between diffuse and specular scattering) and $p = 1$ (totally specular scattering). Again for the case of some specular scattering there are limiting expressions for thick films ($k \gg 1$) and thin films ($k \ll 1$), given respectively by

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx 1 - \frac{3(1-p)}{8k} \tag{28.83}$$

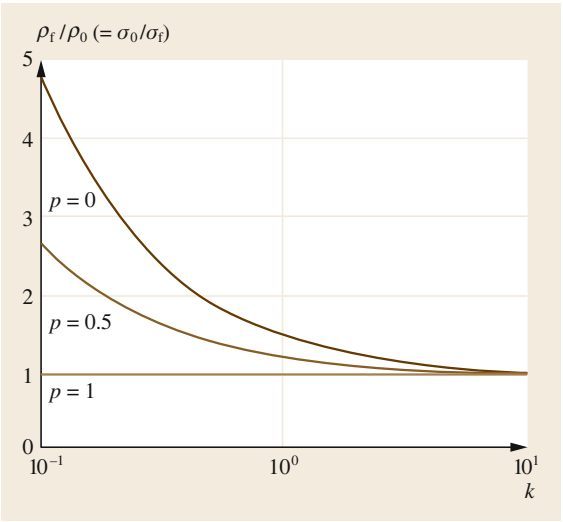


Fig. 28.26 Dependence of the ratio of film to bulk resistivity ρ_f/ρ_0 on the ratio of film thickness to bulk mean free path $k = d/\lambda_0$ predicted by the model of *Fuchs* [28.168]; p is the specularity parameter, with $p = 0$ representing totally diffuse scattering and $p = 1$ representing totally specular scattering, where for the latter $\rho_f/\rho_0 = 1$. The *curve* labelled $p = 0.5$ is intermediate between these two values

and

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4}(1+2p)\left[\ln\left(\frac{1}{k}\right) + 0.4228\right], \tag{28.84}$$

where p is small in (28.84). For totally diffuse scattering, where $p = 0$, these two expressions reduce to the previous results of (28.80) and (28.81). An alternative expression for thin films is also given by *Sondheimer* [28.171],

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4}\left(\frac{1+p}{1-p}\right)\ln\left(\frac{1}{k}\right). \tag{28.85}$$

This reduces to (28.82) for $p = 0$, and for small k and p is consistent with (28.84).

Both *Lucas* [28.172] and *Juretschke* [28.173, 174] have extended the specularity analysis by assuming a different specularity parameter for each side of the film, i. e., p and q. This is of course quite reasonable, in that the specularity parameter of the lower (substrate) surface is likely to be heavily influenced by the substrate surface. The thick- ($k \gg 1$) and thin-film ($k \ll 1$) approximations in this case are then

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx 1 - \frac{3}{8k}\left[1 - \left(\frac{p+q}{2}\right)\right] \tag{28.86}$$

and

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4}\frac{(1+p)(1+q)}{(1-pq)} \ln\left(\frac{1}{k}\right), \qquad (28.87)$$

respectively. For identical surfaces with $p = q$, these expressions reduce to (28.83) and (28.85) respectively. Tabulated values of ρ_f/ρ_0 are given by *Chopra* [28.18] as functions of k for various combinations of p and q. The whole topic of the size-dependent electrical conductivity in thin metallic films and wires is thoroughly reviewed by *Larson* [28.175], but omitting the important work on grain-boundary scattering of *Mayadas* and *Shatzkes* [28.176]. This is, however, covered in a more recent book reviewing size effects in thin films generally [28.177], which also includes useful tabulated conductivity data.

A recent clear example in which surface scattering controls the effective resistivity of a thin film is shown in Fig. 28.27 where the authors have examined the resistivity of single crystal layers of Cu(100) grown epitaxially on insulating crystals TiN(001) and MgO(001) at substrate temperatures of 25 and 80 °C [28.178]. The measurements were made at two temperatures, 298 K (in situ in vacuum) and 77 K in liquid nitrogen; as expected at 77 K the resistivity is much lower. The solid

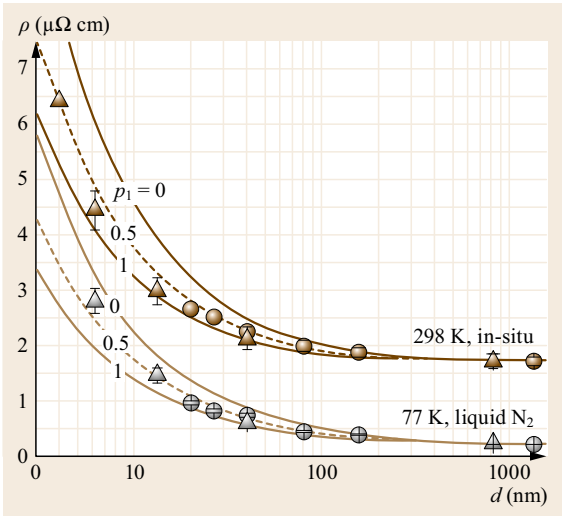


Fig. 28.27 The resistivity ρ measured at 298 K (in situ in vacuum) and 77 K versus layer thickness d of epitaxial Cu(001) layers grown on 2.5 nm TiN(001)/MgO(001) at a substrate temperature of 25 °C and MgO(001) at 80 °C. The *lines* represent the theoretical calculations based on the Fuchs–Sondheimer model at 298 and 77 K for $p_2 = 0$ and $p_1 = 0$, 0.5, and 1, as labeled; p_1 and p_2 are the upper surface and Cu–substrate interface secularity parameters. The *circles* are Cu on TiN and *triangles* are Cu on MgO. (Data extracted from [28.178])

curves are a theoretical prediction based on the Fuchs–Sondheimer formalism. The top and bottom surfaces of the film are assigned p_1 and p_2 for the specularity parameter with $p_2 = 0$ (Cu surface on the substrate) and $p_1 = 0$, 0.5, and 1. It is clear that within experimental error, it is possible to describe the resistivity within the Fuch–Sondheimer model. The distinct advantage of using an epitaxial thin film layer is that there are no grain boundaries within the film that would introduce an additional scattering process and thereby increase the resistivity.

As mentioned above, an important development in size-effect models came about with the inclusion of grain-boundary scattering [28.176]. In a bulk sample, the individual grains are normally considerably larger than the bulk mean free path λ_0, and do not therefore affect the conductivity. However, in thin films, the grain sizes and orientations depend on the preparation method and conditions. Typical grain sizes are of the same order as λ_0 and therefore scattering at their boundaries will influence the conductivity. Mathematically, the model takes account of the fact that grains tend to grow in a columnar structure, with the axes normal to the film plane. Single grains usually extend between both film surfaces, and thus their upper and lower boundaries do not affect the conductivity. Mayadas and Shatzkes argue that only grain boundaries normal to the film plane influence the conductivity. The grain boundaries are represented by δ-function potentials, and are used in the model in the solution of the Boltzmann transport equation, subject to suitable boundary conditions. The model is developed in terms of a dimensionless parameter α, given by

$$\alpha = \frac{\lambda_0}{D}\frac{R}{1-R}, \qquad (28.88)$$

where, as before λ_0 is the bulk mean free path, D is the mean grain dimension, and R is the grain-boundary electron reflection coefficient. The ratio of the grain-boundary conductivity σ_g (and the corresponding resistivity ρ_g) to the bulk values are then given by

$$\frac{\sigma_g}{\sigma_0} = \frac{\rho_0}{\rho_g} = 3\left[\frac{1}{3} - \frac{1}{2}\alpha + \alpha^2 - \alpha^3 \ln\left(1 + \frac{1}{\alpha}\right)\right]. \qquad (28.89)$$

This is easy to evaluate, and the dependence of ρ_g/ρ_0 $(= \sigma_0/\sigma_g)$ on α is shown in Fig. 28.28. For small values of α (i. e., large grain size D, small bulk mean free path λ_0, or small reflection coefficient R) the grain boundaries do not affect the overall conductivity, which remains at the bulk value. However, for large α the

grain-boundary resistivity increases (conductivity decreases). For very large ($\alpha \gg 1$) and very small ($\alpha \ll 1$) values of α, (28.89) has the limiting approximations

$$\frac{\sigma_0}{\sigma_g} = \frac{\rho_g}{\rho_0} \approx \frac{4}{3}\alpha \qquad (28.90)$$

and

$$\frac{\sigma_0}{\sigma_g} = \frac{\rho_g}{\rho_0} \approx 1 + \frac{3}{2}\alpha\,, \qquad (28.91)$$

respectively. Note that following the original work [28.176] the latter two expressions are quoted as the ratio ρ_g/ρ_0 $(= \sigma_0/\sigma_g)$, i. e., the reciprocal of the ratio in (28.89).

Mayadas and *Shatzkes* [28.176] also considered the case of grain-boundary scattering in conjunction with scattering at the film surfaces, and followed the *Fuchs* [28.168] formulation to arrive at an expression for the overall film resistivity ρ_f. Unfortunately, the resulting expression could not be evaluated analytically, requiring numerical solution. Detailed comparisons between the basic Fuchs model with $p = 0$ and $p = 0.5$, and their model using the same values of p and including grain-boundary scattering with $\alpha = 1$, led these workers to interpret their results as showing an enhancement of the intrinsic resistivity ρ_g in polycrystalline samples in which grain size is constant and independent of film thickness. However, discrepancies between the Fuchs and Mayadas–Shatzkes models become most apparent for very large α (very small grain size, D). For this situation the deviation corresponds to the additional resistivity arising from grain boundaries rather than from surface scattering [28.179, 180]. Useful tabulated data of resistivity ratios in samples with both grain-boundary and surface scattering are given elsewhere in the literature [28.177, 181].

The above description of thin film resistivity neglects the roughness of the film surfaces, which can increase the resistivity further. The surfaces are simply assigned a specularity parameter p depending on the nature of scattering at the surface. *Namba* [28.182] considered the resistivity of thin films with surface roughness and derived an expression for the resistivity in terms of the film thickness and the ratio of the roughness h to the mean free path length λ_0 in the bulk. The film was assumed to have a certain mean thickness d and a sinusoidal variation of the top surface that follows $h\sin(kx)$, so that the total thickness was modulated as $d + h\sin(kx)$ where x is the distance parallel to the current flow and k is the spatial frequency of the sinusoidal variation or $2\pi/\Lambda$ where Λ is the periodicity of roughness, i. e., a roughness wavelength. (Roughness variations are implicitly assumed to be on the mesoscopic scale.) Grain boundary scattering was neglected. Figure 28.29 shows the results of his calculations for the normalized resistivity ρ/ρ_0 in terms of the film normalized thickness d/λ_0 for different values of h/λ_0. The case for $h/\lambda_0 = 0$ reduces to the Fuchs–Sondheimer formalism, as expected, and the description is often referred to as the Fuchs–Namba model. There have been several discussions on the effect of surface roughness on measured resistivity, and potential improvements to Namba's approach, see for example, the discussions by

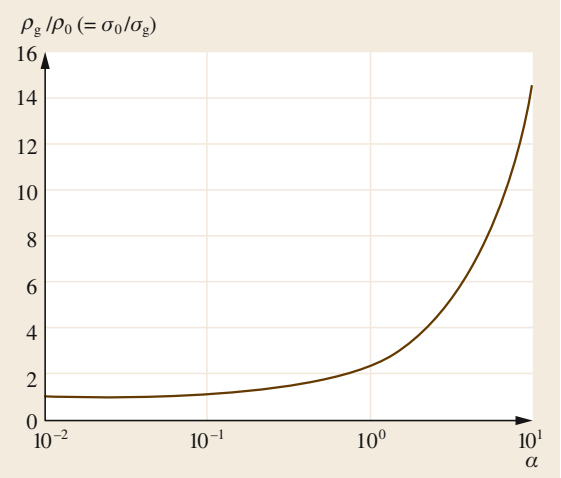


Fig. 28.28 Dependence of the ratio of grain-boundary to bulk resistivity ρ_g/ρ_0 on the parameter α from the model of *Mayadas* and *Shatzkes* [28.176]. The data are calculated from (28.89)

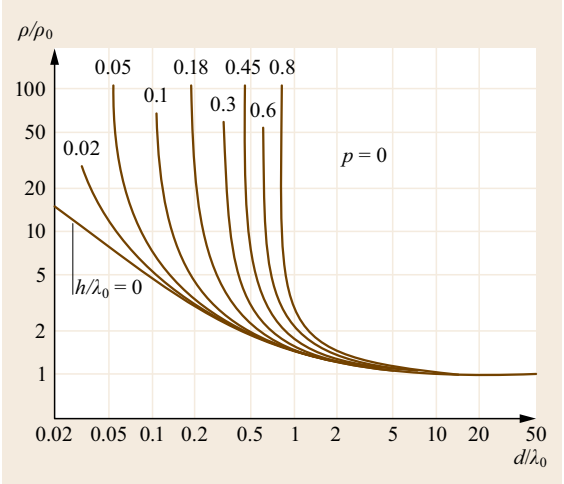


Fig. 28.29 Normalized resistivity (ρ/ρ_0) versus normalized film thickness d/λ_0 for various values of normalized surface roughness h/λ_0. (Data from [28.182])

Trivedi and *Ashcroft* [28.183] and *Jacob* et al. [28.184]. There are several examples in the literature in which good agreements have been reported between the thickness dependence of the thin film resistivity and the Fuchs–Namba model as in references [28.185, 186].

Finally, we briefly discuss the thickness variations of the temperature coefficient of resistivity. This is a measure of the response of the resistivity to changes in temperature, and is usually denoted by the symbol α. To avoid confusion with the Mayadas–Shatzkes parameter, here we denote it by α_T. The temperature coefficient of resistivity is given for the bulk (α_{T0}) and for thin films (α_{Tf}) by

$$\alpha_{T0} = \frac{1}{\rho_0}\frac{d\rho_0}{dT}; \quad \alpha_{Tf} = \frac{1}{\rho_f}\frac{d\rho_f}{dT} \tag{28.92}$$

where ρ_0 and ρ_f are the bulk and thin-film resistivities, as given in (28.73) and (28.74). From the above relationships, and Matthiesen's rule, it is possible to derive expressions for α_{Tf}/α_{T0} in a similar manner as ρ_f/ρ_0. Referring to (28.74), ρ_d is essentially temperature-independent, while the temperature dependence of ρ_s depends on the film thickness. For thicker films ($k \gg 1$), ρ_s is also temperature-independent, and thus

$$\frac{d\rho_f}{dT} = \frac{d\rho_{ph}}{dT} \tag{28.93}$$

Similarly, from (28.73) for the bulk material

$$\frac{d\rho_0}{dT} = \frac{d\rho_{ph}}{dT} \tag{28.94}$$

and therefore both the differentials in (28.92) are equal, and we obtain

$$\alpha_{T0}\rho_0 = \alpha_{Tf}\rho_f \,. \tag{28.95}$$

This useful result implies that

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} = \frac{\alpha_{Tf}}{\alpha_{T0}} \tag{28.96}$$

and that for the case of thicker films the ratio of the film-to-bulk temperature coefficient of resistivity is equal to the ratio of film-to-bulk conductivity. For the case of thinner films ($k \ll 1$), the situation is complicated by the fact that ρ_s is also temperature-dependent, because the mean free path decreases with increasing temperature and becomes significant for small k. Various calculated data for α_{Tf} are given elsewhere in the literature [28.6, 18, 170, 175, 177], to which the reader is referred for a fuller discussion. Discussions concerning the applicability of the various models described above to several types of metallic film are also given in some of these references, and a wider coverage including additional contributions to the size effect models, which cannot be included in the present work, is given in the book by *Tellier* and *Tosser* [28.177].

High-Field Conduction in Insulating and Semiconducting Thin Films

Here electrical conductivity is considered for the case where a voltage is applied across an insulating or semiconducting thin film prepared in a sandwich configuration between metallic electrodes. This situation is in direct contrast to that described in the previous section, in that the intrinsic film conductivity is significantly lower, and the electric field applied can be extremely high, even when only low voltages are applied. The latter point is clearly illustrated by the fact that in a film of thickness 1 μm subjected to a voltage of only 10 V, the mean electric field established across the film is 10^7 V/m, approaching the dielectric breakdown strength of many materials. Several high-field conduction mechanisms that may in principle occur in thin dielectric films have been thoroughly reviewed in the literature [28.187]. These have also been observed in relatively wide-band-gap semiconducting films. These processes are discussed later in this section, together with the appropriate current density–voltage expressions which aid in their identification and analysis. First, however, it is useful to discuss the types of electrical contact which may be applied to dielectric films, since the dominant conduction process is frequently determined by the type of electrical contact and its interfacial properties.

Simmons [28.187] has demonstrated that the type of electrode used significantly influences the conduction processes which may be observed; the various conduction mechanisms were classified as either *bulk-limited* or *electrode-limited*. For the former the charge carriers are generated in the bulk of the material and the electrodes merely serve to apply a potential which generates a drift current. In the latter, at the interface between the electrode and the insulator or semiconductor, a potential barrier to the flow of charge is established, which limits the current. Contacts themselves are said to be either *ohmic* or *blocking*, with an intermediate *neutral* contact. Blocking contacts are conventionally known as Schottky barriers. In principle, the type of contact formed depends on the relative work functions of the insulator or semiconductor ψ_i and that of the metal contact ψ_m. Excellent energy-band diagrams of the various types of contact are given in the literature [28.187], and are reproduced elsewhere in descriptions of contacts to cadmium compound films [28.38] and to nanostructures [28.11]. In equilibrium the vacuum and the Fermi

levels are continuous across the interface, and with no applied voltage the Fermi level also must be flat, in order that no current flows. Away from the interface in the bulk of the material, the energy difference between the vacuum and Fermi levels will be equal to the work function ψ_i, and it can easily be seen that the potential barrier height at the interface φ_0 is given by

$$\varphi_0 = \psi_m - \chi \,, \tag{28.97}$$

where χ is the film electron affinity and ψ_m is the metal work function. However, it is the *type* of contact rather than its precise barrier height which determines the type of conductivity observed. We consider primarily electrons as the charge carriers, as it has been argued that for insulators conduction by holes may normally be ignored, owing to relatively low mobilities and to immobilisation by trapping effects. Similar arguments are also applicable in the case of wide-band-gap semiconductors, and thus for both insulators and some semiconductors n-type conductivity is usual. For p-type conductivity, the conditions concerning the relative values of the work functions are reversed [28.188].

Ohmic contacts occur for electron injection when $\psi_m < \psi_i$. Electrons are injected from the electrode into the insulator or semiconductor in order to comply with thermal equilibrium requirements. They are subsequently located within the insulator or semiconductor conduction band and penetrate a distance λ_0 below the interface, forming a region of negative space charge which is termed an *accumulation region*. Upwards band bending of the conduction band occurs in the accumulation region, beyond which the dielectric is shielded from the electrode. The accumulation region acts as a charge reservoir, and can supply electrons to the material as required by the bias conditions. It has been pointed out by *Simmons* [28.187] that the conductivity is limited by the rate at which electrons can flow into the bulk of the material, and is therefore bulk-limited. Conversely, blocking contacts or Schottky barriers occur when $\psi_m > \psi_i$. In this case the initial flow of charge is reversed, with electrons flowing from the insulator or semiconductor to the metal electrode. In contrast to the case for ohmic contacts, this results in a positive space-charge region near the interface in the insulator or semiconductor, which, owing to the absence of electrons, is termed a *depletion region*. An equal and opposite negative charge is induced on the electrode, and the interaction between the two charges establishes a local electric field near the interface. In this case the conduction band in the insulator or semiconductor bends downwards, and the free-electron concentration in the interfacial region is much lower than in the bulk of the film material. The rate of flow of electrons is thus limited by the rate at which they can flow over the interfacial barrier, and the conductivity is electrode-limited.

Although, in principle, the type of contact established depends on the relative values of ψ_m and ψ_i, in reality such considerations are usually inadequate in predicting the type of contact formed, because of the existence of *surface states*. These are due either to the effects of unsaturated bonds and impurities at the interface [28.189] or to the departure from periodicity in the structure at the interface. It is very difficult to accurately predict a priori the presence and concentration of surface states. If a significant concentration of surface states does exist, (28.97) is inapplicable and the type of contact and the conduction process are determined almost entirely by their presence.

In the remainder of this section the bulk-limited conduction processes are first considered, where the conductivity does not depend on the barrier height at the interface. Following this the electrode-limited processes are considered, where the barrier height becomes a significant feature in the current density–voltage (J–V) equations.

A particularly important conduction process in insulating and semiconducting thin films is that of space-charge-limited conductivity (SCLC) (the same acronym is also used for space charge limited current). For SCLC to occur there is a requirement that the injecting electrode is an ohmic contact. Thus there is a charge reservoir available in the accumulation region, and electrons do not need to be excited over a potential barrier. At low voltages the thermally generated carrier concentration exceeds the injected concentration, and the current density J is proportional to the applied voltage V, following Ohm's law,

$$J = n_0 e \mu \frac{V}{d} \,, \tag{28.98}$$

where n_0 is the thermally generated carrier concentration (electrons), e is the electronic charge, μ the mobility and d the film thickness. When the injected electron concentration exceeds that of the thermally generated concentration the SCLC current becomes dominant. SCLC current density–voltage relationships are derived by taking into account both the drift and diffusion currents, and solving Poisson's equation. This results in an expression for the electric field as a function of distance from the interface. Charge-carrier traps normally exist within the insulating or semiconducting film, and these are effective in mopping up or immobilising most of the injected electrons. For shallow traps located at a discrete energy E_t below the conduction band edge the SCLC current density is given

by [28.187]

$$J = \frac{9}{8}\varepsilon_r\varepsilon_0\theta\mu\frac{V^2}{d^3}\,, \tag{28.99}$$

where ε_r is the relative permittivity of the material, ε_0 the permittivity of free space and θ is the ratio of free to trapped charge. This expression predicts that $J \propto V^2$ and $J \propto d^{-3}$, and the search for these relationships is an important tool in identifying SCLC dominated by a discrete trap level. For perfect materials which do not contain traps, θ is replaced by unity in (28.99). The ratio of free to trapped charge is given by [28.190]

$$\theta = \frac{N_c}{N_{t(s)}}\exp\left(-\frac{E_t}{k_BT}\right), \tag{28.100}$$

where N_c is the effective density of states at the conduction band edge, $N_{t(s)}$ is the shallow trap concentration located at the discrete energy level, k_B is Boltzmann's constant and T is the absolute temperature. A transition between ohmic conductivity and SCLC takes place when the injected carrier concentration first exceeds the thermally generated carrier concentration. An expression for the transition voltage V_t at which this occurs is obtained by simultaneously solving (28.98) and (28.99) to yield

$$V_t = \frac{8en_0d^2}{9\theta\varepsilon_r\varepsilon_0}\,. \tag{28.101}$$

SCLC of this type, showing square-law J–V behaviour, has been observed in many types of thin films. In SbS_3 dielectric films, with bismuth oxide ohmic contacts, the current was at first ohmic, before a transition to SCLC [28.191]. Current density was also found to be proportional to $1/d^3$ at constant voltage, and the transition voltage V_t showed the characteristic linear variation with d^2. It has also been observed in CdS films [28.40, 192, 193], particularly with In injecting electrodes, and a discrete-trap concentration of $2.5\times10^{20}\,m^{-3}$ was deduced [28.40]. In CdTe, a square-law SCLC behaviour was recorded in electro-deposited Ni-CdTe-Au structures [28.194]. The work also showed a systematic increase in current density with temperature over the range 273–338 K, together with the expected $J \propto 1/d^3$ behaviour. *Dharmadhikari* [28.195] observed similar behaviour in evaporated CdTe films in the lower voltage range, although power-law behaviour became apparent at higher voltages. CdSe films have also been reported as showing square-law SCLC [28.196, 197]. In organic films, the phthalocyanines are well know for exhibiting SCLC behaviour [28.120]. Square-law behaviour with discrete trap levels has been seen in several of these materials, with typical activation energies of 0.77 eV [28.198] and 0.62 eV [28.199] in copper phthalocyanine, 0.42 eV in lead phthalocyanine [28.200], 0.58 eV in nickel phthalocyanine [28.201] and 0.45 and 0.69 eV in cobalt phthalocyanine [28.202]. These values appear to vary according to the central metal atomic species in the molecule, and also with the oxygen content. In general these materials are p-type, having a trap concentration for discrete levels in the range 5×10^{16}–$7.1\times10^{25}\,m^{-3}$.

For amorphous and polycrystalline materials, where a single shallow trap level is unlikely to be present, several poorly defined trap levels can lead to an overall exponential distribution of traps $N(E)$ [28.190] given by

$$N(E) = N_0\exp\left(-\frac{E}{k_BT_t}\right), \tag{28.102}$$

where $N(E)$ is the trap concentration per unit energy range at an energy E below the conduction band edge, N_0 is the value of $N(E)$ at the conduction band edge and $T_t > T$ is a temperature parameter which characterises the distribution. Furthermore, it can be shown that the total concentration of traps comprising the distribution $N_{t(e)}$ is [28.203]

$$N_{t(e)} = N_0k_BT_t\,. \tag{28.103}$$

Lampert [28.204] obtained an expression for the current density by assuming an exponential trap distribution as described by (28.102) to yield

$$J = e\mu N_c\left(\frac{\varepsilon_r\varepsilon_0}{eN_0k_BT_t}\right)^l\frac{V^{l+1}}{d^{2l+1}}\,. \tag{28.104}$$

This expression predicts a power-law dependence of J on V, with the exponent $n = l+1$, where l represents the ratio T_t/T. J is also expected to be proportional to $d^{-(2l+1)}$. Similarly to the case for a single shallow trap level, there is also a transition voltage between ohmic conduction and SCLC dominated by an exponential trap distribution. The transition voltage in this case is given by [28.205]

$$V_t = \left(\frac{n_0}{N_c}\right)^{1/l}\frac{eN_0k_BT_td^2}{\varepsilon_r\varepsilon_0}\,. \tag{28.105}$$

Many instances of power-law SCLC are also reported in the literature. *Zuleeg* [28.192] observed a cubic power-law dependence at higher voltages in his CdS films, which suggested that an exponential trap distribution was present. *Dharmadhikari* [28.195] appears to have been the first to observe a power-law J–V dependence in CdTe films at higher voltage levels, in addition to the square-law dependence at lower voltages. Very

high exponents were recorded, although at the time these was not analysed using the exponential trap distribution expressions given above. This connection was made in a later series of measurements [28.206], where the intermediate square-law behaviour was absent. This J–V dependence is shown in Fig. 28.30 [28.206] for three different film thicknesses, and is typical of many results on various different materials showing an exponential trap distribution. Similar exponential trap distributions have also been identified in some CdSe films [28.207–209], notwithstanding the earlier results which indicated discrete trap levels [28.196, 197]. The differences between these two classes of results appear to depend on the different deposition conditions and processing used. Exponential trap distributions are extremely prevalent in phthalocyanine films, and were originally investigated by *Sussman* [28.210] for copper phthalocyanine films with ohmic Au electrodes. The material was shown to be p-type, and exhibited exponent values n in the J–V characteristics in the range 2.6–4.0. The predicted thickness dependence for this type of conductivity was also observed. Many other phthalocyanine films have subsequently shown this type of conductivity, and data on these are tabulated and compared in the literature [28.120]. Total trap concentrations for the exponential distributions are in the range 6×10^{20}–$9.3\times10^{26}\,\mathrm{m}^{-3}$.

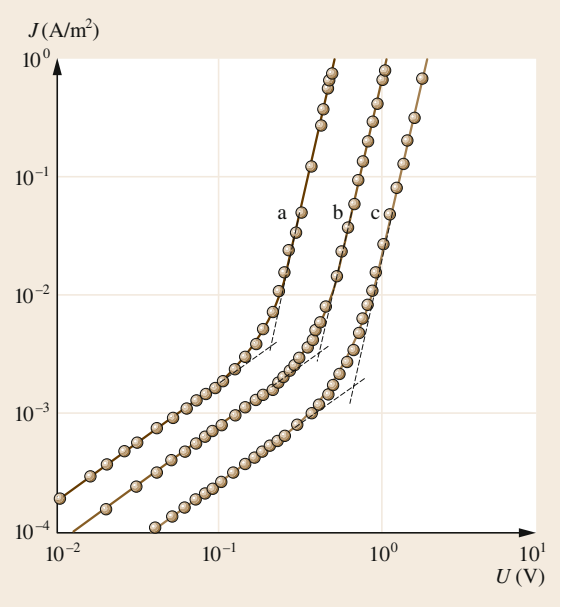


Fig. 28.30 Room-temperature current density–voltage characteristics for evaporated Al-CdTe-Al sandwich structures of thickness 0.31 μm (*a*), 0.45 μm (*b*) and 0.56 μm (*c*), showing ohmic conduction and SCLC. U is the applied voltage. (Data extracted from [28.206])

Using the above Eqs. (28.102)–(28.104) as a starting point, it was shown by the present author [28.211] that for SCLC dominated by an exponential trap distribution, measurements of J as a function of temperature at constant applied voltage in the SCLC region are sufficient to determine both the mobility and the trap concentration. If the data are plotted in the form $\log_{10} J$ against $1/T$ the curves should be linear, and when extrapolated to negative values of $1/T$, should all intersect at a common point irrespective of the applied voltage. The coordinates of this point are given by

$$\log_{10} J = \log_{10}\left(\frac{e^2\mu d N_c N_{t(e)}}{\varepsilon_r\varepsilon_0}\right); \quad \frac{1}{T} = -\frac{1}{T_t}\,. \tag{28.106}$$

Other useful results enabling data to be extracted from these plots are that for the gradient, given by

$$\frac{\mathrm{d}(\log_{10} J)}{\mathrm{d}(1/T)} = T_t \log_{10}\left(\frac{\varepsilon_r\varepsilon_0 V}{e d^2 N_{t(e)}}\right) \tag{28.107}$$

and for the intercept $\log_{10} J_0$ on the $\log_{10} J$ axis, which is

$$\log_{10} J_0 = \log_{10}\left(\frac{e\mu N_c V}{d}\right) \tag{28.108}$$

This set of equations has been used to determine mobility and trap concentration in several different materials showing this form of SCLC, including copper phthalocyanine, where these quantities were measured as functions of the ratio of evaporation background pressure to deposition rate [28.212]. In another example the method was used to determine these quantities in semiconducting CdTe films [28.213].

We might mention that *Rose* [28.190] also explored the possibility of a uniform trap distribution, in addition to discrete energy levels and the exponential distribution. In this case $J \propto V\exp(tV)$, where t is a temperature-dependent constant, but this is observed far less frequently than the square-law and power-law behaviour described above. An example is in the case of electrodeposited p-type CdTe films [28.214], where Rose's expression was followed and the trap concentration per unit energy range was $10^{21}\,\mathrm{eV}^{-1}\,\mathrm{m}^{-3}$. It was also observed in early work on copper phthalocyanine films with ohmic Au electrodes [28.215]. The discussion above for all types of SCLC considers carrier injection of a single type only. Double injection, when one electrode is ohmic for electrons and the other ohmic for holes, is covered in detail in the text by *Lampert* and *Mark* [28.216] on current injection in solids, and is also discussed by *Lamb* [28.189].

The assignment of the observed I–V characteristics to SCLC usually involves testing the current and voltage data for different film thicknesses. The examination of (28.104) and (28.111) shows that in both cases J/d is a function of V/d^2 i. e., in (28.104) $J/d \propto (V/d^2)^2$ and in (28.111) $J/d \propto (V/d^2)^{l+1}$. This observation is generally true for SCLCs that are due to one type of carrier; either electrons or holes are injected into the film and contribute to the observed current. The *scaling law* generalizes this observation in the form,

$$\frac{J}{d} = f\left(\frac{V}{d^2}\right), \tag{28.109}$$

where f represents the SCLC function for the particular film. The scaling law provides a means of testing for SCLCs, provided that the film thickness can be varied without altering any other properties. Figure 28.31 shows an application of the scaling law to Bi_2O_3 thin films in which I/d has been plotted against (V/d^2) on a log–log plot. The electrode area is the same (A in $J = I/A$) but the samples have different thicknesses with d in the range 75–680 nm (almost an order of magnitude). The scaling law is at least obeyed approximately for the wide range of thicknesses and over the entire range of currents; and the observed I–V characteristics on different films can all be interpreted in terms of a single SCLC mechanism similar to (28.111). Note that the function f in (28.109) depends on the exact nature of transport involved through the space charge, with (28.104) and (28.111) being two special examples in which the transport is shallow trap controlled or the deep trap is exponential in energy distribution as in (28.108). If both electrons and holes are injected, the analysis becomes more complicated and the reader is referred to [28.216].

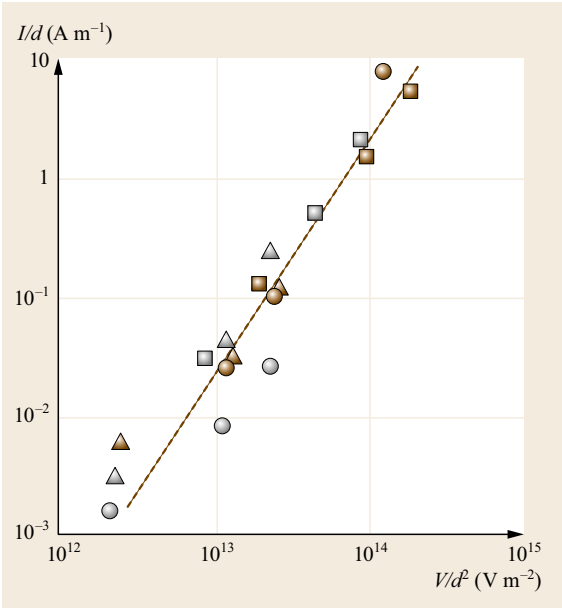


Fig. 28.31 Plot of I/d versus V/d^2 for BiO films with d in the range 75 to 680 nm. All samples have the same sample electrode area. *Each symbol* corresponds to a different thickness of film. A single *line* through all points suggests SCLC. The counter electrodes were Zn and Au. (Data extracted from [28.217, Fig. 4])

The second bulk-limited conduction process to be discussed is the Poole–Frenkel effect, which is essentially the field-assisted lowering of the Coulombic potential barrier φ between electrons located at impurity sites and the edge of the conduction band. Electrons in such centres are unable to contribute to the conductivity until they overcome the potential barrier φ and are promoted into the conduction band. For an applied electric field F the potential is reduced by an amount eFx where e is the electronic charge and x is the distance from the centre. For high electric fields the lowering of the potential can be significant. Such a variation is illustrated in the literature, usually for the case of a donor level located an energy E_d below the bottom of the conduction band [28.187]. The potential energy of the electron in the Coulombic field is $-e^2/(4\pi\varepsilon_r\varepsilon_0 x)$, and there is also another contribution to the potential energy, $-eFx$, resulting from the applied field F. The combined effect of both these contributions to the potential energy is that there is a maximum in the potential energy for emission at a distance x_m from the centre. The effective potential barrier for Poole–Frenkel emission is then lowered by an amount $\Delta\varphi_{PF}$, which depends on the electric field according to the relationship

$$\Delta\varphi_{PF} = \beta_{PF} F^{1/2}\,, \tag{28.110}$$

where

$$\beta_{PF} = \left(\frac{e^3}{\pi\varepsilon_r\varepsilon_0}\right)^{1/2} \tag{28.111}$$

is referred to as the Poole–Frenkel field-lowering coefficient. The conductivity, and consequently the current density, depend exponentially on the potential barrier for excitation of electrons into the conduction band, and since this is lowered by $\Delta\varphi_{PF} = \beta_{PF}F^{1/2}$ the current density is increased by an exponential factor. The current density is given by

$$J = J_0 \exp\left(\frac{\beta_{PF}F^{1/2}}{k_B T}\right), \tag{28.112}$$

where $J_0 = \sigma_0 F$ is the low-field current density. Although J_0 in this expression depends linearly on F,

it is frequently ignored in comparison with the exponential term, which increases far more rapidly with F. Equation (28.112) is usually applicable for thin-film insulators and semiconductors in the presence of traps. If, rarely, traps are absent and the low-field conductivity σ_0 shows an intrinsic dependence proportional to the factor $\exp(-E_g/(2k_BT))$, where E_g is the energy gap, then a factor of 2 appears in the denominator of the exponential term in (28.112) [28.187]. Since $F = V/d$, the above expression can be written

$$J = J_0 \exp\left(\frac{\beta_{PF} V^{1/2}}{k_B T d^{1/2}}\right). \qquad (28.113)$$

Therefore a plot of $\log J$ against $V^{1/2}$ should show a linear relationship in the case of Poole–Frenkel conductivity, from which a value of the field-lowering coefficient can be deduced and compared with the theoretical value of (28.111). If it is necessary for the variation of the pre-exponential factor to also be taken into account, then a linear plot of $\log(J/V)$ against $V^{1/2}$ should be obtained.

Modifications to (28.112) have been proposed to account for the somewhat different current density dependence in the presence of various combinations of donors and traps. *Simmons* [28.187] mentioned the case where the material contains donor levels below the Fermi level as well as shallow neutral traps, in which the coefficient of $F^{1/2}/(k_BT)$ is one half of the theoretical value given by (28.111). For the case of a nonuniform electric field, having a maximum value of $\alpha^2 F$, where F is the mean field, the following expression has been proposed [28.218]

$$J = \frac{2J_0 k_B T}{\alpha\beta_{PF} F^{1/2}} \exp\left(\frac{\alpha\beta_{PF} F^{1/2}}{k_B T}\right). \qquad (28.114)$$

Although this expression was originally proposed for the case of CdTe films, it has also been adopted for the interpretation of results in other materials.

The Poole–Frenkel effect has been observed in many types of thin film, including insulating oxides, semiconductors and organic materials. The main indicator of Poole–Frenkel conductivity is a linear dependence of $\log J$ on $V^{-1/2}$, although it is also seen in the case of the Schottky effect, as will be discussed later. Care must be taken in distinguishing between the two processes. In SiO films, good agreement with (28.113) was obtained [28.219], but the value of β derived exceeded the theoretical Poole–Frenkel value. This type of discrepancy had previously been discussed in the literature [28.220], and it was later proved that in this model the J–V characteristics were symmetric over five orders of magnitude of applied voltage, irrespective of electrode material, which ruled out the Schottky mechanism [28.221]. In another important insulator, silicon nitride, *Sze* [28.222] observed Poole–Frenkel behaviour in metal–insulator–semiconductor (MIS) diodes fabricated on a degenerate Si substrate, over a range of nearly four decades in current. It was found that the current density versus electric field characteristics were independent of film thickness, sample area, electrode materials and polarity, suggesting that the conductivity was bulk-limited rather than electrode-limited, and therefore eliminating the Schottky effect. Poole–Frenkel conductivity has also been observed in several semiconducting cadmium compounds, such as RF-sputtered CdS films [28.223, 224] and more widely in CdTe films [28.218, 225, 226]. In Al-CdTe-Al samples, Poole–Frenkel-type conductivity was consistently observed over the temperature range 0–108 °C. These results are typical of Poole–Frenkel behaviour, and are illustrated in Fig. 28.32 [28.226]. However, the value of the experimental field-lowering coefficient increased from 4.95×10^{-5} eV m$^{1/2}$ V$^{-1/2}$ at 0 °C to 7.7×10^{-5} eV m$^{1/2}$ V$^{-1/2}$ at 108 °C, and was about twice the theoretical β_{PF} for CdTe. The Schottky effect was eliminated, as the results did not appear to depend on the electrode workfunction when the Al electrodes were

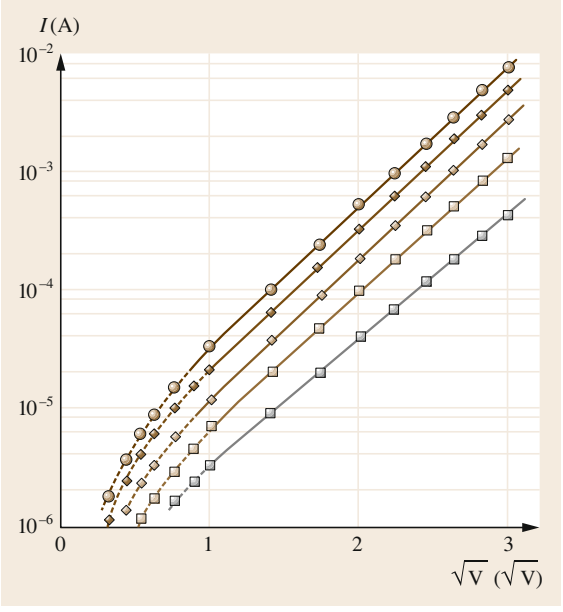


Fig. 28.32 Linear dependence of the logarithm of current on the square root of the applied voltage in an evaporated Al-CdTe-Al sandwich structure of thickness 722.5 nm at 0 °C (*bottom curve*), 30 °C, 55 °C, 80 °C and 108 °C (*top curve*). The behaviour was identified with a form of the Poole–Frenkel effect. (Data extracted from [28.226])

replaced with either In or Ag. These workers adopted the suggestion of *Jonscher* and *Ansari* [28.221] that electrons can hop between sites as the result of thermal activation; on this assumption good agreement was found with the predicted β_{PF}, with an activation energy between hopping sites of 0.15 eV. Very similar results to these were also observed subsequently [28.218], and again the measured value of β exceeded the theoretical Poole–Frenkel value. This coefficient was also calculated from the gradients of the $\ln J$–$V^{1/2}$ characteristics as a function of temperature, and consistent values were obtained. It was suggested that the similarity between the results of this work and the earlier work [28.226] was related to individual centres experiencing varying electric fields, i. e., the electric field was nonuniform. Equation (28.114) was proposed to account for this effect, and has been successful in accounting for the enhanced coefficients sometimes obtained. Poole–Frenkel conductivity was also identified in p-type CdTe, which also shows a change in conductivity type from SCLC to Poole–Frenkel conductivity [28.227]. This conduction mechanism has also been observed in other cadmium compounds, namely CdSe [28.209] and Cd_3As_2 [28.228]. In phthalocyanine films several instances of this effect have been reported. In oxygen-doped copper phthalocyanine films with a Au and an Al electrode, slightly enhanced values of β were obtained and attributed to a nonuniform electric field, whereas in annealed films the field appeared to be uniform [28.229]. Agreement with the basic model was also found for samples with a Au and a Pb electrode [28.198]. Lead phthalocyanine films with Au and Al electrodes exhibited two different values of β [28.200] consistent with Poole–Frenkel conductivity at higher fields and the Schottky effect at lower fields. However other workers had not considered that Poole–Frenkel conductivity was apparent in the same system [28.230], preferring an interpretation based on two different Schottky barrier widths. Both Schottky and Poole–Frenkel behaviour were observed in oxygen-containing lead phthalocyanine films with two Au electrodes [28.231].

Hopping is a third bulk-limited conduction process which is observed in thin films. This type of conductivity is observed particularly in noncrystalline materials, and is thoroughly discussed in various texts concerned with this topic [28.130, 232]. In this type of material the lack of long-range order results in a phenomenon known as localisation, where the energy levels do not merge into a continuum, particularly in the band-tail regions at the edge of the energy bands. Electrons are transported through the material in a series of jumps or *hops* from one localised energy level to another. The hopping process can occur between the localised energy levels even when only small amounts of thermal energy are available, because the localised levels are normally very closely spaced in energy. For this reason, the process can take place at very low temperatures when other processes are energetically impossible. *Mott* and *Davis* [28.130] argued that the conductivity σ exhibits different behaviour in different temperature regions. At higher temperatures thermal excitation of carriers to the band edges is possible and extended-state or free-band conductivity can take place, while at lower temperatures, where less thermal energy is available, hopping may occur. Different varieties of hopping may take place, and these are distinguished by the length of the hop. Nearest-neighbour hopping is self-explanatory, while in variable-range hopping the hops are on average further than to the nearest neighbour, but the energy difference between the states involved is lower. In the case of variable-range hopping the conductivity has been shown to follow a relationship of the form [28.232]

$$\sigma = \sigma_0 \exp\left(-\frac{A}{T}\right)^{1/4} \qquad (28.115)$$

where σ_0 and A are constants. This relationship is known as the Mott $T^{1/4}$ law, since a plot of log σ against $T^{-1/4}$ shows a linear characteristic with negative slope.

Hopping conduction has been observed in many materials, and is extensively reviewed in the literature [28.130, 232, 233]. The $T^{-1/4}$ law is observed in amorphous Ge [28.234], with the results of *Walley* and *Jonscher* [28.235] also showing this behaviour [28.130]. There is also an example of its appearance in amorphous carbon films [28.236]. AC hopping effects have been observed in several cadmium chalcogenides, and this work is reviewed elsewhere [28.38]. DC hopping was identified in CdTe films with Al electrodes [28.237] following earlier work on AC conductivity [28.195]. The results suggested a hopping process at temperatures below about 175 K, with a hopping centre concentration of about $9\times10^{12}\,\mathrm{m}^{-3}$ and a hopping mobility of $2.16\times10^{-5}\,\mathrm{m}^2/(\mathrm{V\,s})$, the latter considerably lower than the free-band mobility applicable at higher temperatures. The extent of the electron wavefunction was estimated to be 0.69 nm, corresponding to the lattice parameter in CdTe, and supporting the existence of a nearest-neighbour hopping process. Hopping at lower fields was also identified in CdSe films [28.238] and associated with an impurity conduction process proposed previously [28.239]. Variable-range hopping following the $T^{-1/4}$ law has been observed in iron phthalocyanine and in triclinic lead phthalocyanine films [28.240, 241], although the temperature range was restricted. In the work on iron

phthalocyanine the law was followed over the range 140–220 K, with particularly good agreement in the range 156–175 K. In the lead phthalocyanine films similar very good linear behaviour was observed in the temperature range of approximately 220–260 K. Hopping at low temperatures has also been indicated from AC measurements in various phthalocyanine systems [28.242–245].

The first electrode-limited conduction process to be considered is that of tunnelling, which is a quantum-mechanical effect, in which the electron wavefunction is attenuated only moderately by a *thin* barrier, resulting in there being a finite probability of its existence on the opposite side of the barrier. Tunnelling directly from the Fermi level of one electrode to the conduction band of the other is normally possible only for very thin films of thickness less than about 10 nm when subjected to a high electric field. *Simmons* [28.246] investigated the effects of tunnelling between similar electrodes separated by a thin insulating film of thickness s, where the barrier presented by the film was of arbitrary shape. The Wentzel–Kramers–Brillouin (WKB) approximation was used to predict the probability of an electron penetrating the barrier. Different approximations for the current density were found, depending on the relative values of the applied voltage V and φ_0/e, where φ_0 is the barrier height at the electrode interface and e is the electronic charge. These approximations are too unwieldy to be included here, however for the case of $V > \varphi_0/e$ it is only necessary for electrons tunnelling from the Fermi level of one electrode to penetrate a distance $\Delta s < s$ to reach unoccupied levels in the second electrode. This worker also suggested [28.247] that for $V > \varphi_0/e$ a modified *Fowler–Nordheim* [28.248] expression may be applicable for tunnelling through an interfacial region. In this case, the electric field at the barrier is sufficiently high to reduce the barrier width, measured at the Fermi level, to about 5 nm. Under these circumstances the current density is related to the voltage and barrier thickness according to the expression

$$J = \frac{e^3 V^2}{8\pi h \varphi_0 d_t^2} \exp\left(-\frac{8\pi (2m)^{1/2} \varphi_0^{3/2} d_t}{3ehV}\right), \quad (28.116)$$

where h is Planck's constant, φ_0 and d_t are the barrier height and effective thickness, respectively, of the tunnelling barrier and m is the free electron mass. The tunnelling current is largely controlled by the barrier height φ_0, which appears both in the pre-exponential and in the exponential terms. Fowler–Nordheim tunnelling is indicated by a linear dependence of $\log J/V^2$ on $1/V$. A further interesting point for tunnelling in general is that there is very little temperature dependence in the current density, a feature allowing it to be distinguished from other temperature-dependent mechanisms, such as the Schottky effect. There is in fact a very slight quadratic dependence on temperature [28.187], but this is normally negligible within the limitations of experimental measurement.

Early measurements on tunnelling were made by *Fisher* and *Giaever* [28.249] on Al_2O_3 produced by oxidation of Al. There was good agreement with an early tunnelling model [28.250], providing the electron effective mass in the insulator was about $1/9$ of the free electron mass. Very thin films of Al_2O_3 and BeO showed ohmic and faster-than-exponential dependencies of J on V at lower and higher voltages respectively [28.251]. These results were correlated with direct metal-to-metal electrode tunnelling at low voltages, and from the Al electrode to the insulator conduction band at higher voltages. The model of *Simmons* [28.187] was tested by measurements on thermally grown Al_2O_3 films on Al, which used several different types of evaporated counter electrodes [28.252]. An extended tunnelling model [28.253], which is applicable to samples with dissimilar electrodes, was followed over nine decades of current. Measurements on thermally grown SiO_2 films, of thickness 65–500 nm with Al or Au electrodes, showed strong evidence of Fowler–Nordheim tunnelling [28.254]. Since these films were relatively thick, high voltages could be applied, which reduced the Fermi level barrier thickness to that suitable for tunnelling. A linear dependence of $\log J/V^2$ on $1/V$ was obtained, but the measured currents were somewhat lower than those predicted by the theory, probably as the result of trapping effects. Effective masses of $0.48m$ for Ag, $0.39m$ for Al and $0.42m$ for Si were estimated, where m is the free electron mass. Since these early tunnelling measurements, the mechanism has been observed in many different insulating films.

Tunnelling is semiconductors forms the basis of the operation of the tunnel diode. However, in this case a doubly degenerate p-n junction is used, and tunnelling occurs across the depletion region. The operation of this device is described in a standard text on semiconductor devices [28.255]. In sandwich samples of nondegenerate materials, tunnelling has been observed in CdTe [28.225]. In this case there was a linear dependence of the logarithm of the drift mobility on reciprocal electric field for fields in excess of 1.2×10^6 V/m, which is consistent with (28.116). Tunnelling has also been observed in sandwich structures consisting of a copper phthalocyanine film with In electrodes [28.256]. A modified Fowler–Nordheim expression for tunnelling through an interfacial barrier, when the electric field is sufficient to reduce the barrier width to about 5 nm, has been proposed [28.247], and was applied in this case to the In/Cu phthalocyanine

interfacial region. Depending on the sample structure, barrier heights of 0.27 and 0.36 eV were derived. Current densities were somewhat lower than predicted by theory, but were accounted for by the fact that tunnelling areas are often considerably lower than the geometric area [28.18] and that in the presence of traps and space charge the current density is also reduced [28.257].

The second electrode-limited conduction process to be considered may occur when the insulating or semiconducting film is too thick for tunnelling to take place and the concentration of allowed states in the forbidden gap is too low for the hopping process to occur. It is known as the Schottky effect, and is the field-assisted lowering of a potential barrier at the injecting electrode. It is similar in origin to the Poole–Frenkel effect, the latter effect having previously been termed the bulk analogue of the Schottky effect [28.258]. The potential barrier at the injecting electrode interface is reduced by an amount $\Delta\varphi_S$ which is given by

$$\Delta\varphi_S = \beta_S F^{1/2} , \tag{28.117}$$

where

$$\beta_S = \left(\frac{e^3}{4\pi\varepsilon_r\varepsilon_0}\right)^{1/2} \tag{28.118}$$

is the Schottky field-lowering coefficient. The barrier lowering process is similar to that for the Poole–Frenkel effect, and is also illustrated in various literature reviews [28.187–189]. Differences between the standard Poole–Frenkel and Schottky coefficients given by (28.111) and (28.118) are related to the dissimilar symmetry of the potential barriers in the two cases. Clearly the two field-lowering coefficients are related, with $\beta_{PF} = 2\beta_S$. The fundamental thermionic emission equation of Richardson

$$J = AT^2 \exp\left(-\frac{\varphi}{k_B T}\right) \tag{28.119}$$

gives the current density flowing by a process of electron emission over a potential barrier of height φ at a temperature T. In this expression, the constant $A = 1.2\times10^6\,\mathrm{A/m^2}$, and is known as the Richardson constant. The current density depends only on the potential barrier height and the temperature, and does not require an applied voltage to flow, the electrons acquiring thermal energy only. For the case of Schottky emission, if φ_0, given by (28.97), is the zero-voltage barrier height then the reduced barrier height φ is given by $(\varphi_0 - \Delta\varphi_S)$, so that (28.119) becomes

$$J = AT^2 \exp\left(-\frac{\varphi_0}{k_B T}\right) \exp\left(\frac{\beta_S F^{1/2}}{k_B T}\right) \tag{28.120}$$

or

$$J = AT^2 \exp\left(-\frac{\varphi_0}{k_B T}\right) \exp\left(\frac{\beta_S V^{1/2}}{k_B T d^{1/2}}\right) . \tag{28.121}$$

Thus, the Schottky effect should show a linear dependence of $\log J$ on $V^{1/2}$, as for the Poole–Frenkel effect described by (28.113). In principle the two effects can be distinguished by the measured value of the field-lowering coefficient, which should be twice as high for the Poole–Frenkel effect as for the Schottky effect.

The Schottky effect has been reported in many inorganic materials, both insulators and semiconductors. *Emtage* and *Tantraporn* [28.259] observed the effect in both Al_2O_3 and GeO_2 sandwich structures having Al electrodes, while *Pollack* [28.260] made similar observations in Al_2O_3 films with Pb electrodes. In this case, not only the constant voltage J–V characteristics were as predicted by (28.121), but the gradients of $\log J$–$V^{1/2}$ curves increased linearly with reciprocal temperature for temperatures above 235 K, as also predicted by this expression. Below this temperature the variation with temperature was very small, and therefore tunnelling was identified in this region. The dependence of the Schottky effect on the species of electrode was determined in Ta-Ta_2O_5-Au samples prepared by electron beam evaporation and plasma oxidation [28.261]. In this system the Schottky effect was observed when the Ta electrode was biassed negatively and tunnelling when it was biassed positively. Similar results were obtained in CdS films [28.192] using In (ohmic) and Au (blocking) contacts. For electron injection from the Au electrode Schottky-type behaviour was observed, with a Schottky barrier height of 0.88 eV obtained from differential capacitance measurements, close to the value expected for Au contacts to CdS single crystals. Field-lowering behaviour has also been observed in Cd_3As_2 films with a bottom Ag electrode and a top electrode of Ag, Al or Au [28.228]. Although the Schottky barrier height is unlikely to be consistent with the simple model of (28.97) owing to the very narrow band gap and the presence of surface states, the experimental value of β was consistent with the value of β_S predicted by (28.118) for a relative permittivity value of 12.

The Schottky effect has also been observed in several organic thin-film materials. Examples are given here only for the phthalocyanines, which are invariably p-type semiconductors. The Schottky effect occurs for hole injection at the blocking contact. Both metal-free [28.262] and copper phthalocyanine films [28.263] having Al electrodes have shown this effect. Behaviour consistent with the Schottky effect has also been observed in planar copper phthalocyanine films with Al electrodes [28.264], with a depletion region width

of thickness 96 nm. In sandwich structures having a copper phthalocyanine film and a positively biased Al electrode, values of β several times higher than the theoretical $\beta_S = 2\times10^{-5}\,\mathrm{eV\,m^{1/2}\,V^{-1/2}}$ were obtained [28.229]. However, if the theoretical value of β_S were assumed and also that most of the voltage was dropped over a depletion region of thickness d_S as suggested previously [28.264], values of $d_S = 120\,\mathrm{nm}$ and $\varphi_0 = 0.88\,\mathrm{eV}$ were derived. This type of assumption has also been subsequently applied in the case of hole injection from a Pb electrode into copper phthalocyanine films [28.198] and for hole injection in triclinic lead phthalocyanine films from an Al electrode [28.200, 230]. In both cases the Schottky barrier width was 50 nm with a barrier height of either 1.11 or 1.0 eV.

In this section the main high-field conduction processes that might be observed in insulating and semiconducting films have been reviewed, and illustrative examples quoted from the literature. It should be emphasised that identification of a particular mechanism is nontrivial, particularly in distinguishing between the Poole–Frenkel and Schottky effects which have very similar J–V dependencies. Furthermore J–V data obtained over a specified voltage range may also show a good correlation with both SCLC and the Poole–Frenkel effect. It is essential to measure temperature dependencies and also to determine the effects of electrode species on the conductivity, making a full consideration of the type of contacts (ohmic or blocking). As devices and structures become ever smaller with the advance of nanotechnology, some of these conduction processes will inevitably become important in nanoelectronic devices. A discussion concerning this aspect is given elsewhere in the literature [28.11].

28.4 Concluding Remarks

In this chapter, an overview of the deposition methods, structures and major properties of thin films has been given. In this section these are briefly summarised and some possible future developments outlined.

Films may be prepared by many different methods. However, these may usually be classified into chemical or physical methods. In the physical methods the film is deposited from a vapour, maintained at a pressure considerably below that of the atmosphere. Vacuum evaporation is the first of these, where the material is transformed into the vapour phase by heating at reduced pressure. Various types of process have been developed to ensure stoichiometry of the deposited material, such as reactive evaporation and co-evaporation. Taken to its limit of sophistication, the method may be operated under UHV conditions using several sources, each of which may be independently controlled. In this case epitaxial films may be deposited, and the method is known as molecular-beam epitaxy (MBE). The second widely used physical method is that of sputtering, which involves the removal of particles (atoms, molecules or larger fragments) from the surface of a target. This is normally achieved using a gas discharge, which delivers high-energy ions to the target surface. Sputtering has several advantages over evaporation, including deposition of stoichiometric materials, better adhesion to the substrate and the capability of depositing most materials. There are several varieties of sputtering, including DC (diode) sputtering, reactive sputtering and RF sputtering, the latter enabling the deposition of insulators as well as conducting materials. Reactive and co-sputtering give greater flexibility, and higher deposition rates are possible using magnetron sputtering. Various chemical methods may also be utilised for the deposition of films from materials having specific chemical properties. Electrodeposition and electroplating involve passing a high electric current through a solution containing the material, while chemical vapour deposition (CVD) entails the use of various chemical reactions to generate a vapour. The latter method is frequently used in the deposition of electronic materials, and substantial work has been performed in developing processes for several of these. Again there are several different subdivisions of this method, operating at different pressures and with different types of precursor materials. A third chemical method that has found favour in recent years is Langmuir–Blodgett deposition, where molecular layers are applied individually using a dipping process. It has the advantage that high vacuum and elevated temperatures are not required, and film thicknesses are easily derived from the number of layers deposited. It is, however, limited to certain defined materials, and does not have the flexibility of the other methods.

In common with bulk materials, the crystal structure and morphology of thin films may be investigated using diffraction methods and various microscopic techniques (electron microscopy, scanning tunnelling microscopy, etc.). Diffraction methods allow the type of crystal structure to be determined by measurement of the unit cell dimensions in the case of epitaxial and polycrystalline films. Frequently a material may exist in two or more crystalline modifications, and generally these may be distinguished by comparison of diffraction data

from the film with tabulated standard data. The type of crystalline modification that is deposited can depend sensitively on the deposition conditions such as the substrate temperature and deposition rate, and this dependence may be established by systematically depositing films under a wide combination of conditions. Many films have a preferred crystalline orientation, and this too may sometimes be correlated with deposition conditions and the substrate morphology. In polycrystalline films, the crystallites are generally preferentially oriented and show a fibre texture. Estimates of the mean microcrystallite size may be obtained from the width of the diffraction peaks using the Scherrer method. The morphological characteristics of thin films commonly have a columnar grain structure when physical deposition methods are used, and the orientation of the grains may sometimes be related to the incident direction of the depositing atoms or molecules. Larger-scale features of the morphology may often be accounted for in terms of a zone model, where the observed features are identified with the ratio of the substrate temperature to the melting temperature of the film material. Examples of some of these effects in different thin-film materials have been given, and although these are representative of behaviour shown in many thin-film systems, it is clear that both crystal structure and morphology require determination in any specific process used to deposit a particular film material. These features are likely to be particularly important in influencing the optical and electrical properties.

The optical and electrical properties of thin films are those most relevant to this Handbook, and examples of some of the basic properties have been considered as a foundation for some of the more advanced applications described in other chapters. The interference properties of light are fundamental to most of the phenomena observed. In principle the properties of light propagating through media are described by Maxwell's equations, which allow a full analysis for both nonabsorbing and absorbing films for various states of polarisation and optical frequencies. The Fresnel coefficients for reflection and transmission of light at the boundaries between different optical media allow expressions for the energy reflectance and transmittance to be derived. Some examples of these have been given, together with an introduction to matrix methods which may be used to calculate the electric and magnetic field vectors for optical designs of considerable complexity, involving both nonabsorbing and absorbing films. The operation of optical filters, other optical components and fibre-optic waveguides may be predicted by advanced application of these methods. Examples of the basic electrical properties of thin films, which are directly related to the film thickness, are lateral electrical conductivity and high-field conduction processes. Lateral electrical conduction in thin films is a very important feature, which is determined primarily by electron scattering within the film. For metallic films, conductivity–film thickness relationships have been derived in terms of a parameter k, the ratio of the film thickness to the bulk electron mean free path. One would intuitively expect the conductivity to decrease (or the resistivity to increase) with decreasing k, since this implies a curtailment of the mean free path. This problem was first tackled over a century ago by Thomson, and more realistic expressions have been obtained subsequently by solving the Boltzmann transport equation. Unfortunately, the derived general expressions are somewhat complex, but a wide variety of approximations have been obtained for various ranges of k, and the conductivity is predicted to reduce to that of the bulk material for thick films. Further refinements in the analysis were made by introducing a specularity parameter, p, which takes into account the proportion of electrons which are specularly scattered at the surfaces. Better agreement with experimental results is generally obtained if the value of p is intermediate between that for totally diffuse scattering ($p = 0$) and totally specular scattering ($p = 1$). Another development of the theory includes assigning different specularity parameters to the two film surfaces, since scattering at the substrate surface is clearly different from scattering at the top surface. In general the various expressions derived from this type of model are consistent with each other, in that those derived from the more complex assumptions normally reduce to those obtained using simpler assumptions if suitable limits are chosen. Scattering by internal grain boundaries has also been investigated by solving the Boltzmann transport equation, and allows contributions to the conductivity resulting from internal features, as well as the surfaces, to be included in the analysis. The ratio of the thin-film to bulk temperature coefficient of resistivity has been shown to be equal to the ratio of thin-film to bulk conductivity for the case of thicker films, although no such convenient relationship applies for thinner films owing to the dependence of the bulk mean free path on temperature, which is particularly significant for very thin films. Considerations of these types of effect need to be fully assessed when designing interconnects to integrated circuits operating in particular temperature ranges. In insulating or semiconducting thin films, very high electric fields may be applied across the thickness of the film. Under these circumstances a variety of different conduction processes may occur, none of which are observed at lower fields. Crucial to the type of conduction process which develops is the type of contact at the interface. Ohmic contacts allow currents to pass unimpeded by the contact, and

the conduction process is bulk-limited, while blocking contacts effectively limit the conductivity, which is then electrode-limited. Bulk-limited conduction processes include SCLC, Poole–Frenkel conductivity and hopping. Varies types of SCLC have been observed, depending on the type of trap distribution present, if any, and may be distinguished by their characteristic voltage and temperature dependencies. Poole–Frenkel conductivity results from field-dependent lowering of the potential barrier for emission of electrons into the conduction band. Again, various varieties of this process have been observed depending on the appearance of donors and/or traps, and also for the case of nonuniform fields. Hopping is observed particularly in noncrystalline solids, where electrons hop from one localised energy level to another. It is characterised by low mobility, but requires little thermal energy, and is therefore often observed at low temperatures. Different types of hopping are characterised by the mean distance of the electron hopping process, and show different characteristic behaviours. Electrode-limited conduction processes include tunnelling and the Schottky effect. Tunnelling is a quantum-mechanical effect, where for a thin barrier the electron wavefunction has a finite probability of existence on the opposite side of the barrier. Tunnelling may be either directly from one electrode to the other, or may be from the Fermi level of the injecting electrode into the conduction band of the insulator or semiconductor. The type of tunnelling depends on the relative values of the applied voltage and the barrier height at the interface. The tunnelling current is almost independent of temperature, a feature which allows it to be identified relatively easily. For structures where the barrier is too thick for tunnelling to occur, the Schottky effect may take place instead. This is the field-assisted lowering of a potential barrier at the interface, and has several similarities with the bulk Poole–Frenkel effect. Similar J–V characteristics are predicted for both, but may usually be distinguished from the value of a measured field-lowering coefficient, or from considerations of the current level drawn with varying electrode species.

In conclusion, it should be noted that in this chapter only the main established deposition methods and some of the structural, optical and electrical properties of thin films have been reviewed. Although it is hoped that the references included herein are sufficient to allow the reader to explore the subject in some considerable depth, it is clear that these can only provide a snapshot of knowledge at the present time. During the past half century, there have been tremendous developments in all areas of thin-film knowledge, although the basic principles and properties stressed in this chapter underpin most of them. The main physical methods of deposition have been known for a century or so, but improvements have been contingent on advances in vacuum technology, gas-handling capabilities and theoretical work concerning atomic interactions in solids. MBE deposition systems of considerable complexity and sophistication now exist, enabling the deposition of a considerable range of materials in a variety of structures and designs, although only some of the most important processes have been developed to the level where they are employed in automated manufacture. Control of many factors are vital in successful deposition processes, and progress in this area has gone hand in hand with the development of computerised monitoring and data-handling systems. Progress in the manipulation of individual atoms and molecules has been made by the adoption of techniques originally developed for very high-resolution surface imaging, such as scanning tunnelling and atomic force microscopy. Chemical methods of film deposition have developed into many dedicated CVD systems, exploiting reactions which are specific to a particular material or class of materials. Additionally many types of films may be deposited using the Langmuir–Blodgett technique, although this relies on the ingenuity of chemists to develop suitable substituted organic compounds. Further development of many of these deposition methods will undoubtedly progress in future years, particularly as the power of computer systems increases. The most significant advances are likely to be in the area of nanotechnology, where MBE systems, equipped with facilities to build nanostructures atom by atom, are currently under development. The structure and morphology adopted by thin films are likely to have a significant influence on all their properties; this is not restricted to just the optical and electrical properties described in this chapter. Diffraction methods are becoming more precise, and can now be used for the investigation of real-time changes when synchrotron radiation is used. Scanning tunnelling and atomic force microscopic techniques have enabled the acquisition of surface images with resolution of atomic dimensions, and the increased use and development of these and associated techniques will be necessary for future investigation of nanostructures. Optical properties described in this chapter will no doubt be exploited in more complex and sophisticated filter and mirror designs, as well as their employment in fibre-optic cables and applications in photonics. Recent developments in the area of negative-refractive-index materials will no doubt also be exploited in due course. The electrical properties of thin films under both low- and high-field conditions are reasonably well understood, and it is anticipated that applications are likely to be made in the area of nanoelectronics. Some possible applications have been

proposed elsewhere [28.11]. The provision of electrical contacts will remain of particular importance, although it may not be possible to consider the contact region separately from the remaining nanostructure. The efficient injection of charge into nanostructures using tunnelling and Schottky barrier (blocking) contacts may be achieved, and high current densities may be obtained with trap-free SCLC contacts. It is nevertheless clear, however, that in all areas of future development in thin-film technology, the control of deposition processes and materials processing aspects are of paramount importance. This is likely to remain a considerable challenge in the future.

Acknowledgments. We thank Jan Mistrik for his help in checking the section on opical properties.

References

28.1 L. Eckertová: *Physics of Thin Films*, 2nd edn. (Plenum, New York 1986)
28.2 J. Thewlis: *Concise Dictionary of Physics and Related Subjects*, 2nd edn. (Pergamon, Oxford 1979) p. 336
28.3 R.W. Hoffman: The mechanical properties of thin condensed films. In: *Physics of Thin Films*, Vol. 3, ed. by G. Hass, R.E. Thun (Academic, San Diego 1966) p. 211
28.4 D.S. Campbell: Mechanical properties of thin films. In: *Handbook of Thin Film Technology*, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970), Chap. 12
28.5 C.A. Neugebauer, M.B. Webb: J. Appl. Phys. **33**, 74 (1962)
28.6 L.I. Maissel: Electronic properties of metallic thin films. In: *Handbook of Thin Film Technology*, Vol. 13, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970)
28.7 K.H. Behrndt: Film-thickness and deposition-rate monitoring devices and techniques for producing films of uniform thickness. In: *Physics of Thin Films*, Vol. 3, ed. by G. Hass, R.E. Thun (Academic, San Diego 1966) p. 1
28.8 W.A. Pliskin, S.J. Zanin: Film thickness and composition. In: *Handbook of Thin Film Technology*, Vol. 11, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970)
28.9 J.P. Hirth, K.L. Moazed: Nucleation processes in thin film formation. In: *Physics of Thin Films*, Vol. 4, ed. by G. Hass, R.E. Thun (Academic, San Diego 1967) p. 97
28.10 C.A. Neugebauer: In: *Condensation, Nucleation and Growth of Thin Films*, Vol. 8, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970)
28.11 R.D. Gould: High field conduction in nanostructures. In: *Encyclopedia of Nanoscience and Nanotechnology*, Vol. 3, ed. by H.S. Nalwa (American Scientific, Stevenson Ranch 2004) pp. 891–915
28.12 L.I. Maissel, R. Glang (Eds.): *Handbook of Thin Film Technology* (McGraw–Hill, New York 1970)
28.13 S.M. Rossnagel, A. Ulman, M.H. Francombe (Eds.): *Thin Films*, Vol. 25 (Academic, San Diego 1998)
28.14 H.S. Nalwa (Ed.): *Handbook of Thin Film Materials* (Academic, San Diego 2001)
28.15 O.S. Heavens: *Thin Film Physics* (Methuen, London 1970)
28.16 L.I. Maissel, M.H. Francombe: *An Introduction to Thin Films* (Gordon and Breach, New York 1973)
28.17 T.J. Coutts (Ed.): *Active and Passive Thin Film Devices* (Academic, New York 1978)
28.18 K.I. Chopra: *Thin Film Phenomena* (Krieger, New York 1979), (first published by McGraw–Hill, New York, 1969)
28.19 M. Ohring: *The Materials Science of Thin Films* (Academic, San Diego 1992)
28.20 S. Dushman, J.M. Lafferty (Eds.): *Scientific Foundation of Vacuum Technique*, 2nd edn. (Wiley, New York 1962)
28.21 R. Glang: Vacuum evaporation. In: *Handbook of Thin Film Technology*, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970), Chap. 1
28.22 O. Kubaschewski, S.L. Evans, C.B. Alcock: *Metallurgical Thermochemistry*, 4th edn. (Pergamon, Oxford 1967)
28.23 R.E. Honig: RCA Rev. **23**, 567 (1962)
28.24 R.E. Honig, D.A. Kramer: RCA Rev. **30**, 285 (1969)
28.25 H. Hertz: Ann. Phys. Chem. **17**, 177 (1882)
28.26 M. Knudsen: Ann. Phys. **47**, 697 (1915)
28.27 J.P. Hirth, G.M. Pound: *Condensation and Evaporation* (Pergamon, Oxford 1963)
28.28 L. Holland, W. Steckelmacher: Vacuum **2**, 346 (1952)
28.29 A. von Hippel: Ann. Phys. **81**, 1043 (1926)
28.30 H.A. Macleod: *Thin-Film Optical Filters* (Adam Hilger, London 1969)
28.31 T.C. Tisone, J.B. Bindell: J. Vac. Sci. Technol. **11**, 72 (1974)
28.32 R. Brown: Thin film substrates. In: *Handbook of Thin Film Technology*, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970), Chap. 6
28.33 G. Sauerbrey: Z. Phys. **155**, 206 (1959)
28.34 C.D. Stockbridge: Resonance frequency versus mass added to quartz crystals. In: *Vacuum Microbalance Techniques*, Vol. 5, ed. by K.H. Behrndt (Plenum, New York 1966) p. 193
28.35 C.J. Bowler, R.D. Gould: J. Vac. Sci. Technol. A **5**, 114 (1987)
28.36 R. Hultgren, R.L. Orr, P.D. Anderson, K.K. Kelley: *Selected Values of Thermodynamic Properties of Metals and Alloys* (Wiley, New York 1963)
28.37 G. Zinsmeister: Vakuum-Tech. **8**, 223 (1964)
28.38 R.D. Gould: Electrical conduction properties of thin films of cadmium compounds. In: *Handbook of Thin Film Materials*, Chap, Vol. 4, ed. by

H.S. Nalwa (Academic, San Diego 2001) p. 187
28.39 D.R. Stull: *JANAF Thermochemical Tables* (Dow Chemical Co., U.S. Clearinghouse, Springfield 1970)
28.40 F.A. Pizzarello: J. Appl. Phys. **35**, 2730 (1964)
28.41 P.A. Timson, C.A. Hogarth: Thin Solid Films **8**, 237 (1971)
28.42 A.Y. Cho: Thin Solid Films **100**, 291 (1983)
28.43 B.A. Joyce: Rep. Prog. Phys. **48**, 1637 (1985)
28.44 Y. Ota: Thin Solid Films **106**, 3 (1983)
28.45 G.K. Wehner, G.S. Anderson: The nature of physical sputtering. In: *Handbook of Thin Film Technology*, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970), Chap. 3
28.46 O. Almén, G. Bruce: Nucl. Instrum. Methods **11**, 279 (1961)
28.47 F. Keywell: Phys. Rev. **97**, 1611 (1955)
28.48 P. Sigmund: Sputtering by ion bombardment: Theoretical concepts. In: *Sputtering by Particle Bombardment I*, ed. by R. Behrisch (Springer, Berlin, Heidelberg 1981) p. 9, Chap. 2
28.49 R.S. Pease: *Rendiconti della Scuola Internatzionale di Fisica "Enrico Fermi", Corso XIII* (Società Italiana di Fisica, Bologna 1959) p. 158
28.50 P. Sigmund: Phys. Rev. **184**, 383 (1969)
28.51 L.I. Maissel: The deposition of thin films by cathode sputtering. In: *Physics of Thin Films*, Vol. 3, ed. by G. Hass, R.E. Thun (Academic, San Diego 1966) p. 61
28.52 G.H. Kinchin, R.S. Pease: Rep. Prog. Phys. **18**, 1 (1955)
28.53 H.H. Andersen, H.L. Bay: Sputtering yield measurements. In: *Sputtering by Particle Bombardment I*, Chap, Vol. 4, ed. by R. Behrisch (Springer, Berlin, Heidelberg 1981) p. 145
28.54 J. Lindhard, V. Nielsen, M. Scharff: Kgl. Danske Videnskal. Selskab, Mat.-Fys. Medd. **36**(10), 1 (1968)
28.55 A. von Engel: *Ionized Gases*, 2nd edn. (Clarendon, Oxford 1965)
28.56 F. Llewellyn-Jones: *The Glow Discharge* (Methuen, London 1966)
28.57 L.I. Maissel: Applications of sputtering to the deposition of films. In: *Handbook of Thin Film Technology*, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970), Chap. 4
28.58 G. Perny, M. Samirant, B.M. Laville Saint: Compt. Rend. Acad. Sci. Ser. C **262**, 265 (1966)
28.59 G. Perny, B.M. Laville Saint: Proceedings of the International Symposium on Basic Problems in Thin Film Physics, Clausthal 1965, ed. by R. Niedermayer, H. Mayer (Vandenhoeck and Ruprecht, Göttingen 1966), 709
28.60 S.A. Awan, R.D. Gould: Thin Solid Films **423**, 267 (2003)
28.61 W.D. Westwood: Reactive sputtering. In: *Physics of Thin Films*, Vol. 14, ed. by M.H. Francombe, J.C. Vossen (Academic, San Diego 1989) p. 1
28.62 W.R. Sinclair, F.G. Peters: Rev. Sci. Instrum. **33**, 744 (1992)
28.63 L.I. Maissel, P.M. Schaible: J. Appl. Phys. **36**, 237 (1965)
28.64 R. Frerichs: J. Appl. Phys. **33**, 1898 (1962)
28.65 H.C. Theuerer, J.J. Hauser: J. Appl. Phys. **35**, 554 (1964)
28.66 H.E. Winters, E. Kay: J. Appl. Phys. **38**, 3928 (1967)
28.67 G.S. Anderson, W.N. Mayer, G.K. Wehner: J. Appl. Phys. **33**, 2991 (1962)
28.68 P.D. Davidse, L.I. Maissel: J. Appl. Phys. **37**, 574 (1966)
28.69 E.W. Williams: *The CD-ROM and Optical Disc Recording Systems* (Oxford Univ. Press, Oxford 1996)
28.70 R.A. Powell, S.M. Rossnagel: *PVD for Microelectronics: Sputter Deposition Applied to Semiconductor Manufacturing*, Thin Films, Vol. 26 (Academic, San Diego 1999) p. 51
28.71 J.S. Logan, N.M. Mazza, P.D. Davidse: J. Vac. Sci. Technol. **6**, 120 (1969)
28.72 S. Rohde: Unbalanced magnetron sputtering. In: *Physics of Thin Films*, Vol. 18, ed. by M.H. Francombe, J.L. Vossen (Academic, San Diego 1994) p. 235
28.73 B. Window, N. Savvides: J. Vac. Sci. Technol. A **4**, 196 (1986)
28.74 Y. Ochiai, K. Aso, M. Hayakawa, H. Matsuda, K. Hayashi, W. Ishikawa, Y. Iwasaki: J. Vac. Sci. Technol. A **4**, 19 (1986)
28.75 D.S. Campbell: The deposition of thin films by chemical methods. In: *Handbook of Thin Film Technology*, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970), Chap. 5
28.76 D.S. Campbell: Preparation methods for thin film devices. In: *Active and Passive Thin Film Devices*, ed. by T.J. Coutts (Academic, New York 1978) p. 23
28.77 A. Brenner: *Electrodeposition of Alloys, Vol. 1, 2* (Academic, New York 1963)
28.78 K.R. Lawless: The growth and structure of electrodeposits. In: *Physics of Thin Films*, Vol. 4, ed. by G. Hass, R.E. Thun (Academic, San Diego 1967) p. 191
28.79 L. Young: *Anodic Oxide Films* (Academic, New York 1961)
28.80 C.J. Dell'Oca, D.L. Pulfrey, L. Young: Anodic oxide films. In: *Physics of Thin Films*, Vol. 6, ed. by M.H. Francombe, R.W. Hoffman (Academic, San Diego 1971) p. 1
28.81 W.M. Feist, S.R. Steele, D.W. Readey: The preparation of films by chemical vapor deposition. In: *Physics of Thin Films*, Vol. 5, ed. by G. Hass, R.E. Thun (Academic, San Diego 1969) p. 237
28.82 J.R. Knight, D. Effer, P.R. Evans: Solid State Electron. **8**, 178 (1965)
28.83 M.H. Francombe, J.E. Johnson: The preparation and properties of semiconductor films. In: *Physics of Thin Films*, Vol. 5, ed. by G. Hass, R.E. Thun (Academic, San Diego 1969) p. 143
28.84 S.M. Ojha: Plasma-enhanced chemical vapor deposition of thin films. In: *Physics of Thin Films*, Vol. 12, ed. by G. Hass, M.H. Francombe, J.L. Vossen (Academic, San Diego 1982) p. 237
28.85 A.C. Adams: Dielectric and polysilicon film deposition. In: *VLSI Technology*, 2nd edn., Chap, Vol. 6, ed. by S.M. Sze (McGraw–Hill, New York 1988) p. 233

28.86 O.A. Popov: Electron cyclotron resonance plasma sources and their use in plasma-assisted chemical vapor deposition of thin films. In: *Physics of Thin Films*, Vol. 18, ed. by M.H. Francombe, J.L. Vossen (Academic, San Diego 1994) p. 121
28.87 N.G. Dhere: High-T_c superconducting thin films. In: *Physics of Thin Films*, Vol. 16, ed. by M.H. Francombe, J.L. Vossen (Academic, San Diego 1992) p. 1
28.88 I. Langmuir: Trans. Faraday Soc. **15**, 62 (1920)
28.89 K.B. Blodgett: J. Am. Chem. Soc. **56**, 495 (1934)
28.90 G.L. Gaines: Thin Solid Films **99**, ix (1983)
28.91 C.W. Pitt, L.M. Walpitta: Thin Solid Films **68**, 101 (1980)
28.92 S. Baker, M.C. Petty, G.G. Roberts, M.V. Twigg: Thin Solid Films **99**, 53 (1983)
28.93 M.C. Petty: *Langmuir–Blodgett Films: An Introduction* (Cambridge Univ. Press, Cambridge 1996)
28.94 L.S. Miller, P.J.W. Stone: Thin Solid Films **210/211**, 19 (1992)
28.95 L.S. Miller, A.L. Rhoden: Thin Solid Films **243**, 339 (1994)
28.96 P.S. Vincett, G.G. Roberts: Thin Solid Films **68**, 135 (1980)
28.97 A.N. Broers, M. Pomerantz: Thin Solid Films **99**, 323 (1983)
28.98 A.C. Pierre: *Introduction to Sol–Gel Processing* (Kluwer, Boston 1998)
28.99 J. Brinker, G.W. Scherer: *Sol–Gel Science: The Physics and Chemistry of Sol–Gel Processing* (Academic Press, Boston 1990)
28.100 A.M. Buckley, M. Greenblatt: J. Chem. Educ. **71**(7), 599 (1994)
28.101 U. Schubert, N. Hüsing: *Synthesis of Inorganic Materials*, 3rd edn. (VCH-Wiley, Weinheim 2012)
28.102 Y. Cheng, M.S. Xia, F. Luo, N. Li, C.B. Guo, C.D. Wei: Colloid Surf. A **490**, 200 (2016) doi:10.1016/j.colsurfa.2015.11.055
28.103 A.K. Hassan, N.B. Chaure, A.K. Ray, A.V. Nabok, S. Habesch: J. Phys. D **36**(9), 1120 (2003) doi:10.1088/0022-3727/36/9/310
28.104 M. Zahedi, A.K. Ray, D.S. Barratt: J. Phys. D **41**(3), 035404 (2008) doi:10.1088/0022-3727/41/3/035404
28.105 P.H. Mutin, G. Guerrero, J.G. Alauzun: J. Ceram. Soc. Jpn. **123**(1441), 709 (2015) doi:10.2109/jcersj2.123.709
28.106 N. Baccile, F. Babonneau, B. Thomas, T. Coradin: J. Mater. Chem. **19**(45), 8537 (2009) doi:10.1039/b911123a
28.107 R.E. Thun: Structure of thin films. In: *Physics of Thin Films*, Vol. 1, ed. by G. Hass (Academic, San Diego 1963) p. 187
28.108 C.A. Neugebauer: Structural disorder phenomena in thin metal films. In: *Physics of Thin Films*, Vol. 2, ed. by G. Hass, R.E. Thun (Academic, San Diego 1964) p. 1
28.109 I.H. Khan: The growth and structure of single-crystal films. In: *Handbook of Thin Film Technology*, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970), Chap. 10
28.110 G.B. Stringfellow: Rep. Prog. Phys. **43**, 469 (1982)
28.111 D. Walton: J. Chem. Phys. **37**, 2182 (1962)
28.112 D. Walton, T.N. Rhodin, R. Rollins: J. Chem. Phys. **38**, 2695 (1963)
28.113 V.N.E. Robinson, J.L. Robins: Thin Solid Films **20**, 155 (1974)
28.114 J.A. Venables, G.D.T. Spiller: Rep. Prog. Phys. **47**, 399 (1984)
28.115 I.P. Kalinkin, K.K. Muravyeva, L.A. Sergeyewa, V.B. Aleskowsky, N.S. Bogomolov: Krist. Tech. **5**, 51 (1970)
28.116 K.K. Muravjeva, I.P. Kalinkin, V.B. Aleskowsky, N.S. Bogomolov: Thin Solid Films **5**, 7 (1970)
28.117 D.B. Holt: Thin Solid Films **24**, 1 (1974)
28.118 K. Zanio: Cadmium telluride. In: *Semiconductors and Semimetals*, Vol. 13, ed. by R.K. Willardson, A.C. Beer (Academic, New York 1978)
28.119 A. Ashour, R.D. Gould, A.A. Ramadan: Phys. Status Solidi (a) **125**, 541 (1991)
28.120 R.D. Gould: Coord. Chem. Rev. **156**, 237 (1996)
28.121 A.K. Hassan, R.D. Gould: Phys. Status Solidi (a) **132**, 91 (1992)
28.122 S.I. Shihub, R.D. Gould: Phys. Status Solidi (a) **139**, 129 (1993)
28.123 P. Scherrer: Gott. Nachr. **2**, 98 (1918)
28.124 M. Ashida, N. Uyeda, E. Suito: Bull. Chem. Soc. Jpn. **39**, 2616 (1966)
28.125 L. Żdanowicz, S. Miotkowska: Thin Solid Films **29**, 177 (1975)
28.126 G.A. Steigmann, J. Goodyear: Acta Crystallogr. B **24**, 1062 (1968)
28.127 S.R. Elliott: *Physics of Amorphous Materials*, 2nd edn. (Longman, London 1990)
28.128 A. Bienenstock, F. Betts, S.R. Ovshinsky: J. Non-Cryst. Solids **2**, 347 (1970)
28.129 J.R. Bosnell: Amorphous semiconducting films. In: *Active and Passive Thin Film Devices*, ed. by T.J. Coutts (Academic, New York 1978) p. 245
28.130 N.F. Mott, E.A. Davis: *Electronic Processes in Non-Crystalline Materials*, 2nd edn. (Oxford Univ. Press, Oxford 1979)
28.131 D.B. Dove: Electron diffraction analysis of the local atomic order in amorphous films. In: *Physics of Thin Films*, Vol. 7, ed. by G. Hass, M.H. Francombe, R.W. Hoffman (Academic, San Diego 1973) p. 1
28.132 K.H. Norian, J.W. Edington: Thin Solid Films **75**, 53 (1981)
28.133 J.M. Nieuwenhuizen, H.B. Haanstra: Philips Tech. Rev. **27**, 87 (1966)
28.134 S.B. Hussain: Thin Solid Films **22**, S5 (1974)
28.135 J.I.B. Wilson, J. Woods: J. Phys. Chem. Solids **34**, 171 (1973)
28.136 A.G. Stanley: Cadmium sulphide solar cells. In: *Applied Solid State Science*, Vol. 5, ed. by R. Wolfe (Academic, New York 1975) p. 251
28.137 B. Goldstein, L. Pensak: J. Appl. Phys. **30**, 155 (1959)
28.138 J. Saraie, M. Akiyama, T. Tanaka: Jpn. J. Appl. Phys. **11**, 1758 (1972)
28.139 Y.K. Yezhovsky, I.P. Kalinkin: Thin Solid Films **18**, 127 (1973)
28.140 K. Tanaka: Jpn. J. Appl. Phys. **9**, 1070 (1970)
28.141 T.M. Ratcheva-Stambolieva, Y.D. Tchistyakov, G.A. Krasulin, A.V. Vanyukov, D.H. Djoglev: Phys.

Status Solidi (a) **16**, 315 (1973)
28.142 H.W. Lehmann, R. Widmer: Thin Solid Films **33**, 301 (1976)
28.143 J. Jurusik, L. Żdanowicz: Thin Solid Films **67**, 285 (1980)
28.144 J. Jurusik: Thin Solid Films **214**, 117 (1992)
28.145 J. Jurusik: Thin Solid Films **248**, 178 (1994)
28.146 J.R. Ferraro, K. Nakamoto: *Introductory Raman Spectroscopy*, 2nd edn. (Elsevier Science, San Diego 2003) p. 16
28.147 T. Li, T. Patel, I. Banerjee, R. Pearce-Hill, J. Gallop, L. Hao, A.K. Ray: J. Mater. Sci. Mater. Electron. **26**(7), 4810 (2015) doi:10.1007/s10854-015-3122-0
28.148 S. Paul, P.G. Harris, C. Pal, A.K. Sharma, A.K. Ray: Mater. Lett. **130**, 40 (2014) doi:10.1016/j.matlet.2014.05.071
28.149 Z. Knittl: *Optics of Thin Films* (Wiley, New York 1976)
28.150 O.S. Heavens: *Optical Properties of Thin Solid Films* (Dover, New York 1965), (first published by Butterworths, London, 1955)
28.151 A. Vasicek: *Optics of Thin Films* (North Holland, Interscience, Amsterdam 1960), Reworked and supplemented ed.
28.152 O.S. Heavens: Measurement of optical constants of thin films. In: *Physics of Thin Films*, Vol. 2, ed. by G. Hass, R.E. Thun (Academic, San Diego 1964) p. 193
28.153 O.S. Heavens: Optical properties of thin films. In: *Encyclopaedic Dictionary of Physics*, *Supplementary*, Vol. 3, ed. by J. Thewlis (Pergamon, Oxford 1969) p. 412
28.154 A. Yariv, P. Yeh: *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (Wiley, Hoboken 2002)
28.155 S. Visnovsky: *Optics in Magnetic Multilayers and Nanostructures* (CRC Press, Boca Raton 2006)
28.156 H.A. Macleod: Thin film optical devices. In: *Passive and Active Thin Film Devices*, ed. by T.J. Coutts (Academic, New York 1978) p. 321
28.157 A. Thelen: Design of multilayer interference filters. In: *Physics of Thin Films*, Vol. 5, ed. by G. Hass, R.E. Thun (Academic, San Diego 1969) p. 47
28.158 M. Kildemo, O. Hunderi, B. Drevillon: J. Opt. Soc. Am. A **14**, 931 (1997) doi:10.1364/JOSAA.14.000931
28.159 F. Abelès: Optical properties of metallic thin films. In: *Physics of Thin Films*, Vol. 6, ed. by M.H. Francombe, R.W. Hoffman (Academic, San Diego 1971) p. 151
28.160 P.H. Berning: Theory and calculations of optical thin films. In: *Physics of Thin Films*, Vol. 1, ed. by G. Hass (Academic, San Diego 1963) p. 69
28.161 J.T. Cox, G. Hass: Antireflection coatings for optical and infrared optical materials. In: *Physics of Thin Films*, Vol. 2, ed. by G. Hass, R.E. Thun (Academic, San Diego 1964) p. 239
28.162 G. Hass, M.H. Francombe, R.W. Hoffman: Metal-dielectric interference filters. In: *Physics of Thin Films*, Vol. 9, ed. by G. Hass, M.H. Francombe, R.W. Hoffman (Academic, San Diego 1977) p. 74
28.163 J.M. Eastman: Scattering in all-dielectric multilayer bandpass filters and mirrors for lasers. In: *Physics of Thin Films*, Vol. 10, ed. by G. Hass, M.H. Francombe (Academic, San Diego 1978) p. 167
28.164 G. Hass, J.B. Heaney, W.R. Hunter: Reflectance and preparation of front surface mirrors for use at various angles of incidence from the ultraviolet to the far infrared. In: *Physics of Thin Films*, Vol. 12, ed. by G. Hass, M.H. Francombe, J.L. Vossen (Academic, San Diego 1982) p. 2
28.165 J. Tauc, A. Menth: J. Non-Cryst. Solids **8**(10), 569 (1972)
28.166 J.N. Zemel: Transport phenomena in hetero-epitaxial semiconductor films. In: *The Use of Thin Films in Physical Investigations*, ed. by J.C. Anderson (Academic, New York 1966) p. 319
28.167 J.J. Thomson: Proc. Cambridge Phil. Soc. **11**, 120 (1901)
28.168 K. Fuchs: Proc. Cambridge Phil. Soc. **34**, 100 (1938)
28.169 A.C.B. Lovell: Proc. R. Soc. A **157**, 311 (1936)
28.170 D.S. Campbell: The electrical properties of single-crystal metal films. In: *The Use of Thin Films in Physical Investigations*, ed. by J.C. Anderson (Academic, New York 1966) p. 299
28.171 E.H. Sondheimer: Adv. Phys. **1**, 1 (1952)
28.172 M.S.P. Lucas: J. Appl. Phys. **36**, 1632 (1965)
28.173 H.J. Juretschke: Surf. Sci. **2**, 40 (1964)
28.174 H.J. Juretschke: J. Appl. Phys. **37**, 435 (1966)
28.175 D.C. Larson: Size-dependent electrical conductivity in thin metal films and wires. In: *Physics of Thin Films*, Vol. 6, ed. by M.H. Francombe, R.W. Hoffman (Academic, San Diego 1971) p. 81
28.176 A.F. Mayadas, M. Shatzkes: Phys. Rev. B **1**, 1382 (1970)
28.177 C.R. Tellier, A.J. Tosser: *Size Effects in Thin Films* (Elsevier Science, Amsterdam 1982)
28.178 J.S. Chawla, X.Y. Zhang, D. Gall: J. Appl. Phys. **110**, 043714 (2011)
28.179 E.E. Mola, J.M. Heras: Thin Solid Films **18**, 137 (1973)
28.180 T.J. Coutts: Electrical properties and applications of thin metallic and alloy films. In: *Active and Passive Thin Film Devices*, ed. by T.J. Coutts (Academic, New York 1978) p. 57
28.181 E.E. Mola, J.M. Heras: Electrocomp. Sci. Technol. **1**, 77 (1974)
28.182 Y. Namba: Jpn J. Appl. Phys. **9**, 1327 (1970)
28.183 N. Trivedi, N.W. Ashcroft: Phys. Rev. B **38**, 12298 (1988)
28.184 U. Jacob, J. Vancea, H. Hoffmann: Phys. Rev. B **41**, 11852 (1990)
28.185 H. Hoffmann, J. Vancea: Thin Films **85**, 147 (1981)
28.186 H.-D. Liu, Y.-P. Zhao, G. Ramanath, S.P. Murarka, G.-C. Wang: Thin Solid Films **384**, 151 (2001)
28.187 J.G. Simmons: J. Phys. D **4**, 613 (1971)
28.188 E.H. Rhoderick: *Metal–Semiconductor Contacts* (Clarendon, Oxford 1978)
28.189 D.R. Lamb: *Electrical Conduction Mechanisms in Thin Insulating Films* (Metheun, London 1967)
28.190 A. Rose: Phys. Rev. **97**, 1538 (1955)
28.191 T. Budinas, P. Mackus, A. Smilga, J. Vivvakas: Phys. Status Solidi **31**, 375 (1969)
28.192 R. Zuleeg: Solid State Electron. **6**, 645 (1963)

28.193 R. Zuleeg, R.S. Muller: Solid State Electron. **7**, 575 (1964)
28.194 B.M. Basol, O.M. Stafsudd: Solid State Electron. **24**, 121 (1981)
28.195 V.S. Dharmadhikari: Int. J. Electron. **54**, 787 (1983)
28.196 R. Glew: Thin Solid Films **46**, 59 (1977)
28.197 R.K. Pandey, R.B. Gore, A.J.N. Rooz: J. Phys. D **20**, 1059 (1987)
28.198 R.D. Gould, A.K. Hassan: Thin Solid Films **193/194**, 895 (1990)
28.199 A.K. Hassan, R.D. Gould: J. Phys. Condens. Matter **1**, 6679 (1989)
28.200 T.S. Shafai, R.D. Gould: Int. J. Electron. **73**, 307 (1992)
28.201 A.K. Hassan, R.D. Gould: Int. J. Electron. **74**, 59 (1993)
28.202 S. Gravano, A.K. Hassan, R.D. Gould: Int. J. Electron. **70**, 477 (1991)
28.203 R.D. Gould, M.S. Rahman: J. Phys. D **14**, 79 (1981)
28.204 M.A. Lampert: Rep. Prog. Phys. **27**, 329 (1964)
28.205 R.D. Gould, B.A. Carter: J. Phys. D **16**, L201 (1983)
28.206 B.B. Ismail, R.D. Gould: Phys. Status Solidi (a) **115**, 237 (1989)
28.207 K.N. Sharma, K. Barua: J. Phys. D **12**, 1729 (1979)
28.208 A.O. Oduor, R.D. Gould: Thin Solid Films **270**, 387 (1995)
28.209 A.O. Oduor, R.D. Gould: Thin Solid Films **317**, 409 (1998)
28.210 A. Sussman: J. Appl. Phys. **38**, 2738 (1967)
28.211 R.D. Gould: J. Appl. Phys. **53**, 3353 (1982)
28.212 R.D. Gould: J. Phys. D **9**, 1785 (1986)
28.213 R.D. Gould, B.B. Ismail: Int. J. Electron. **69**, 19 (1990)
28.214 S.S. Ou, O.M. Stafsudd, B.M. Basol: Thin Solid Films **112**, 301 (1984)
28.215 G.M. Delacote, J.P. Fillard, F.J. Marco: Solid State Commun. **2**, 373 (1964)
28.216 M.A. Lampert, P. Mark: *Current Injection in Solids* (Academic, New York 1970)
28.217 M. Ieda, Y. Suzuki, M. Nakgawa, T. Mizutani: IEEE Trans. Electr. Insul. **25**, 599 (1990)
28.218 R.D. Gould, C.J. Bowler: Thin Solid Films **164**, 281 (1988)
28.219 A. Servini, A.K. Jonscher: Thin Solid Films **3**, 341 (1969)
28.220 A.K. Jonscher: Thin Solid Films **1**, 213 (1967)
28.221 A.K. Jonscher, A.A. Ansari: Philos. Mag. **23**, 205 (1971)
28.222 S.M. Sze: J. Appl. Phys. **38**, 2951 (1967)
28.223 H. Murray, A. Tosser: Thin Solid Films **36**, 247 (1976)
28.224 A. Piel, H. Murray: Thin Solid Films **44**, 65 (1977)
28.225 C. Canali, F. Nava, G. Ottaviani, K. Zanio: Solid State Commun. **13**, 1255 (1973)
28.226 S. Gogoi, K. Barua: Thin Solid Films **92**, 227 (1982)
28.227 R.D. Gould, B.B. Ismail: Vacuum **50**, 99 (1998)
28.228 M. Din, R.D. Gould: Thin Solid Films **340**, 28 (1999)
28.229 A.K. Hassan, R.D. Gould: Int. J. Electron. **69**, 11 (1990)
28.230 A. Ahmad, R.A. Collins: Phys. Status Solidi (a) **126**, 411 (1991)
28.231 J. Kašpar, I. Emmer, R.A. Collins: Int. J. Electron. **76**, 793 (1994)
28.232 N.F. Mott: *Metal–Insulator Transitions* (Taylor and Francis, London 1974)
28.233 H. Böttger, V.V. Bryksin: *Hopping Conduction in Solids* (VCH, Weinheim 1985)
28.234 A.H. Clark: Phys. Rev. **154**, 750 (1967)
28.235 P.A. Walley, A.K. Jonscher: Thin Solid Films **1**, 367 (1968)
28.236 C.J. Adkins, S.M. Freake, E.M. Hamilton: Philos. Mag. **22**, 183 (1970)
28.237 R.D. Gould, B.B. Ismail: Phys. Status Solidi (a) **134**, K65 (1992)
28.238 D.S.H. Chan, A.E. Hill: Thin Solid Films **35**, 337 (1976)
28.239 J. Mycielski: Phys. Rev. **123**, 99 (1961)
28.240 J. Le Moigne, R. Even: J. Chem. Phys. **83**, 6472 (1985)
28.241 A. Ahmad, R.A. Collins: Thin Solid Films **217**, 75 (1992)
28.242 S.A. James, A.K. Ray, J. Silver: Phys. Status Solidi (a) **129**, 435 (1992)
28.243 A.M. Saleh, R.D. Gould, A.K. Hassan: Phys. Status Solidi (a) **139**, 379 (1993)
28.244 R.D. Gould, A.K. Hassan: Thin Solid Films **223**, 334 (1993)
28.245 S.I. Shihub, R.D. Gould: Thin Solid Films **254**, 187 (1995)
28.246 J.G. Simmons: J. Appl. Phys. **34**, 1793 (1963)
28.247 J.G. Simmons: Phys. Rev. **166**, 912 (1968)
28.248 R.H. Fowler, L.W. Nordheim: Proc. R. Soc. A **119**, 173 (1928)
28.249 J.C. Fisher, I. Giaever: J. Appl. Phys. **32**, 172 (1961)
28.250 R. Holm: J. Appl. Phys. **22**, 569 (1951)
28.251 D. Meyerhofer, S.A. Ochs: J. Appl. Phys. **34**, 2535 (1963)
28.252 S.R. Pollack, C.E. Morris: J. Appl. Phys. **35**, 1503 (1964)
28.253 J.G. Simmons: J. Appl. Phys. **34**, 2581 (1964)
28.254 M. Lenzlinger, E.H. Snow: J. Appl. Phys. **40**, 278 (1969)
28.255 S.M. Sze: *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York 1981)
28.256 A.K. Hassan, R.D. Gould: J. Phys. D **22**, 1162 (1989)
28.257 D.V. Geppert: J. Appl. Phys. **33**, 2993 (1962)
28.258 J.G. Simmons: Electronic conduction through thin insulating films. In: *Handbook of Thin Film Technology*, ed. by L.I. Maissel, R. Glang (McGraw–Hill, New York 1970), Chap. 14
28.259 P.R. Emtage, W. Tantraporn: Phys. Rev. Lett. **8**, 267 (1962)
28.260 S.R. Pollack: J. Appl. Phys. **34**, 877 (1963)
28.261 W.E. Flannery, S.R. Pollack: J. Appl. Phys. **37**, 4417 (1966)
28.262 M. Fusstöss-Wegner: Thin Solid Films **36**, 89 (1976)
28.263 B.Sh. Barkhalov, Y.A. Vidadi: Thin Solid Films **40**, L5 (1977)
28.264 A. Wilson, R.A. Collins: Sens. Actuators **12**, 389 (1987)

29. Thick Films

Neil White

Thick film technology is an example of one of the earliest forms of microelectronics-enabling technologies and it has its origins in the 1950s. At that time it offered an alternative approach to printed circuit board technology and the ability to produce miniature, integrated, robust circuits. It has largely lived in the shadow of silicon technology since the 1960s. The films are deposited by screen printing (stenciling), a graphic reproduction technique that can be dated back to the great Chinese dynasties of around a thousand years ago. Indeed, there is evidence that even early Palaeolithic cave paintings from circa 15 000 BC may have been created using primitive stenciling techniques. With the advent of surface-mounted electronic devices in the 1980s, thick film technology again became popular because it allowed the fabrication of circuits without through-hole components.

This chapter will review the main stages of the thick film fabrication process and discuss some of the commonly used materials and substrates. It will highlight the way in which the technology can be used to manufacture hybrid microelectronic circuits. The latter stages of the chapter will demonstrate how the technology has evolved over the past twenty years or so to become an important method in the production of solid state sensors.

29.1 **Thick Film Processing** 708
29.1.1 Screen Printing 708
29.1.2 The Drying and Firing Process 709

29.2 **Substrates** 710
29.2.1 Alumina 710
29.2.2 Stainless Steel 710
29.2.3 Polymer Substrates 710

29.3 **Thick Film Materials** 711
29.3.1 Conductors 711
29.3.2 Resistors 712
29.3.3 Dielectrics 713
29.3.4 Polymer Thick Films 713

29.4 **Components and Assembly** 714
29.4.1 Passive Components 714
29.4.2 Active Components 714
29.4.3 Trimming 715
29.4.4 Wire Bonding 716
29.4.5 Soldering of Surface-Mounted Components 716
29.4.6 Packaging and Testing 717

29.5 **Sensors** 717
29.5.1 Mechanical 718
29.5.2 Thermal 719
29.5.3 Optical 719
29.5.4 Chemical 719
29.5.5 Magnetic 720
29.5.6 Actuators 720

References 720

The term *thick film* is often misinterpreted, and so it is worth noting from the outset that it has little to do with the actual thickness of the film itself. The preferred definition encompasses the fabrication process, namely screen printing, used to deposit the films. The typical range of thicknesses for thick film layers is $0.1-100\,\mu m$. Screen printing is one of the oldest forms of graphic art reproduction and involves the deposition of an ink (or paste) onto a base material (or substrate) through the use of a finely-woven screen with an etched pattern of the desired geometry. The process is commonly used for the production of graphics and text onto items such as T-shirts, mugs, pencils, textiles and so on, and is very similar to that used for microelectronic thick films. The degree of sophistication for the latter is, however, significantly greater, resulting in high-quality, reproducible films for use in a variety of electronic systems.

The technology used to manufacture thick film hybrid microelectronic circuits was introduced in the 1950s. Such circuits typically comprised semiconductor devices, monolithic integrated circuits, discrete passive components and the thick films themselves [29.1, 2]. In the early days of the technology, the thick films

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_29

were mainly resistors, conductors or dielectric layers. Evidence of thick film circuits can still be found in many of today's commercial devices such as televisions, calculators and telephones. The use of thick film technology was overshadowed in the 1960s by the impact of silicon technology. It found popularity again in the 1980s as a result of the advent of surface-mounted devices, which can be attached to circuit boards using screen-printed solder layers. Nowadays, thick films are not only used in hybrid circuits but also in advanced solid state sensors and actuators, as we shall see later. In broad terms, thick films can be classified as either *cermets* (ceramic/metallics), requiring high-temperature processing, or *polymer* thick films, which are cured at significantly lower temperatures. Polymer thick films will be covered in more detail in Sect. 29.3.4.

29.1 Thick Film Processing

Thick film technology is sometimes referred to as an *additive* technology in that the layers are built up in sequence without the need to remove (or subtract) parts of the film by techniques such as etching. Compare this with, say, a standard printed circuit board (PCB), where the conductive tracks are formed by selectively etching away the undesired areas (gaps) from a continuous copper layer. In the early days of the technology, the ability to fabricate together components made with different enabling technologies to produce a *hybrid* circuit was seen as a way of opening up a whole new field of electronics. This is still one of the most endearing features of the technology today.

29.1.1 Screen Printing

This is the distinguishing feature of thick film technology and is the method of depositing the desired films. The fundamental aspects of the process, traditionally recognized as an art, have been modernized and updated to provide a scientific tool for microelectronic technology. Figure 29.1 illustrates the main aspects of the screen printing process [29.3].

The screen fabric is permanently attached to the screen frame, which is firmly held within the screen printing machine. The distance between the screen and the substrate is typically around 0.5 mm, although the precise gap size is dependent on the overall screen dimensions. The substrate is held in place in the holder by either a vacuum or a mechanical clamp. The position of the screen can be finely adjusted to ensure good registration between consecutive layers. The paste is applied to the upper surface of the screen and the flexible *squeegee* is pulled across the screen over the ink, which is forced through the open areas of the screen mesh. At a point immediately behind the squeegee, the screen peels away from the substrate and, due to the surface tension between the ink and the substrate, leaves a deposit of the paste in the desired pattern on the substrate.

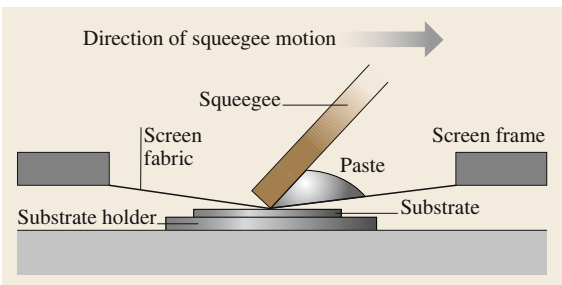


Fig. 29.1 The basic screen printing process

The squeegee is a flexible blade whose function is to transfer the ink through the screen and onto the substrate. It is usually made from materials such as polyurethane or neoprene. The squeegee pressure is adjustable and facilitates accurate and repeatable print thicknesses. The pastes generally exhibit pseudoplastic behavior in that the viscosity varies with the applied shear force. This is a necessary property because the ink must have minimum viscosity to ensure good transfer though the screen, but it also must become more viscous after printing to provide a good definition of the film.

The function of the screen is to define the pattern of the printed film and also to control the amount of paste being deposited. The screens used in graphic artwork are generally made of silk and the process is often termed silk screen printing. For microelectronic circuits, however, the screens tend to be made from polyester, nylon or stainless steel. The resolution of the printed line widths is largely determined by the mesh count (in lines per inch or centimeter) and the mesh filament diameter. Nylon has very good elastic properties but relatively poor resilience. Stainless steel screens produce excellent line definition and durability and are particularly well suited to flat substrates. Polyester screens have the highest resilience and offer a long lifetime and low squeegee wear. Figure 29.2 shows a plain weave pattern on a typical thick film screen. An ultraviolet (UV)-sensitive emulsion covers the underside of the screen. The desired

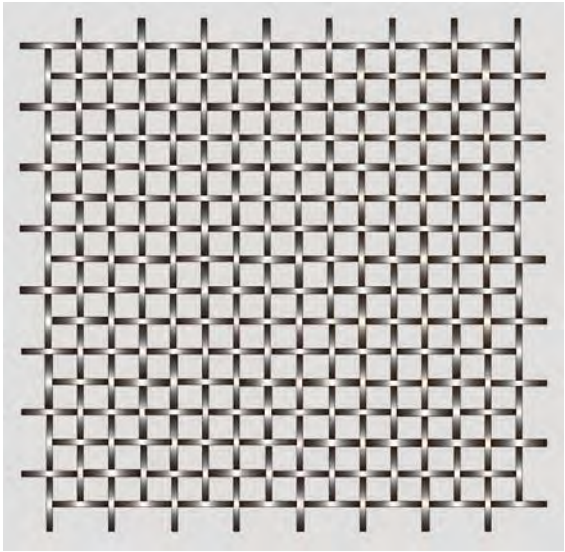


Fig. 29.2 A plain weave screen mesh

pattern is formed onto screen emulsion using a photographic process. The screen printing process can be used to deposit both cermet and polymer thick films; the processing steps that follow are, however, different.

29.1.2 The Drying and Firing Process

The final form of a cermet thick film is a fired, composite layer that is firmly attached to the substrate. Essentially, there are three main stages in the production process: screen printing, drying and firing. At each of these stages, the film is in a slightly different state. Commercial thick film pastes are purchased from the manufacturer in plastic jars and are similar to standard printing inks in many respects. The three main components in a thick film paste are:

- The active material
- A glass frit
- An organic vehicle.

The active material is a finely ground powder with a typical particle size of a few microns. A conductor paste, for example, will contain a precious metal or metal alloy. The glass frit serves as a binder that holds the active particles together and also bonds the film to the substrate. The organic vehicle is necessary to give the paste the correct viscosity for screen printing. It usually contains a resin dissolved in a solvent together with a surfactant that ensures good dispersion of the solid particles.

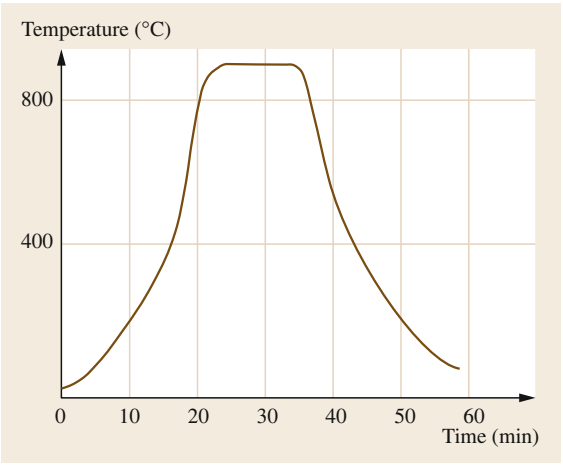


Fig. 29.3 A typical thick film firing profile with a peak temperature of 850 °C

After screen printing, it is usual to allow the film to stand in air for a few minutes to let the paste level off. The film is then dried in an infrared belt drier or a conventional box oven at temperatures up to 150 °C. The purpose of drying is to remove the organic solvents so that the film and substrate can be handled during further processing steps. After drying, the films then proceed to the firing stage. In some circumstances, however, they may be overprinted with another thick film layer if the nature of the film allows this.

The high-temperature firing cycle performs three main functions: the remaining organic carrier is removed from the film, the electrical properties are developed, and the film is bonded to the substrate. In order to achieve these, it is necessary to subject the films to temperatures of up to 1000 °C in a moving belt furnace. Inside the furnace there are several heating zones, which can be independently controlled to a profile specified by the paste manufacturer. Figure 29.3 shows a typical firing profile that is used with commercial thick film resistor pastes. The substrates enter the furnace on a metal chain mesh belt at ambient temperature and slowly (at a rate of a few centimeters per minute) travel through the heating zones, which have their temperatures set to form the desired profile. The substrates remain at the peak temperature for about ten minutes and the electrical properties of the film are formed during this phase. Most thick film materials are fired in a clean, filtered air atmosphere, as this produces high-quality films with repeatable characteristics. Occasionally, however, it is necessary to provide an inert atmosphere such as nitrogen, so that materials such as copper can be processed.

29.2 Substrates

The main functions of the substrate are to provide mechanical support and electrical insulation for the thick films and hybrid circuits. Some of the main considerations for selecting substrates are listed below:

- Dielectric constant: This determines the capacitance associated with different elements fabricated onto the substrate. The dielectric strength will also determine the breakdown properties of the substrate.
- Thermal conductivity: Substrates with a high thermal conductivity can be used in applications where the circuit generates significant amounts of heat.
- Thermal coefficient of expansion (TCE): In general terms, the TCE of the substrate should be closely matched to the thick film materials and other components mounted on it. In some cases this cannot be assured, and in this case the consequences (in terms of thermal strains) must be fully considered.

The main substrate materials used in thick film technology are the ceramic materials alumina (Al_2O_3), beryllia (BeO) and aluminium nitride (AlN). Enamelled or insulated stainless steel substrates are sometimes used in some applications. Silicon has also been used in specialist transducer applications. Alumina, however, is the most common substrate material and it possesses desirable physical and chemical properties in addition to providing an economical solution. Alumina of 96% purity is used in the vast majority of worldwide commercial circuits. The remaining 4% weight fraction of the content is made up of magnesia and silica, which improve the densification and electrical properties. Beryllia has a high thermal conductivity and is used in applications where rapid heat removal is required. It is, however, a very toxic material and is therefore only used in limited application areas. Aluminium nitride is, essentially, an alternative for beryllia, with a high thermal conductivity and also improved mechanical properties such as higher flexural strength. Insulated stainless steel substrates are sometimes used in applications where a high thermal dissipation and mechanical ruggedness are required. They are particularly well suited to mechanical sensor applications.

29.2.1 Alumina

Alumina substrates are manufactured by blending alumina powder, with an average particle size of around 1 μm, together with small amounts of silica, magnesia and calcia. These are either ball- or roll-milled for about 10 h with lubricants, binders and solvents that ensure thorough mixing. Most thick film substrates are less than one millimeter thick, and the preferred method of fabrication is sheet casting. A slurry is allowed to flow out onto a smooth belt, and it passes under a metal doctor blade which controls the resultant thickness. The material is then dried in air to remove the solvent and, at this stage, it is sometimes referred to as the *green state* because of its color. The substrates are then fired in a kiln for at least 12 h. A peak temperature of around 1500 °C ensures that the materials are properly sintered. During firing, the substrates can shrink by up to 20%, and this needs to be taken into consideration for the formation of the final substrate. The surface finish can be improved by coating the surface with a thin, glassy layer (glazing), which is done as an additional step at a lower temperature.

29.2.2 Stainless Steel

Stainless steel is a strong, elastic material with a relatively high thermal conductivity. Being a good electrical conductor, the steel must be coated with an insulating layer before it can be used as a substrate for thick film circuits. Porcelain enamelled steel substrates are made by coating a stainless steel plate with a glassy layer between 100 and 200 μm in thickness. The steel is enamelled either by dipping, electrostatic spraying or electrophoretic deposition of a low-alkali glass and subsequent firing at several hundred degrees Celsius. Some commercial paste manufacturers produce an insulating dielectric thick film ink that can be screen printed directly onto various types of stainless steel. The substrates are fired at a temperature of around 900 °C. The paste contains a devitrifying glass that does not recrystallize on further firings and therefore provides compatibility with other standard thick film materials. Insulated stainless steel substrates also offer the advantages of having a built-in ground plane (the steel itself) and excellent electromagnetic and electrostatic shielding properties. It is also possible to machine the substrate, using conventional workshop facilities, prior to the circuit fabrication.

29.2.3 Polymer Substrates

In some applications it is desirable to fabricate a circuit onto a flexible substrate; evidence of these can be found in mobile telephones, calculators and notebook computers. Cermet thick film materials are not compatible with flexible substrates and hence special polymer thick film materials are used (Sect. 29.3.4). Polymer substrate materials are mainly based on polyesters, polycarbonates and polyimide plastics. The maximum processing temperature is usually limited to around 200 °C.

29.3 Thick Film Materials

29.3.1 Conductors

Thick film conductors are the most widely used material in thick film hybrid circuits. Their main function is to provide interconnection between the components in the circuit. For a multilayered circuit, the conductor tracks are separated by dielectric layers and connection between each layer is achieved with metallized vias. Conductors are also used to form attachment pads for surface-mounted components such as integrated circuits or discrete passive components (resistors, capacitors and inductors). They can also be used as bond pads for naked dice, which may be attached directly to the thick film circuit. Another function of conductors is to provide the terminations for thick film resistors. With such a diverse range of applications, it is no surprise that a wide range of conductor materials is available.

The characteristics of thick film conductors are dependent upon the composition of the functional phase of the paste. Typically, these comprise finely divided particles of precious metals such as silver, gold, platinum or palladium. Base metals such as aluminium, copper, nickel, chromium, tungsten or molybdenum are also used. The particle size, distribution and shape also have an effect on the electrical and physical properties of the fired film.

The resistance of a conductor film is given by

$$R = \frac{\rho l}{wt} ,$$

where ρ is the bulk resistivity (Ω cm), l and w are the length and width of the conductor respectively, and t is the fired thickness of the film. The term *sheet resistivity* is often used for thick films and is defined as the bulk resistivity at a given thickness (ρ/t), expressed in ohms per square ($\Omega/\square$). This is convenient because the resistance of the track can then be calculated by simply multiplying the sheet resistivity by the aspect ratio (l/w) of the film. Table 29.1 summarizes some of the most common metals and metal alloys used in thick film conductors together with their sheet resistivities.

Table 29.1 Most of the common metals and metal alloys used in thick film conductors and their sheet resistivities

Metallurgy	Sheet resistivity (m$\Omega/\square$)
Silver (Ag)	1–3
Gold (Au)	3–5
Copper (Cu)	2–3
Silver/palladium (Ag/Pd)	10–50
Gold/palladium (Au/Pd)	10–80
Gold/platinum (Au/Pt)	50–100

Silver Conductors

Silver pastes were one of the earliest thick film conductors to be developed. They possess good bond strength and high conductivity. There are, however, several disadvantages to using pure silver which prevent it being widely used in many applications. These include:

- Poor leach resistance to solder
- Oxidation in air over time
- Susceptible to electromigration in the presence of moisture, elevated temperature and bias voltage.

Silver/Palladium Conductors

The alloy of silver and palladium is the most common type of thick film conductor and is the alloy most widely used in the hybrid circuit industry. It can be used for interconnecting tracks, attachment pads and resistor terminations, but is generally not recommended for wiring bonding pads. Silver/palladium conductors overcome many of the problems associated with pure silver, and low-migration formulations are available from several commercial suppliers.

Gold Conductors

Gold pastes have a high conductivity and are mainly used in applications where high reliability is required. Gold is a particularly good material for wire bonding pads, although it has relatively poor solderability. Gold is a precious material and hence very expensive; it is therefore not used for general purpose applications and is limited to those areas that can justify the higher costs.

Copper Conductors

Copper thick films have to be processed in an inert atmosphere; nitrogen is typically used in this case. Copper is widely used as the conductor material for printed circuit boards owing to its good electrical conductivity and solderability. These features are also applicable to thick film copper conductors, which also offer ability to handle larger currents than most other thick film conductors.

Platinum Conductors

Not too surprisingly, perhaps, these are the most expensive of all the commercial thick film conductor materials. Platinum films have a very high resistance to solder leaching and exhibit similar electrical properties to those of the bulk material: a linear, well-defined temperature coefficient of resistance (TCR). Platinum films are therefore used in specialist applications such as heaters, temperature sensors and screen-printed chemical sensors.

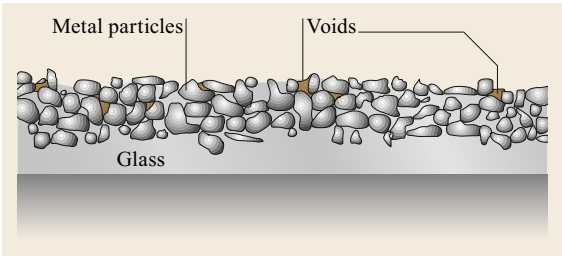


Fig. 29.4 Idealized cross-sectional view of a cermet thick film conductor

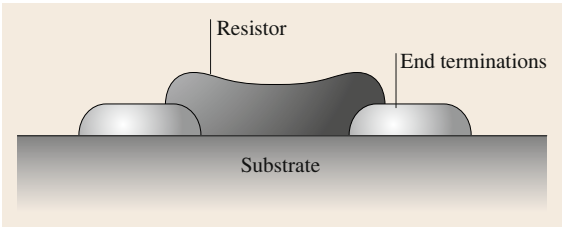


Fig. 29.5 An idealized cross-section of a thick film resistor

Gold Alloy Conductors (Au/Pd, Au/Pt)

These have good bond strength, good solder leach resistance and are relatively easy to solder. The conductivity of the gold alloys tends to be inferior to that of the other type of thick film conductor, but they can be used with both ultrasonic (aluminium) and thermosonic (gold) wire bonds.

Once the conductor has been fired, a composite structure is formed. Figure 29.4 shows a simplified view of the cross-section of a fired thick film conductor on a substrate. Many of the metallic particles have joined together to make a continuous chain. The glass is mainly evident at the interface between the bulk of the film and the substrate. The presence of voids (gaps), both within the film and on its surface, can also be noted.

29.3.2 Resistors

Resistor inks consist of a mixture of the three main phases listed earlier. The relative proportions of the active material to glass frit can have a dramatic effect on the electrical properties of the fired film. The earliest resistors were made from materials such as carbon, silver and iron oxide and were found to suffer from poor long-term stability, unpredictability of fired resistivity, and unacceptably high temperature coefficients. Modern thick film resistor systems are mainly based on ruthenium dioxide (RuO_2) and have much-improved characteristics. This material has a high conductivity

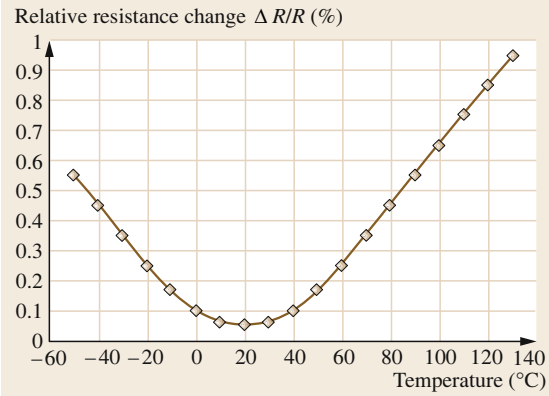


Fig. 29.6 A plot of relative resistance change against temperature for a thick film resistor

and is extremely stable at high temperatures. Within the resistive paste, the conductive phase comprises submicron particles, and these are mixed with larger glass particles (several microns in diameter). Various additives are added to the formulation to improve the stability and electrical properties of the fired film. The nature of these additives is proprietary knowledge of the paste manufacturer.

Figure 29.5 shows an idealized cross-section of a thick film resistor fabricated onto a substrate. The end terminations are first printed and fired and then the resistor is formed in the same manner, allowing for a slight overlap on the terminations so that small misalignments during the printing of the resistor can be accounted for. The actual surface profile of a real resistor would show a nonuniform thickness across its length.

Commercial thick film resistor pastes are available in a range of sheet resistivities from $1-10^9\,\Omega/\square$, and the value of the fired resistor is determined by the selected sheet resistivity and the ratio of the length to the width, as previously described for conductors. For example, a $10\,k\Omega$ resistor could be made by printing a square ($l = w$) of $10\,k\Omega/\square$ resistor paste. Similarly, a $5\,k\Omega$ resistor requires an aspect ratio of $1:2$ ($l:w$) if printed with the same paste. Note that the absolute value of a thick film resistor is not limited to a preferred value, as is the case for discrete resistors.

Figure 29.6 shows the temperature dependence of resistance for a typical thick film resistor. The shape of the plot is unusual because there is a point at which the resistance is a minimum (T_{min}). In general terms, the sensitivity of a resistor to temperature is denoted by the temperature coefficient of resistance (TCR). Mathemat-

ically, this can be expressed by

$$\mathrm{TCR} = \frac{\Delta R/R}{\Delta T} ,$$

where $\Delta R/R$ is the relative change in resistance and ΔT is a small change in temperature. Metals generally exhibit an increase in resistance with increasing temperature, so the TCR is positive. A thick film resistor appears to have both a positive and negative TCR at different regions of the resistance versus temperature curve [29.4]. The point of minimum resistance occurs around room temperature, which means that the resistance is stable under normal operating conditions. It is common practice to refer to the *hot* and *cold* TCR for the regions above and below T_{min}, respectively.

The long-term stability of thick film resistors in air at room temperature is excellent, with a typical change of less than 0.2% over 10^5 h (about 100 years!). At elevated temperatures the stability is still impressive, with most resistance values changing by only 0.5% over 1000 h at 150 °C. The tolerance of fired resistors is around ±20% due to a variety of reasons, including variations in resistor thickness, the tolerance on the quoted sheet resistivity and the effects of the firing process. If accurate values of resistance are required, it is therefore necessary to trim the resistor. This is achieved by removing areas of the resistor with either an air-abrasive jet or a laser. Resistor values can only be increased by trimming, so the printed value is always designed to be less than that of the post-trimmed value.

29.3.3 Dielectrics

Dielectric pastes have four main uses in thick film hybrid circuit applications:

- Cross-over insulators for multilayer circuits
- Thick film capacitor dielectrics
- Passivation layers
- Insulation layers for stainless steel substrates.

Cross-over dielectrics comprise a ceramic material such as alumina, together with the usual glass frit and organic vehicle. These films are required to have a low dielectric constant in order to minimize capacitive coupling between the conducting tracks. They must also have a good insulation resistance and a smooth, pinhole-free surface finish.

There is an occasional requirement to fabricate thick film capacitors, although the wide availability of surface-mounted capacitors has largely precluded their need. It is also difficult to ensure that post-fired values of thick film capacitors closely match those for which they were designed. An expensive trimming process is needed to ensure that the correct valves are obtained.

Overglazes are mainly used to protect resistors from overspray during trimming and also the circuit from environmental attack during operation. The passivation layer is usually the last one to be processed and it is therefore fired at a reduced temperature in order to minimize any adverse refiring effects on other layers. Overglaze materials are therefore almost exclusively made from low-temperature glasses that fire at temperatures of 450–500 °C.

The use of insulated stainless steel substrates was mentioned earlier. The dielectric materials used for this purpose must provide a high insulation resistance (in excess of $10^{12}\,\Omega$) and also possess a high breakdown voltage (greater than 2 kV/mm). There is an inherent mismatch between the thermal coefficient of expansion of the metal substrate and that of the insulation layer, which limits the type of steels that can be used for this purpose.

29.3.4 Polymer Thick Films

The processing temperatures for polymer thick films are significantly less than those needed for cermet materials. Rather than being *fired*, polymer materials are cured at temperatures below 200 °C. One advantage of polymer thick films over cermets is a reduction in processing and material costs. As with cermet pastes, the formulation of polymer thick films comprises the active material, a polymer matrix and various solvents. The polymer matrix acts as *glue* for the active component. Three types of polymer organic composition are used in polymer thick films:

- Thermoplastic
- Thermosetting
- Ultraviolet (UV)-curable.

With thermoplastics, the required viscosity for screen printing is achieved via solvents. The polymer material is typically acrylic, polyester, urethane or vinyl. After printing, the paste is hardened by drying in a belt or box oven. These types of film have relatively poor resistance to environmental conditions and are not resistant to elevated temperatures and solvents. Thermosetting pastes have polymers that are partially cured and are typically epoxy, silicone or phenolic resin. After printing, the polymer is fully cured, providing a strong and stable matrix. Solvents are still needed to provide the correct rheology for printing. The UV curable pastes are generally used for dielectric inks and can be cured at room temperature under an ultraviolet light source.

For conductors, the most commonly used active phases are silver, copper and nickel. Owing to their poor stability at high temperatures, polymer thick film

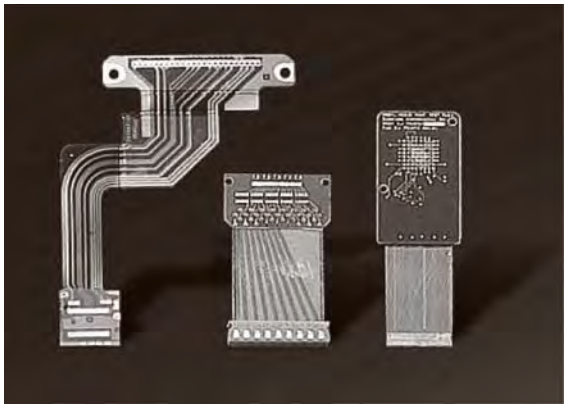


Fig. 29.7 A selection of flexible circuits. (Courtesy of Flex Interconnect Technologies, Milpitas, USA)

conductors cannot be soldered and alternative forms of attachment must be adopted. Carbon is typically used as the active material in polymer thick film resistors. The performance of these resistors is inferior to that of their cermet counterparts and they are therefore seldom used in critical applications. The dielectric pastes are similar in nature to conductors and resistors except that the conducting phase is omitted from the formulation. Some manufacturers add minerals to improve the electrical and mechanical properties of the dielectric films. Polymer thick films are a popular choice of material for disposable biosensors such as those used in the home testing of levels of glucose in human blood samples.

The process for fabricating polymer thick films is similar to that used with cermet materials. Once they have been screen printed, the layers are left to stand in air for a few minutes to ensure that the surface is level and contains no residual mesh patterns. The curing process is achieved in a box oven or an infrared belt dryer at temperatures in the range 150–200 °C. For a thermoset polymer, the higher the temperature and longer the curing time, the greater the cross-linking of the polymer chains in the matrix. This can lead to improved film stability and increased shrinkage.

An early and successful application of polymer thick film technology was the fabrication of membrane switches for keyboards. Today, examples of polymer thick film circuits can be found in many consumer products, such as mobile phones, portable computers, personal digital assistants and calculators. Figure 29.7 shows some typical flexible polymer circuits.

29.4 Components and Assembly

29.4.1 Passive Components

Passive electronic components are those that do not require an external energy source to function. Examples are resistors, capacitors and inductors. As we have seen earlier, thick film technology allows the fabrication of high-quality, stable resistors. It is also possible to add resistors to a thick film circuit in the form of an additional chip component. Such devices are available as surface-mounted devices, which do not require holes to be drilled into the circuit board. Interestingly, chip resistors are often manufactured as multiple parts using thick film techniques on ceramic substrates. These are then diced and the terminations are added. An example of a typical chip component is shown in Fig. 29.8. Resistor values can range 1 Ω to 10 MΩ, with typical tolerances of between ±1 to ±20%.

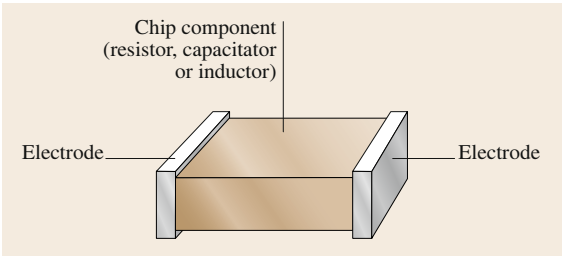


Fig. 29.8 A typical surface-mounted chip component

Planar screen-printed capacitors are rarely used in thick film hybrid circuits owing to their poor stability and high production costs. They typically comprise at least three layers (two electrodes and one dielectric layer), and trimming is often required to achieve the target value of capacitance. Chip capacitors offer improved performance at a lower cost. Capacitances in the range 1 pF to 100 μF are readily available in a range of sizes from the 0201 series (0.6 mm × 0.3 mm × 0.3 mm) to the 2220 series (5.7 mm × 5.0 mm × 3.2 mm). Chip inductors are also available (typical range: 0.1–1000 μH) in a variety of package sizes. There is also a wide choice of variable passive components such as potentiometers and variable capacitors/inductors that are currently available from many major component suppliers.

29.4.2 Active Components

Active components are those that require an external energy source to function. Examples are transistors, diodes and semiconductor integrated circuits (ICs). A wide variety of semiconductor components are avail-

able to the thick film circuit designer. Transistors and diodes, being relatively small devices, are obtained in a small plastic package with three terminals. This is known as a SOT-23 package and is shown in Fig. 29.9a. Many standard integrated circuits that are available in dual-in-line (DIL) packages for through-hole printed circuit boards are also available in small outline (SO) packages for hybrid circuits. An example of a small outline device is shown in Fig. 29.9b, and these usually have between 8 and 40 pins. For devices with higher pin counts such as microprocessor, gate arrays and so on, it is usual to place the leads on all four sides of the package, as with the flatpack device shown in Fig. 29.9c. Occasionally, ICs may be obtained in the form of a naked die, without any packaging. In such cases, it is necessary to glue the chip to the board and to bond very fine wires from the chip to the board. Care has to be taken to ensure that the naked device is suitably encapsulated for use afterwards.

Other forms of IC include flip chips, which are essentially naked chips with raised connection contacts (bumps) made of solder, gold or aluminium. These are mounted by turning over the chip (flipping) and bonding directly to the substrate. Beam lead chips also examples of naked dice with either gold or aluminium leads protruding from the edge. The leads (beams) are an integral part of the chip metallization process. Such devices are usually passivated with a layer of silicon nitride during processing. Tab automated bonding (TAB) refers to a technique by which the naked chip is attached to metallized fingers on a continuous strip of film. Large quantities of devices can be produced on a single roll and the leads of the devices are welded onto the boards before the carrier film is removed. In this manner, it is possible to achieve good yields on high-density circuit populations.

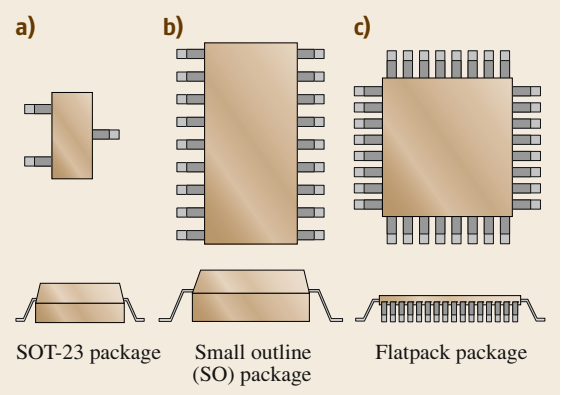


Fig. 29.9a–c Examples of surface-mounted packages **(a)** SOT-23 package **(b)** Small outline (SO) package **(c)** Flatpack package

29.4.3 Trimming

The tolerance on the printed value of thick film components such as resistors is around ±20% of the desired value. Many applications require a much tighter tolerance, and so the components need to be trimmed. The two most popular techniques used are trimming by laser or by an air-abrasive jet. Both of these are capable of producing resistors with a tolerance of ±0.1%. Laser trimming, however, is more amenable to large-scale component adjustment.

Air-abrasive trimming uses a pressurized jet of air containing a fine abrasive powder to remove a small area of the fired thick film. The diameter of the jet nozzle is around 0.5 to 1 mm. Alumina particles of average diameter 25 μm are often used as the abrasive medium. The substrate containing the component to be trimmed is held underneath the nozzle at a distance of around 0.6 mm and electrical probes are attached to the device. During trimming, the particles in the jet stream remove material from the component. The debris is removed from the substrate by a vacuum exhaust system. In the case of resistor trimming, the value can only increase because the material is being removed. Resistor values cannot be reduced by the trimming process. For this reason, resistors requiring trimming are designed to be between 25 and 30% lower than the post-trimmed value.

Laser trimming has the advantage of offering a fully automated, high-speed way of adjusting component values. The thick film is vaporized with high-energy laser beam pulses. The laser is typically a Q-switched neodymium-doped YAG (yttrium aluminium garnet) type. A single pulse removes a hole of material and a line is achieved by overlapping consecutive pulses.

Resistors are the most common component requiring trimming. Sometimes conductor tracks need to be adjusted in special circumstances. For example, a platinum conductor being used as a classic Pt 100 resistance thermometer must have a resistance of 100 Ω at 0 °C. In rare circumstances, thick film capacitors can be trimmed by removing an area of one of the plates, although the post-trimmed stability is poor.

Figure 29.10 shows some examples of different types of cut that can be used to trim resistors. The straight cut is the fastest way to trim, but it does not provide a reliable way of achieving a high accuracy. The L-cut overcomes this problem; the resistor is first trimmed straight and then the cut runs parallel to its length, providing a finer adjustment of the value. Both the serpentine and double cut require additional cuts perpendicular to an initial straight cut. The top hat structure is used in situations where a large change in resistance is needed.

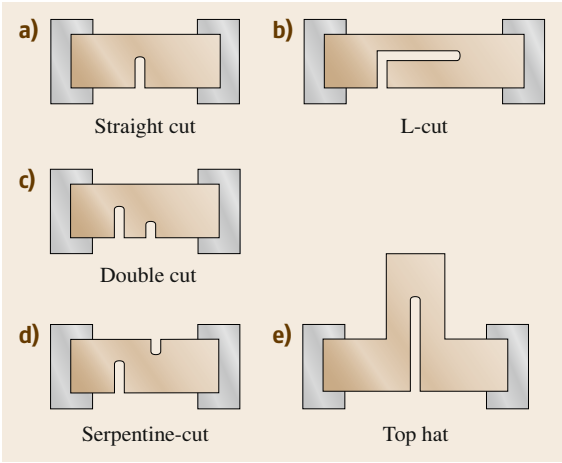


Fig. 29.10a–e Examples of trim cuts for resistors **(a)** Straight cut **(b)** L-cut **(c)** Double cut **(d)** Serpentine cut **(e)** Top hat

29.4.4 Wire Bonding

When a naked integrated circuit is needed as part of a hybrid circuit, connection must be made directly to the bond pads on the die. The chip can be attached to the substrate using an epoxy or by a gold/silicon eutectic bond. Once the chip is firmly held in position, wire bonding can commence. Three main methods are:

- Thermocompression
- Ultrasonic
- Thermosonic.

Thermocompression bonding relies on a combination of heat and pressure. The wire is usually made of gold with a diameter of around 25 μm. Gold or palladium/gold pads are deposited onto the hybrid substrate prior to bonding. The wire is fed through a ceramic capillary and a ball is formed at the end of the wire by a flame or spark discharge. A temperature of 350 °C is required for the bond and is achieved by heating either the substrate or the capillary. An epoxy chip bond cannot be used with this technique as it will soften during the bonding process. The first part of the bond is made on the aluminium bond pad on the chip; the capillary is lowered onto the pad and a force is applied to form a ball shape on the pad. The wire feeds out of the capillary and is then positioned over the desired pad on the substrate. As the capillary is lowered, a combination of heat and pressure forms the bond and the wire is then broken so that the process can be repeated for further bonds.

Ultrasonic bonding uses either a gold or aluminium wire and does not require external heat. The ultrasonic energy is supplied from a 40 kHz transducer. The combination of pressure and ultrasonic vibration causes the materials to bond together at the interface of the wire and the bond pad. This is generally a faster technique than the thermocompression method.

The final category, thermosonic bonding, is a combination of the other techniques. The substrate is heated to around 150 °C and the bond is made using the ultrasonic vibrations. This method is amenable to multilevel and multidirectional bonding and is therefore the preferred method, allowing bonding of up to 100 wires per minute.

29.4.5 Soldering of Surface-Mounted Components

Several techniques exist for the attachment of surface-mounted components to a thick film hybrid circuit. Surface-mounted components are generally much smaller than their through-hole counterparts. Of course, soldering by hand is also possible, although this is a tricky task requiring good operator skill and is often impractical because of the relatively long length of time needed.

Solder dipping requires the components to be placed on the board, either by hand or by a special pick-and-place machine. The components are fixed in position on the substrate by adding a small dot of glue and elevating the temperature to between 120 and 180 °C, which is sufficient to cure the adhesive. The board can then be dipped into a bath of molten solder at a temperature of 200 °C and then withdrawn at a sufficient rate to ensure that an adequate solder coating is obtained.

Wave soldering also requires the components to be fixed in position prior to the soldering process. Wave soldering machines were originally used for soldering through-hole components onto printed circuit boards, but they can also be used effectively with surface-mounted devices. The substrates are placed on a moving belt component side-down and initially pass through a flux bath before entering a solder bath. A wave of molten solder then flows over the substrate and creates a good joint at the desired location. This process can expose the components to a great thermal shock and it is therefore common to have a preheating phase which minimizes such effects. With both of these techniques, it is also necessary to ensure that a solder resist layer is applied to the substrate to cover all the areas that are not required to be soldered. With very densely populated circuits, there can also be a masking effect where some areas are not sufficiently coated with solder.

Reflow soldering is the preferred method of attaching surface-mounted devices. A solder cream is deposited onto the component pads either by screen

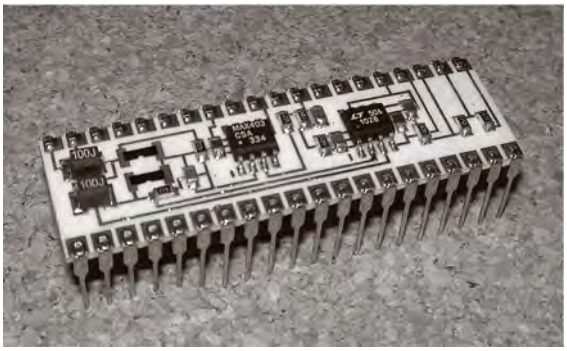


Fig. 29.11 A thick film hybrid circuit

printing or by a solder dispenser. The flux within the cream is sufficiently tacky to hold the component in place so that handling of the substrate is possible. After all of the components have been positioned on the circuit, the solder cream is dried and then reflowed. This process takes place by belt reflow, vapor phase or infrared belt system. A typical belt reflow system comprises a thermally conducting belt upon which the substrates are placed. The belt then travels through a number of heating stages, which causes the solder cream to melt (reflow). Vapor-phase soldering requires the substrates to be lowered into a vessel containing a boiling, inert fluorocarbon. The vapor condenses onto the substrate and raises the temperature uniformly to that of the liquid below. Infrared belt reflow systems are similar to those used for drying thick film materials. The substrate is placed on a wire-mesh belt, which travels through several infrared radiator zones.

29.4.6 Packaging and Testing

Thick film hybrid circuits are very versatile and offer advantages over other forms of enabling technologies. Owing to this flexibility, the circuits have a wide range of shapes and sizes and hence there is no *standard* package type. Selection of a particular form of packaging must therefore involve the consideration of issues such as:

- Protection of the circuit from harsh environmental conditions
- Protection from mechanical damage
- Avoidance of water ingress
- Electrical or mechanical connections to other parts of the system
- Thermal mismatches of different materials.

A simple way of protecting the circuit is to screen print an overglaze layer over the substrate, covering all areas of the substrate except those where components are to be added. A lead frame can then be added to the substrate to allow external connections to be made. An example of a thick film hybrid circuit (without overglaze) is depicted in Fig. 29.11. The two resistors on the left of the circuit have been trimmed and the straight cuts are visible.

Conformal coatings are often used to protect the circuit from environmental attack. These are applied in the form of either a powder or fluid. In the former case, the substrate is heated and immersed into the powder. The temperature is then increased so that the coating dries. For fluids, the substrate is dipped into the coating material and subsequently dried at a temperature of around 70 °C. Typically, the thickness of a conformal coating is between 300 and 1200 μm.

For circuits requiring operation in harsh environments, a special hermetic packaging is needed. The package can be made from ceramics, metals, ceramic/metal or glass/ceramic compositions. The hermetic seal is made by brazing, welding or glass sealing techniques. This form of packaging is often very expensive and is therefore only used in special application areas.

The final stage of the process is to test the circuit to see that its performance matches the design specification. Electrical testing can be difficult if the circuit has been coated or hermetically sealed, as physical access to components may be restricted. It is therefore usual to ensure that key test points are brought out to an external pin on the package. Environmental testing over a range of temperature and humidity may also be required in some circumstances. High-reliability circuits are often subject to a so-called burn-in phase, which involves holding the circuits at an elevated temperature for a given time to simulate the ageing process.

29.5 Sensors

Advances in the field of sensor development are greatly affected by the technologies that are used for their fabrication. The use of thick film processes as an enabling technology for modern-day sensors continues to expand. As we have already seen, the ability to produce miniaturized circuits is clearly one area in which thick film technology excels. The hybrid electronic circuitry can be integrated into the sensor housing to produce the

basis of a smart (or intelligent) sensor [29.5]. Thick film technology also offers the advantage that it can provide a supporting structure onto which other materials can be deposited, possibly using other enabling technologies [29.6].

A major contribution of the technology to sensor development, however, results from the fact that the thick film itself can act as a primary sensing element. As an example, the thick film strain gauge, described below, is merely a conventional thick film resistor that is configured in such a way as to exploit one of its physical characteristics. Commercial thick film platinum conductors can be trimmed and used as calibrated temperature sensors. Most standard pastes, however, have not been specifically developed for sensor applications and do not necessarily have optimum sensing properties. The formulation of special-purpose thick film sensor pastes is the subject of intensive research activity [29.7].

For the purpose of this text, a *sensor* is considered as being a device that translates a signal from one of the common sensing domains (mechanical, thermal, optical, chemical or magnetic) into an electrical signal. An *actuator* is a device that converts an electrical signal into one of the other domains (mainly mechanical).

29.5.1 Mechanical

In broad terms, thick film mechanical sensors are mainly based on piezoresistive, piezoelectric or capacitive techniques. Materials that exhibit a change in bulk resistivity when subjected to deformation by an external force are termed *piezoresistive*. A more common term is the strain gauge, denoting the fact that such devices produce a change in resistance when strained. The effect can be observed in standard cermet thick film resistors [29.6, 8]. The sensitivity of a strain gauge is called the gauge factor (GF) and is defined as

$$\mathrm{GF} = \frac{\Delta R/R}{\varepsilon} ,$$

where $\Delta R/R$ is the relative change in resistance and ε is the applied strain (dimensionless). The gauge factors of metal foil strain gauges and thick film resistors are around 2 and 10 respectively. The former have typical resistance values of either 120 or 350 Ω. As we have already seen, however, it is possible to produce thick film resistors with a wide range of resistance values, and this allows greater flexibility in strain gauge design. It is usual to place the strain gauges in a Wheatstone bridge configuration in order to produce a linear output analog voltage change that is proportional to the mechanical measurand. A wide range of thick film piezoresistive sensors exist, including accelerometers, pressure sensors and load cells.

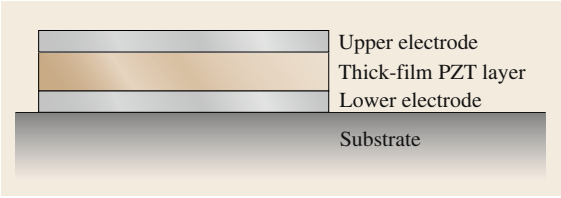


Fig. 29.12 Cross-section of a thick film piezoelectric sample

Piezoelectric materials exhibit the property of producing an electric charge when subjected to an applied mechanical force. They also deform in response to an externally applied electric field. This is an unusual effect, as the material can act as both a sensor and actuator. Certain crystals such as quartz and Rochelle salt are naturally occurring piezoelectrics, whilst others, like the ceramic materials barium titanate, lead zirconate titanate (PZT) and the polymer material polyvinylidene fluoride (PVDF), are *ferroelectric*. Ferroelectric materials are those that exhibit spontaneous polarization upon the application of an applied electric field. This means that ferroelectrics must be poled (polarized) prior to use in order to obtain piezoelectric behavior.

Thick film piezoelectrics have been made by mixing together PZT powder, a glass binder and an organic carrier [29.9]. A conducting layer is first screen printed, dried and fired onto a substrate and then several layers of the PZT film are deposited onto this lower electrode. The piezoelectric layer can be processed in a similar manner to conventional thick films. An upper electrode layer is then deposited onto the PZT in order to make a sandwich structure similar to that shown in Fig. 29.12. This is essentially a planar capacitor and the as-fired film must undergo a poling process by applying a direct current (DC) electric field of around 4 MV/m and elevating the temperature to about 120 °C.

Thick film piezoelectrics have been used in a variety of sensor and actuator applications, including accelerometers, pressure sensors [29.10], micromachined pumps [29.11], surface acoustic wave (SAW) [29.12] and resonant sensors [29.13].

The capacitance C of a parallel plate capacitor is given by

$$C = \frac{\varepsilon_0 \varepsilon_\mathrm{r} A}{d} ,$$

where ε_0 is the permittivity of free space, ε_r is the relative permittivity of the material between the electrodes, A is the area of overlap, and d is the separation of the electrodes. A mechanical sensor exhibiting a change in capacitance can be made by varying A, d or by dis-

placing the dielectric (changing ε_r). The most popular configuration is to vary d in accordance with the desired measurand. This results in a nonlinear relationship between displacement and capacitance. If the variable plate is positioned between two fixed outer electrodes, then a differential structure with a linear response can be obtained. This is a common arrangement in many types of pressure sensor.

29.5.2 Thermal

Devices that exhibit a change in resistance in accordance with variations in temperature are termed *thermoresistive*. For metals, such devices have a linear response to temperature and are known as resistance thermometers. Thermally sensitive semiconductors, typically having a nonlinear response, are termed *thermistors*. Thick film platinum conductor layers can be used as resistance thermometers and exhibit a linear TCR of around 3800 ppm/°C (ppm: parts per million), slightly lower than that of bulk platinum, which is around 4000 ppm/°C. Platinum resistance thermometers (PRTs) are often trimmed so that they have a resistance (R_0) of 100 Ω at 0 °C. Such sensors are sometimes referred to as Pt 100s and are often made of wound platinum wire. Thick film Pt 100s, fabricated onto alumina substrates, are available commercially and are considerably cheaper than the bulk versions. The following expression applies to a PRT over the linear part of the temperature characteristic (−200 to +500 °C)

$$R = R_0(1 + \alpha T) ,$$

where R is the resistance at a temperature T and α is the TCR (ppm/°C).

Thermistors are usually available in the form of discs, rods or beads comprising a sintered composite of a ceramic and a metallic oxide (typically manganese, copper or iron). Thermistor pastes are commercially available and screen-printed sensors are fabricated in a similar manner to conventional thick film resistors. The electrodes are first printed and fired onto an alumina substrate. The thermistor is then deposited across the electrodes and can be trimmed to a specific value if desired. Most thermistors have a negative temperature coefficient (NTC) of resistance; in other words their resistance decreases as the temperature increases. The resistance versus temperature relationship is of the form

$$R = R_0 \exp\left[\beta\left(\frac{1}{T} - \frac{1}{T_0}\right)\right] ,$$

where R_0 is the resistance at a reference temperature T_0 (usually 25 °C).

Positive temperature coefficient (PTC) thermistors are also available, although they are generally not as stable or repeatable as NTCs and are therefore used as simple thermal detectors rather than calibrated devices.

Another type of temperature sensor can be made by joining together two dissimilar metals (or semiconductors). If a temperature difference exists between the joined and open ends, then an open-circuit voltage can be measured between the open ends. This is known as the Seebeck effect and is the basis of a *thermocouple*. A thick film version can be made by overlapping different conductor materials on a substrate, although the thermal sensitivity is much less than traditional wire-based devices.

29.5.3 Optical

Screen-printed photosensors are probably one of the earliest examples of thick film sensors, and their use dates back to the mid 1950s. Materials that exhibit a change in electrical conductivity due to absorbed electromagnetic radiation are known as *photoconductors*. Cadmium sulfide (CdS) is an example of such a material and is notable because of its highly sensitive response in the visible range (450–700 nm). The resistances of such devices can drop from several tens of MΩ in the dark to a few tens of ohms in bright sunlight.

Thick film photoconductor pastes are not widely available commercially, but have been the subject of some research activity [29.14]. Such pastes, based on cadmium sulfide and selenide, are prepared by sintering powdered CdS or CdSe with a small amount of cadmium chloride (which acts as a flux) at a temperature of around 600 °C. This is then ground into a powder and mixed with an organic carrier to make a screen-printable paste. This can then be printed over metal electrodes and fired at a temperature of around 600 °C in air.

29.5.4 Chemical

Thick film materials have been used in a variety of chemical sensor applications for the measurement of gas and liquid composition, acidity and humidity [29.15]. The two main techniques are impedance-based sensors and electrochemical sensors. With the former method, the measurand causes a variation in resistance or capacitance, whilst the latter relies on the sensed quantity changing an electrochemical potential or current.

Impedance-based gas sensor pastes usually comprise a semiconducting metal oxide powder, inorganic

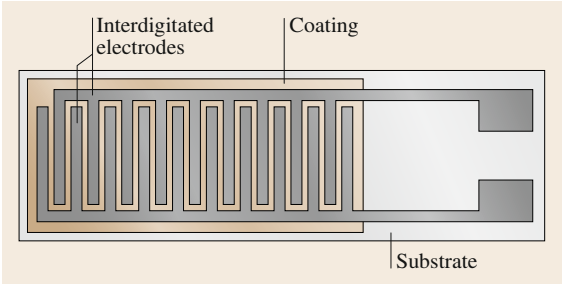


Fig. 29.13 A thick film humidity sensor

additives and organic binders [29.16]. The paste is printed over metal electrodes and a back-heated resistor on an alumina substrate. The heating element is necessary to promote the reaction between the gas being measured and the sensing layer. Figure 29.13 shows an example of a thick film sensor, without a heating element, that can be used to measure humidity. The interlocking *finger* electrodes are often referred to as interdigitated electrodes and are screen printed and fired onto an alumina substrate. A porous dielectric layer is then screen printed onto the electrodes. As the humidity increases, moisture will penetrate the surface of the dielectric layer causing a change in dielectric constant within the sensitive layer. This results in a change in capacitance between the electrodes.

Electrochemical techniques can be used to realise pH sensors. These are often used in biomedical, fermentation, process control and environmental applications. These devices often make use of a solid electrolyte which generates an electrochemical potential between two electrodes in response to the measurand.

Perhaps the most common example of a thick film chemical sensor is the disposable, polymer-based glucose sensor used in many home testing kits for diabetic patients. This illustrates how thick film sensors can offer robust, compact and cost-effective solutions to many modern-day requirements.

29.5.5 Magnetic

Some screen-printable conductors, particularly those containing nickel, exhibit a change in resistivity in response to an applied magnetic field. Such devices are referred to as *magnetoresistive* sensors. Air-fireable nickel-based conductors have been shown to exhibit a nonlinear change in resistance for a linear increase in applied magnetic field [29.17]. A peak change in resistance of around 1% can occur at an applied field of 0.1 T. Researchers have made linear and rotary displacement sensors based on thick film nickel pastes, although it should be noted that such devices are also thermoresistive and therefore the magnetic measurements need to be taken in a temperature-controlled environment.

29.5.6 Actuators

We have previously defined an actuator as a device that converts a signal from the electrical domain into one of the other signal domains. It was noted earlier that piezoelectric materials produce a mechanical stress in response to an electrical charge. Such materials can therefore be used as actuators. Thick film piezoelectric layers have been screen printed onto thin silicon diaphragms in order to form the basis of a micropump [29.11].

Photovoltaic devices convert incident optical radiation into electric current and are often termed *solar cells*. They are used to power devices such as calculators, clocks, pumps and lighting. In general terms, the output power level is proportional to the physical size of the photovoltaic cell. The device is essentially a heterojunction between n-type and p-type semiconductors. Thick film solar cells have been made comprising CdS (n-type) and CdTe (p-type) as the junction materials [29.18]. Such thick film actuators have been shown to have relatively low conversion efficiencies (between 1 and 10%).

References

29.1 R.A. Rikoski: *Hybrid Microelectronic Circuits: The Thick-Film* (Wiley, New York 1973)

29.2 M.A. Topfer: *Thick-Film Microelectronics: Fabrication, Design and Fabrication* (Van Nostrand-Reinhold, New York 1971)

29.3 P.J. Holmes, R.G. Loasby: *Handbook of Thick Film Technology* (Electrochemical Publ., Ayr 1976)

29.4 M. Prudenziati, A. Rizzi, P. Davioli, A. Mattei: Nuovo Cim. **3**, 697 (1983)

29.5 J.E. Brignell: In: *Thick-Film Sensors*, ed. by M. Prudenziati (Elsevier, Amsterdam 1994)

29.6 J.E. Brignell, N.M. White, A.W.J. Cranny: Sensor applications of thick-film technology, IEE Proc. Part I **135**(4), 77 (1988)

29.7 N.M. White, J.D. Turner: Meas. Sci. Technol. **8**, 1 (1997)

29.8 C. Canali, D. Malavisi, B. Morten, M. Prudenziati: J. Appl. Phys. **51**, 3282 (1980)

29.9 H. Baudry: Screen printing piezoelectric devices, 6th Eur. Microelectron., Bournemouth (1987) p. 456

29.10 M. Prudenziati, B. Morten, G. De Cicco: Microelectron. Int. **38**, 5 (1995)

29.11 M. Koch, N. Harris, A.G.R. Evans, N.M. White, A. Brunnschweiler: Sens. Actuat. A **70**(1–2), 98 (1998)
29.12 N.M. White, V.T.K. Ko: Electron. Lett. **29**, 1807 (1993)
29.13 S.P. Beeby, N.M. White: Sens. Actuat. A **88**, 189 (2001)
29.14 J.N. Ross: Meas. Sci. Technol. **6**, 405 (1995)
29.15 M. Prudenziati, B. Morten: Microelectron. J. **23**, 133 (1992)
29.16 G. Martinelli, M.C. Carotta: Sens. Actuat. B **23**, 157 (1995)
29.17 B. Morten, M. Prudenziati, F. Sirotti, G. De Cicco, A. Alberigi-Quaranta, L. Olumekor: J. Mater. Sci. Mater. Electron. **1**, 118 (1990)
29.18 N. Nakayama, H. Matsumoto, A. Nakano, S. Ikegami, H. Uda, T. Yamashita: Jpn. J. Appl. Phys. **19**, 703 (1980)

Part D Materials

Part D Materials for Optoelectronics and Photonics

30 III-V Ternary and Quaternary Compounds
Sadao Adachi, Gunma, Japan

31 Group III Nitrides
Romualdo A. Ferreyra, Kyoto, Japan
Congyong Zhu, Chandler, USA
Ali Teke, Balikesir, Turkey
Hadis Morkoç, Richmond, USA

32 Electron Transport Within III-V Nitride Semiconductors
Stephen K. O'Leary, Kelowna, Canada
Poppy Siddiqua, Kelowna, Canada
Walid A. Hadi, Windsor, Canada
Brian E. Foutz, Charlottesville, USA
Michael S. Shur, New York, USA
Lester F. Eastman†

33 II-IV Semiconductors for Optoelectronics: CdS, CdSe, CdTe
Minoru Isshiki, Sendai, Japan
Jifeng Wang, Tokyo, Japan

34 II-VI Narrow Bandgap Semiconductors: Optoelectronics
Ian M. Baker, Southampton, UK

35 Optoelectronic Devices and Materials
Stephen J. Sweeney, Guildford, UK
Jayanta Mukherjee, Durham, UK

36 Liquid Crystals
Geoffrey Luckhurst, Southampton, UK
David Dunmur, Southampton, UK

37 Organic Photoconductors
David S. Weiss, Rochester, USA
Martin Abkowitz†

38 Luminescent Materials
Andy Edgar, Wellington, New Zealand

39 Nano-Engineered Tunable Photonic Crystals
Harry E. Ruda, Toronto, Canada
Naomi Matsuura, Toronto, Canada

40 Quantum Wells, Superlattices, and Band-Gap Engineering
Mark Fox, Sheffield, UK
Radu Ispasoiu, San José, USA

41 Glasses for Photonic Integration
Ray DeCorby, Edmonton, Canada
Mehrdad Irannejad, Waterloo, Canada

42 Optical Nonlinearity in Photonic Glasses
Keiji Tanaka, Sapporo, Japan

30. III–V Ternary and Quaternary Compounds

Sadao Adachi

III–V ternary and quaternary alloy systems are potentially of great importance for many high-speed electronic and optoelectronic devices, because they provide a natural means of tuning the magnitude of forbidden gaps so as to optimize and widen the applications of such semiconductor devices. Literature on the fundamental properties of these material systems is growing rapidly. Even though the basic semiconductor alloy concepts are understood at this time, some practical and device parameters in these material systems have been hampered by a lack of definite knowledge of many material parameters and properties.

This chapter attempts to summarize, in graphical and tabular forms, most of the important theoretical and experimental data on the III–V ternary and quaternary alloy parameters and properties. They can be classified into six groups:

1. Structural parameters
2. Mechanical, elastic, and lattice vibronic properties
3. Thermal properties
4. Energy band parameters
5. Optical properties
6. Carrier transport properties.

The III–V ternary and quaternary alloys considered here are those of Group III (Al, Ga, In) and V (N, P, As, Sb) atoms. The model used in some cases is based on an interpolation scheme and, therefore, requires that data on the material parameters for the related binaries (AlN, AlP, GaN, GaP, etc.) are known. These data have been taken mainly from the Landolt-Börnstein collection, Vol. III/41, and from the *Handbook on Physical Properties of Semiconductors Volume 2: III–V Compound Semiconductors*, published by Springer in 2004. The material parameters and properties derived here are used with wide success to obtain the general properties of these alloy semiconductors.

30.1 **Introduction to III–V Ternary and Quaternary Compounds** 726
30.2 **Interpolation Scheme** 726
30.3 **Structural Parameters** 727
30.3.1 Lattice Parameters and Lattice-Matching Conditions Between III–V Quaternaries and Binary Substrates 727
30.3.2 Molecular and Crystal Densities 727
30.4 **Mechanical, Elastic and Lattice Vibronic Properties** 729
30.4.1 Microhardness 729
30.4.2 Elastic Constants and Related Moduli 729
30.4.3 Long-Wavelength Phonons 729
30.5 **Thermal Properties** 731
30.5.1 Specific Heat and Debye Temperature 731
30.5.2 Thermal Expansion Coefficient 732
30.5.3 Thermal Conductivity 732
30.6 **Energy Band Parameters** 733
30.6.1 Bandgap Energy 733
30.6.2 Carrier Effective Mass 734
30.6.3 Deformation Potential 736
30.7 **Optical Properties** 737
30.7.1 The Reststrahlen Region 737
30.7.2 The Interband Transition Region 739
30.8 **Carrier Transport Properties** 739
References 741

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_30

30.1 Introduction to III–V Ternary and Quaternary Compounds

III–V semiconducting compound alloys are widely used as materials for optoelectronic devices such as light-emitting diodes, laser diodes and photodetectors, as well as for electronic transport devices such as field effect transistors, high electron mobility transistors and heterojunction bipolar transistors. In a ternary alloy, the bandgap energy E_g and the lattice parameter a are generally both functions of a single composition parameter, so they cannot be selected independently. In quaternary alloys, on the other hand, the two composition parameters allow E_g and a to be selected independently, within the constraints of a given alloy-substrate system. Even though the basic semiconductor alloy concepts are understood at this time, the determination of some practical device parameters has been hampered by a lack of definite knowledge of many material parameters. This chapter provides data on the fundamental material properties of III–V ternary and quaternary alloys. The model used here is based on an interpolation scheme and thus requires that values of the material parameters for the related endpoint binaries are known. We therefore begin with the constituent binaries and gradually move on to alloys. The phenomenon of spontaneous ordering in semiconductor alloys, which can be categorized as a self-organized process, is observed to occur spontaneously during the epitaxial growth of certain alloys, and results in modifications to their structural, electronic and optical properties. This topic is omitted from the coverage [30.1].

30.2 Interpolation Scheme

The electronic energy band parameters of III–V compound alloys and their dependence on alloy composition are very important device parameters, and so they have received considerable attention in the past. Investigations of many device parameters have, however, been hampered by a lack of definite knowledge of various material parameters. This necessitates the use of some kind of interpolation scheme. Although the interpolation scheme is still open to experimental verification, it can provide more useful and reliable material parameters over the entire range of alloy composition [30.2].

If one uses the linear interpolation scheme, the ternary parameter T can be derived from the binary parameters (B) by

$$T_{A_xB_{1-x}C} = xB_{AC} + (1-x)B_{BC} \equiv a + bx \qquad (30.1)$$

for an alloy of the form $A_xB_{1-x}C$, where $a \equiv B_{BC}$ and $b \equiv B_{AC} - B_{BC}$. Some material parameters, however, deviate significantly from the linear relation (30.1), and exhibit an approximately quadratic dependence on the mole fraction x. The ternary material parameter in such a case can be very efficiently approximated by the relationship

$$\begin{aligned} T_{A_xB_{1-x}C} &= xB_{AC} + (1-x)B_{BC} + C_{A-B}x(1-x) \\ &\equiv a + bx + cx^2 \,, \end{aligned} \qquad (30.2)$$

where $a \equiv B_{BC}$ and $b \equiv B_{AC} - B_{BC} + C_{A-B}$, and $c \equiv -C_{A-B}$. The parameter c is called the bowing or nonlinear parameter.

The quaternary material $A_xB_{1-x}C_yD_{1-y}$ is thought to be constructed from four binaries: AC, AD, BC, and BD. If one uses the linear interpolation scheme, the quaternary parameter Q can be derived from the Bs by

$$\begin{aligned} Q(x,y) = {} & xyB_{AC} + x(1-y)B_{AD} + (1-x)yB_{BC} \\ & + (1-x)(1-y)B_{BD} \,. \end{aligned} \qquad (30.3)$$

If one of the four binary parameters (e.g., B_{AD}) is lacking, Q can be estimated from

$$Q(x,y) = xB_{AC} + (y-x)B_{BC} + (1-y)B_{BD} \,. \qquad (30.4)$$

The quaternary material $A_xB_yC_{1-x-y}D$ is thought to be constructed from three binaries: AD, BD, and CD. The corresponding linear interpolation is given by

$$Q(x,y) = xB_{AD} + yB_{BD} + (1-x-y)B_{CD} \,. \qquad (30.5)$$

If the material parameter can be given by a specific expression owing to some physical basis, it is natural to consider that the interpolation scheme may also obey this expression. The static dielectric constant ε_s is just the case that follows the Clausius–Mosotti relation. Then, the interpolation expression for the $A_xB_{1-x}C_yD_{1-y}$ quaternary, for example, has the

form

$$\frac{\varepsilon_s(x,y)-1}{\varepsilon_s(x,y)-2} = xy\frac{\varepsilon_s(\mathrm{AC})-1}{\varepsilon_s(\mathrm{AC})-2} + x(1-y)\frac{\varepsilon_s(\mathrm{AD})-1}{\varepsilon_s(\mathrm{AD})-2} + (1-x)y\frac{\varepsilon_s(\mathrm{BC})-1}{\varepsilon_s(\mathrm{BC})-2} + (1-x)(1-y)\frac{\varepsilon_s(\mathrm{BD})-1}{\varepsilon_s(\mathrm{BD})-2} . \quad (30.6)$$

When bowing from the anion sublattice disorder is independent of the disorder in the cation sublattice, the interpolation scheme is written by incorporating these cation and anion bowing parameters into the linear interpolation scheme as

$$Q(x,y) = xyB_{\mathrm{AC}} + x(1-y)B_{\mathrm{AD}} + (1-x)yB_{\mathrm{BC}} + (1-x)(1-y)B_{\mathrm{BD}} + C_{\mathrm{A-B}}x(1-x) + C_{\mathrm{C-D}}y(1-y) \quad (30.7)$$

for the $A_xB_{1-x}C_yD_{1-y}$ quaternary, or

$$Q(x,y) = xB_{\mathrm{AD}} + yB_{\mathrm{BD}} + (1-x-y)B_{\mathrm{CD}} + C_{\mathrm{A-B-C}}xy(1-x-y) \quad (30.8)$$

for the $A_xB_yC_{1-x-y}D$ quaternary.

If relationships for the ternary parameters Ts are available, the quaternary parameter Q can be expressed either as$(A_xB_{1-x}C_yD_{1-y})$

$$Q(x,y) = \frac{x(1-x)[yT_{\mathrm{ABC}}(x)+(1-y)T_{\mathrm{ABD}}(x)]}{x(1-x)+y(1-y)} + \frac{y(1-y)[xT_{\mathrm{ACD}}(y)+(1-x)T_{\mathrm{BCD}}(y)]}{x(1-x)+y(1-y)} , \quad (30.9)$$

or $(A_xB_yC_{1-x-y}D)$

$$Q(x,y) = \frac{xyT_{\mathrm{ABD}}(u)+y(1-x-y)T_{\mathrm{BCD}}(v)}{xy+y(1-x-y)+x(1-x-y)} + \frac{x(1-x-y)T_{\mathrm{ACD}}(w)}{xy+y(1-x-y)+x(1-x-y)} \quad (30.10)$$

with

$$u = \frac{1}{2}(1-x-y) ,$$
$$v = \frac{1}{2}(2-x-2y) ,$$
$$w = \frac{1}{2}(2-2x-y) . \quad (30.11)$$

30.3 Structural Parameters

30.3.1 Lattice Parameters and Lattice-Matching Conditions Between III–V Quaternaries and Binary Substrates

The lattice parameter a (c) is known to obey Vegard's law well, i. e., to vary linearly with composition. Thus, the lattice parameter for a III–V ternary can be simply obtained from (30.1) using the binary data listed in Table 30.1 [30.3, 4]. Introducing the lattice parameters in Table 30.1 into (30.3) [(30.5)], one can also obtain the lattice-matching conditions for $A_{1-x}B_xC_yD_{1-y}$ ($A_xB_yC_{1-x-y}D$) quaternaries on various III–V binary substrates (GaAs, GaSb, InP and InAs). These results are summarized in Tables 30.2, 30.3, 30.4 and 30.5.

30.3.2 Molecular and Crystal Densities

The molecular density d_{M} can be obtained via

$$d_{\mathrm{M}} = \frac{4}{a^3} \quad (30.12)$$

for zinc blende-type materials, and

$$d_{\mathrm{M}} = \frac{4}{a_{\mathrm{eff}}^3} \quad (30.13)$$

for wurtzite-type materials, where a_{eff} is an effective cubic lattice parameter defined by

$$a_{\mathrm{eff}} = \left(\sqrt{3}a^2c\right)^{1/3} . \quad (30.14)$$

Table 30.1 Lattice parameters a and c and crystal density g for some III–V binaries at 300 K

Binary	Zinc blende a (Å)	Wurtzite a (Å)	Wurtzite c (Å)	g (g/cm^{-3})
AlN	–	3.112	4.982	3.258
	4.38	–	–	3.24
AlP	5.4635	–	–	2.3604
AlAs	5.66139	–	–	3.7302
AlSb	6.1355	–	–	4.2775
α-GaN	–	3.1896	5.1855	6.0865
β-GaN	4.52	–	–	6.02
GaP	5.4508	–	–	4.1299
GaAs	5.65330	–	–	5.3175
GaSb	6.09593	–	–	5.6146
InN	–	3.548	5.760	6.813
	4.986	–	–	6.903
InP	5.8690	–	–	4.7902
InAs	6.0583	–	–	5.6678
InSb	6.47937	–	–	5.7768

Table 30.2 Lattice-matching conditions for some III–V quaternaries of type $A_xB_{1-x}C_yD_{1-y}$ at 300 K. $x = \frac{A_0+B_0y}{C_0+D_0y}$

Quaternary	Substrate	A_0	B_0	C_0	D_0	Remark
$Ga_xIn_{1-x}P_yAs_{1-y}$	GaAs	0.4050	−0.1893	0.4050	0.0132	$0 \le y \le 1.0$
	InP	0.1893	−0.1893	0.4050	0.0132	$0 \le y \le 1.0$
$Al_xIn_{1-x}P_yAs_{1-y}$	GaAs	0.4050	−0.1893	0.3969	0.0086	$0.04 \le y \le 1.0$
	InP	0.1893	−0.1893	0.3969	0.0086	$0 \le y \le 1.0$

Table 30.3 Lattice-matching conditions for some III–V quaternaries of type $A_xB_{1-x}C_yD_{1-y}$ at 300 K. $y = \frac{A_0+B_0x}{C_0+D_0x}$

Quaternary	Substrate	A_0	B_0	C_0	D_0	Remark
$Al_xGa_{1-x}P_yAs_{1-y}$	GaAs	0	0.0081	0.2025	−0.0046	$0 \le x \le 1.0$
$Al_xGa_{1-x}As_ySb_{1-y}$	GaSb	0	0.0396	0.4426	0.0315	$0 \le x \le 1.0$
	InP	0.2269	0.0396	0.4426	0.0315	$0 \le x \le 1.0$
	InAs	0.0376	0.0396	0.4426	0.0315	$0 \le x \le 1.0$
$Al_xGa_{1-x}P_ySb_{1-y}$	GaAs	0.4426	0.0396	0.6451	0.0269	$0 \le x \le 1.0$
	GaSb	0	0.0396	0.6451	0.0269	$0 \le x \le 1.0$
	InP	0.2269	0.0396	0.6451	0.0269	$0 \le x \le 1.0$
	InAs	0.0376	0.0396	0.6451	0.0269	$0 \le x \le 1.0$
$Ga_xIn_{1-x}As_ySb_{1-y}$	GaSb	0.3834	−0.3834	0.4211	0.0216	$0 \le x \le 1.0$
	InP	0.6104	−0.3834	0.4211	0.0216	$0.47 \le x \le 1.0$
	InAs	0.4211	−0.3834	0.4211	0.0216	$0 \le x \le 1.0$
$Ga_xIn_{1-x}P_ySb_{1-y}$	GaAs	0.8261	−0.3834	0.6104	0.0348	$0.52 \le x \le 1.0$
	GaSb	0.3834	−0.3834	0.6104	0.0348	$0 \le x \le 1.0$
	InP	0.6104	−0.3834	0.6104	0.0348	$0 \le x \le 1.0$
	InAs	0.4211	−0.3834	0.6104	0.0348	$0 \le x \le 1.0$
$Al_xIn_{1-x}As_ySb_{1-y}$	GaSb	0.3834	−0.3439	0.4211	0.0530	$0 \le x \le 1.0$
	InP	0.6104	−0.3439	0.4211	0.0530	$0.48 \le x \le 1.0$
	InAs	0.4211	−0.3439	0.4211	0.0530	$0 \le x \le 1.0$
$Al_xIn_{1-x}P_ySb_{1-y}$	GaAs	0.8261	−0.3439	0.6104	0.0616	$0.53 \le x \le 1.0$
	GaSb	0.3834	−0.3439	0.6104	0.0616	$0 \le x \le 1.0$
	InP	0.6104	−0.3439	0.6104	0.0616	$0 \le x \le 1.0$
	InAs	0.4211	−0.3439	0.6104	0.0616	$0 \le x \le 1.0$

Table 30.4 Lattice-matching conditions for some III–V quaternaries of type $A_xB_yC_{1-x-y}D$ at 300 K. $y = A_0 + B_0x$

Quaternary	Substrate	A_0	B_0	Remark
$Al_xGa_yIn_{1-x-y}P$	GaAs	0.5158	−0.9696	$0 \le x \le 0.53$
$Al_xGa_yIn_{1-x-y}As$	InP	0.4674	−0.9800	$0 \le x \le 0.48$

Table 30.5 Lattice-matching conditions for some III–V quaternaries of type $AB_xC_yD_{1-x-y}$ at 300 K. $x = A_0 + B_0y$

Quaternary	Substrate	A_0	B_0	Remark
$AlP_xAs_ySb_{1-x-y}$	GaAs	0.7176	−0.7055	$0 \le y \le 1.0$
	InP	0.3966	−0.7055	$0 \le y \le 0.56$
	InAs	0.1149	−0.7055	$0 \le y \le 0.16$
$GaP_xAs_ySb_{1-x-y}$	GaAs	0.6861	−0.6861	$0 \le y \le 1.0$
	InP	0.3518	−0.6861	$0 \le y \le 0.51$
	InAs	0.0583	−0.6861	$0 \le y \le 0.085$
$InP_xAs_ySb_{1-x-y}$	GaSb	0.6282	−0.6899	$0 \le y \le 0.911$
	InAs	0.6899	−0.6899	$0 \le y \le 1.0$

The x-ray crystal density g can be simply written, using d_M, as

$$g = \frac{Md_M}{N_A} , \tag{30.15}$$

where M is the molecular weight and $N_A = 6.022 \times 10^{23}\,\mathrm{mol}^{-1}$ is the Avogadro constant. We list g for some III–V binaries in Table 30.1. Alloy values of d_M and g can be accurately obtained using Vegard's law, i. e., (30.1), (30.3), and (30.5).

30.4 Mechanical, Elastic and Lattice Vibronic Properties

30.4.1 Microhardness

The hardness test has been used for a long time as a simple means of characterizing the mechanical behavior of solids. The Knoop hardness H_P for $Ga_xIn_{1-x}P_yAs_{1-y}$ lattice-matched to InP has been reported [30.5], and is found to increase gradually from 520 kg/mm^2 for $y = 0$ ($Ga_{0.47}In_{0.53}As$) to 380 kg/mm^2 for $y = 1.0$ (InP). It has also been reported that the microhardness in $Al_xGa_{1-x}N$ thin film slightly decreases with increasing AlN composition x [30.6].

30.4.2 Elastic Constants and Related Moduli

Although the elastic properties of the III–V binaries have been studied extensively, little is known about their alloys. Recent studies, however, suggested that the elastic properties of the alloys can be obtained, to a good approximation, by averaging the binary endpoint values [30.7, 8]. We have, therefore, listed in Tables 30.6 and 30.7 the elastic stiffness (C_{ij}) and compliance constants (S_{ij}) for some III–V binaries with zinc blende and wurtzite structures, respectively. Table 30.8 also summarizes the functional expressions for the bulk modulus B_u, Young's modulus Y, and Poisson's ratio P. Note that Y and P are not isotropic, even in the cubic zinc blende lattice.

30.4.3 Long-Wavelength Phonons

The atoms of a crystal can be visualized as being joined by harmonic springs, and the crystal dynamics can be analyzed in terms of a linear combination of $3N$ normal modes of vibration (N is the number of different types of atoms; different in terms of mass or ordering in space). In alloys, the nature of the lattice optical spectrum depends on the difference between the quantities representing the lattice vibronic properties of the components. If these quantities are similar, then the optical response of an alloy is similar to the response of a crystal with the quantities averaged over the composition (one-mode behavior). In one-mode systems, such as most I–VII alloys, a single set of long-wavelength optical modes appears, as schematically shown in Fig. 30.1. When the parameters differ strongly, the response of a system is more complex; the spectrum contains a number of bands, each of which corresponds to one of the components, and it has an intensity governed by its content in the alloy (*multimode* behavior). For example, a two-mode system exhibits

Table 30.6 Elastic stiffness (C_{ij}) and compliance constants (S_{ij}) for some cubic III–V binaries at 300 K

Binary	C_{ij} (10^{11} dyn/cm^2)			S_{ij} (10^{-12} cm^2/dyn)		
	C_{11}	C_{12}	C_{44}	S_{11}	S_{12}	S_{44}
AlN	31.5[a]	15.0[a]	18.5[a]	0.458[a]	−0.148[a]	0.541[a]
AlP	15.0[a]	6.42[a]	6.11[a]	0.897[a]	−0.269[a]	1.64[a]
AlAs	11.93	5.72	5.72	1.216	−0.394	1.748
AlSb	8.769	4.341	4.076	1.697	−0.5618	2.453
β-GaN	29.1[a]	14.8[a]	15.8[a]	0.523[a]	−0.176[a]	0.633[a]
GaP	14.050	6.203	7.033	0.9756	−0.2988	1.422
GaAs	11.88	5.38	5.94	1.173	−0.366	1.684
GaSb	8.838	4.027	4.320	1.583	−0.4955	2.315
InN	19.2[a]	7.30[a]	9.35[a]	0.659[a]	−0.182[a]	1.07[a]
InP	10.22	5.73	4.42	1.639	−0.589	2.26
InAs	8.329	4.526	3.959	1.945	−0.6847	2.526
InSb	6.608	3.531	3.027	2.410	−0.8395	3.304

[a] Theoretical

Table 30.7 Elastic stiffness (C_{ij}) and compliance constants (S_{ij}) for some wurtzite III–V binaries at 300 K

Binary	C_{ij} (10^{11} dyn/cm^2)						S_{ij} (10^{-12} cm^2/dyn)					
	C_{11}	C_{12}	C_{13}	C_{33}	C_{44}	C_{66}^a	S_{11}	S_{12}	S_{13}	S_{33}	S_{44}	S_{66}^b
AlN	41.0	14.0	10.0	39.0	12.0	13.5	0.285	−0.085	−0.051	0.283	0.833	0.740
α-GaN	37.3	14.1	8.0	38.7	9.4	11.6	0.320	−0.112	−0.043	0.276	1.06	0.864
InN	19.0	10.4	12.1	18.2	0.99	4.3	0.957	−0.206	−0.499	1.21	10.1	2.33

[a] $C_{66} = 1/2(C_{11} - C_{12})$, [b] $S_{66} = 2(S_{11} - S_{12})$

Table 30.8 Functional expressions for the bulk modulus B_u, Young's modulus Y, and Poisson's ratio P in semiconductors with zinc blende (ZB) and wurtzite (W) structures

Parameter	Structure	Expression	Remark
B_u	ZB	$(C_{11}+2C_{12})/3$	
	W	$[(C_{11}+C_{12})C_{33}-2C_{13}^2]/(C_{11}+C_{12}+2C_{33}-4C_{13})$	
Y	ZB	$1/S_{11}$	(100), [001]
		$1/(S_{11}-S/2)$	(100), [011]
		$1/S_{11}$	(110), [001]
		$1/(S_{11}-2S/3)$	(110), [111]
		$1/(S_{11}-S/2)$	(111)
	W	$1/S_{11}$	$c \perp \boldsymbol{l}$
		$1/S_{33}$	$c \parallel \boldsymbol{l}$
P	ZB	$-S_{12}/S_{11}$	(100), $m=[010]$, $n=[001]$
		$-(S_{12}+S/2)/(S_{11}-S/2)$	(100), $m=[011]$, $n=[0\bar{1}1]$
		$-S_{12}/S_{11}$	(110), $m=[001]$, $n=[1\bar{1}0]$
		$-(S_{12}+S/3)/(S_{11}-2S/3)$	(110), $m=[1\bar{1}1]$, $n=[1\bar{1}\bar{2}]$
		$-(S_{12}+S/6)/(S_{11}-S/2)$	(111)
	W	$(1/2)[1-(Y/3B_u)]$	$c \perp \boldsymbol{l}$, $c \parallel \boldsymbol{l}$

$S = S_{11} - S_{12} - (S_{44}/2)$; $\boldsymbol{l}$ = directional vector; m = direction for a longitudinal stress; n = direction for a transverse strain ($n \perp m$)

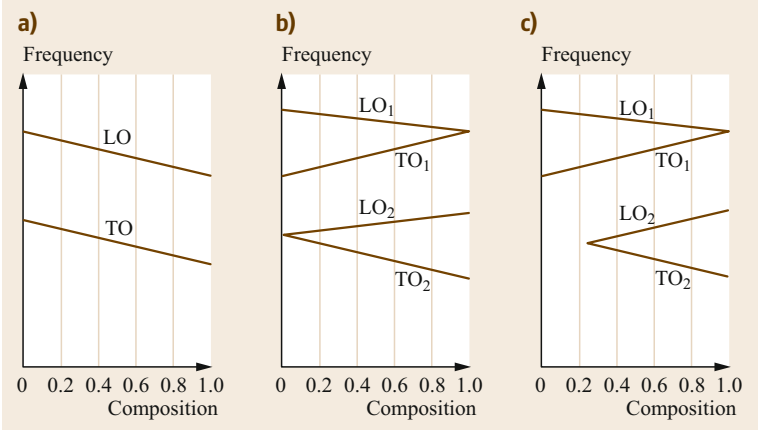


Fig. 30.1a–c Three different types of long-wavelength phonon mode behavior in ternary alloys: **(a)** one-mode; **(b)** two-mode; and **(c)** one–two-mode

Table 30.9 Behavior of the long-wavelength optical modes in III–V ternary and quaternary alloys

Behavior	Alloy
One mode	AlGaN (LO; cubic), AlGaN (except for E_1(TO); wurtzite), GaInN, AlAsSb
Two mode	AlGaN (TO; cubic), AlGaN (E_1(TO); wurtzite), AlInN, AlGaP, AlGaAs, AlGaSb, AlInP, AlInAs, AlInSb, GaInP, GaInAs, GaNP, GaNAs, GaPAs, GaPSb, GaAsSb, InPAs, InPSb
One–two mode	GaInSb, InAsSb
Two–three mode	AlGaInP/GaAs
Three mode	AlGaAsSb, GaInAsSb, AlGaInAs/InP, InPAsSb
Three–four mode	AlGaPAs
Four mode	GaInPSb, GaInPAs/GaAs, GaInPAs/InP

two distinct sets of optical modes with frequencies characteristic of each end member and strengths that are roughly proportional to the respective concentrations.

As seen in Table 30.9, the long-wavelength optical phonons in III–V ternaries exhibit either one-mode or two-mode behavior, or more rigorously, three different types of mode behavior: one-mode, two-mode, and one–two-mode behaviors. The one–two-mode system exhibits a single mode over only a part of the composition range, with two modes observed over the remaining range of compositions.

In a quaternary alloy of the $A_xB_{1-x}C_yD_{1-y}$ type, there are four kinds of unit cells: AC, AD, BC, and BD. On the other hand, in the $A_xB_yC_{1-x-y}D$ type there are three kinds of unit cells: AD, BD, and CD. We can, thus, expect four-mode or three-mode behavior of the long-wavelength optical modes in such quaternary alloys ([30.9]; Table 30.9). However, the $Ga_xIn_{1-x}As_ySb_{1-y}$ quaternary showed three-mode behavior with GaAs, InSb and mixed InAs/GaAs characteristics [30.10]. The $Ga_xIn_{1-x}As_ySb_{1-y}$ quaternary was also reported to show two-mode or three-mode behavior, depending on the alloy composition [30.11].

The long-wavelength optical phonon behavior in the $Al_xGa_{1-x}As$ ternary has been studied both theoretically and experimentally. These studies suggest that the optical phonons in $Al_xGa_{1-x}As$ exhibit the two-mode behavior over the whole composition range. Thus, the $Al_xGa_{1-x}As$ system has two couples of the transverse optical (TO) and longitudinal optical (LO) modes; one is the GaAs-like mode and the other is the AlAs-like mode. Each phonon frequency can be expressed as [30.12]:

- TO (GaAs): $268-14x\ \mathrm{cm}^{-1}$
- LO (GaAs): $292-38x\ \mathrm{cm}^{-1}$
- TO (AlAs): $358+4x\ \mathrm{cm}^{-1}$
- LO (AlAs): $358+71x-26x^2\ \mathrm{cm}^{-1}$.

It is observed that only the AlAs-like LO mode shows a weak nonlinearity with respect to the alloy composition x.

30.5 Thermal Properties

30.5.1 Specific Heat and Debye Temperature

Since alloying has no significant effect on elastic properties, it appears that using the linear interpolation scheme for alloys can provide generally acceptable specific heat values (C). In fact, it has been reported that the C values for InP_xAs_{1-x} [30.13] and $Al_xGa_{1-x}As$ [30.14] vary fairly linearly with alloy composition x. It has also been shown [30.12] that the Debye temperature θ_D for alloys shows very weak nonlinearity with composition. From these facts, one can suppose that the linear interpolation scheme may provide generally acceptable C and θ_D values for III–V semiconductor alloys. We have, therefore, listed in Table 30.10 the III–V binary endpoint values for C and θ_D at $T = 300\,\mathrm{K}$. Using these values, the linearly interpolated C value for $Al_xGa_{1-x}As$ can be obtained from $C(x) = 0.424x + 0.327(1-x) = 0.327 + 0.097x$ (J/(g K)).

Table 30.10 Specific heat C and Debye temperature θ_D for some III–V binaries at 300 K

Binary	C (J/(g K))	θ_D (K)	α_{th} ($10^{-6}K^{-1}$)
AlN	0.728	988	3.2 ($\perp c$), 2.4 ($\parallel c$)
AlP	0.727	687	
AlAs	0.424	450	4.1
AlSb	0.326[a]	370[a]	4.2
α-GaN	0.42	821	6.2 ($\perp c$), 4.9 ($\parallel c$)
GaP	0.313	493[b]	4.89
GaAs	0.327	370	5.73
GaSb	0.344[a]	240[a]	6.35
InN	2.274	674	3.830 ($\perp c$), 2.751 ($\parallel c$)
InP	0.322	420[a]	4.56
InAs	0.352	280[a]	5.50
InSb	0.350[a]	161[a]	5.04

[a] At 273 K, [b] at 150 K

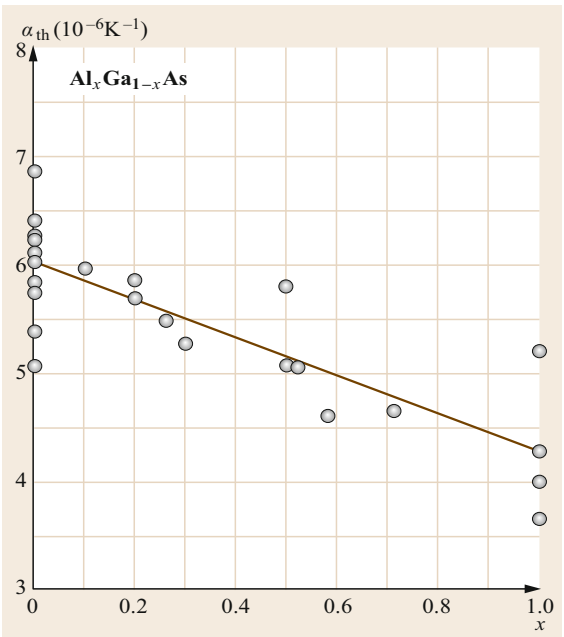


Fig. 30.2 Thermal expansion coefficient α_{th} as a function of x for the $Al_xGa_{1-x}As$ ternary at $T = 300\,\mathrm{K}$. The experimental data are gathered from various sources. The *solid line* is linearly interpolated between the AlAs and GaAs values

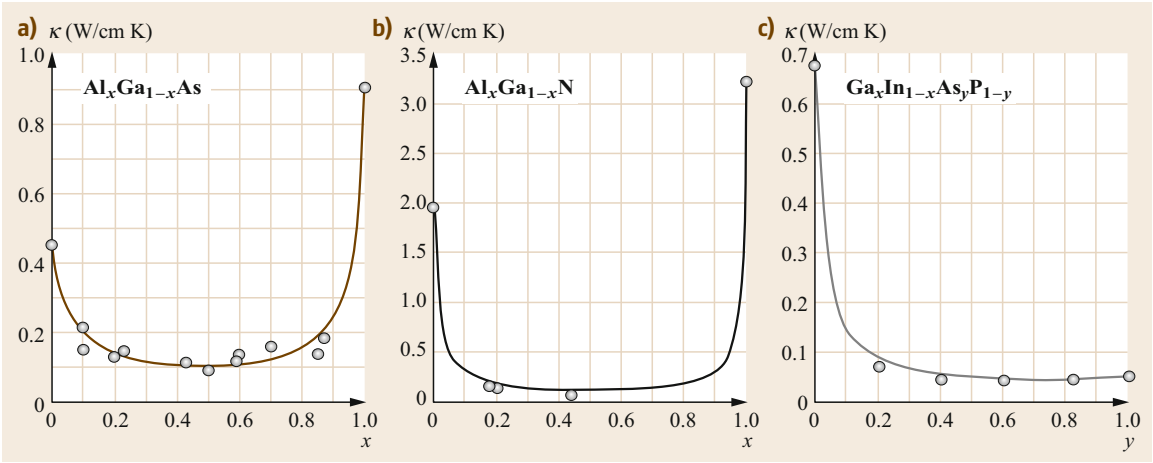


Fig. 30.3a–c Thermal conductivity κ as a function of $x(y)$ for **(a)** $Al_xGa_{1-x}As$, **(b)** $Al_xGa_{1-x}N$, and **(c)** $Ga_xIn_{1-x}As_yP_{1-y}$ lattice-matched to InP at $T = 300\,K$. The experimental data (*solid circles*) are gathered from various sources. The *solid lines* represent the results calculated from (30.2) and (30.6) using the binary endpoint values and nonlinear parameters in Table 30.11

30.5.2 Thermal Expansion Coefficient

The linear thermal expansion coefficient α_{th} is usually measured by measuring the temperature dependence of the lattice parameter. The composition dependence of α_{th} has been measured for many semiconductor alloys, including $Ga_xIn_{1-x}P$ [30.15] and GaP_xAs_{1-x} [30.16]. These studies indicate that the α_{th} value varies almost linearly with composition. This suggests that the thermal expansion coefficient can be accurately estimated using linear interpolation. In fact, we plot in Fig. 30.2 the 300 K value of α_{th} as a function of x for the $Al_xGa_{1-x}As$ ternary. By using the least-squares fit procedure, we obtain the linear relationship between α_{th} and x as $\alpha_{th}(x) = 6.01-1.74x\ (10^{-6}\,K^{-1})$. This expression is almost the same as that obtained using the linear interpolation expression

$$\alpha_{th}(x) = 4.28x + 6.03(1-x)$$
$$= 6.03-1.75(10^{-6}\,K^{-1})\,.$$

The binary endpoint values of α_{th} are listed in Table 30.10.

30.5.3 Thermal Conductivity

The lattice thermal conductivity κ, or the thermal resistivity $W = 1/\kappa$, results mainly from interactions between phonons and from the scattering of phonons by crystalline imperfections. It is important to point out that when large numbers of foreign atoms are added to the host lattice, as in alloying, the thermal conductivity may decrease significantly. Experimental data on various alloy semiconductors, in fact, exhibit strong nonlinearity with respect to the alloy composition. Such a composition dependence can be successfully explained by using the quadratic expression of (30.2) or (30.6) [30.17].

In Fig. 30.3 we compare the results calculated from (30.2) [(30.7)] to the experimental data for $Al_xGa_{1-x}As$, $Al_xGa_{1-x}N$ and $Ga_xIn_{1-x}As_yP_{1-y}$/InP alloys. The binary W values used in these calculations are taken from Table 30.11. The corresponding nonlinear parameters C_{A-B} are also listed in Table 30.11. The agreement between the calculated and experimen-

Table 30.11 Thermal resistivity values W for some III–V binaries at 300 K. Several cation and anion bowing parameters used for the calculation of alloy values are also listed in the last column

Binary	W (cm K/W)	C_{A-B} (cm K/W)
AlN	0.31[a]	
AlP	1.11	$C_{Al-Ga} = 32$
AlAs	1.10	$C_{Al-In} = 15$
AlSb	1.75	$C_{Ga-In} = 72$
α-GaN	0.51[a]	$C_{N-P} = 36$
GaP	1.30	$C_{N-As} = 12$
GaAs	2.22	$C_{N-Sb} = 10$
GaSb	2.78	$C_{P-As} = 25$
InN	2.22[b]	$C_{P-Sb} = 16$
InP	1.47	$C_{As-Sb} = 91$
InAs	3.33	
InSb	5.41–6.06	

[a] Heat flow parallel to the basal plane, [b] ceramics

tal data is excellent. By applying the present model, it is possible to estimate the κ (or W) values of experimentally unknown III–V alloy systems, such as $GaAs_xSb_{1-x}$ and $Al_xGa_yIn_{1-x-y}As$.

30.6 Energy Band Parameters

30.6.1 Bandgap Energy

Lowest Direct and Lowest Indirect Band Gaps

The bandgap energies of III–V ternaries usually deviate from the simple linear relation of (30.1) and have an approximately quadratic dependence on the alloy composition x. Table 30.12 summarizes the lowest direct gap energy E_0 and the lowest indirect gap energies E_g^X and E_g^L for some III–V binaries of interest here. The corresponding nonlinear parameters C_{A-B} are listed in Table 30.13 [30.18]. Note that the E_g^X and E_g^L transitions correspond to those from the highest valence band at the Γ point to the lowest conduction band near X ($\Gamma_8 \rightarrow X_6$) or near L ($\Gamma_8 \rightarrow L_6$), respectively. The E_0 transitions take place at the Γ point ($\Gamma_8 \rightarrow \Gamma_6$).

Figure 30.4 plots the values of E_0 and E_g^X as a function of alloy composition x for the $Ga_xIn_{1-x}P$ ternary at $T = 300$ K. The solid lines are obtained by introducing the numerical values from Tables 30.12 and 30.13 into (30.2). These curves provide the direct–indirect crossover composition at $x \approx 0.7$. Figure 30.5 also shows the variation in composition of E_0 in the $Ga_xIn_{1-x}As$, $InAs_xSb_{1-x}$ and $Ga_xIn_{1-x}Sb$ ternaries. It is understood from Table 30.13 that the bowing parameters for the bandgap energies of III–V ternaries are negative or very small, implying a downward bowing or a linear interpolation to within experimental uncertainty (Figs. 30.4 and 30.5). It should be noted that nitrogen incorporation into (In,Ga)(P,As) results in a giant bandgap bowing of the host lattice for increasing nitrogen concentration [30.19]. We also summarize in Table 30.14 the expressions for the E_0 gap energy of some III–V quaternaries as a function of alloy composition.

Table 30.12 Band-gap energies, E_0, E_g^X and E_g^L, for some III–V binaries at 300 K. ZB = zinc blende

Binary	E_0 (eV)	E_g^X (eV)	E_g^L (eV)
AlN	6.2	–	–
AlN (ZB)	5.2	5.34	8.6[a]
AlP	3.91	2.48	3.57[a]
AlAs	3.01	2.15	2.37
AlSb	2.27	1.615	2.211
α-GaN	3.420	–	–
β-GaN	3.231	4.2[a]	5.5[a]
GaP	2.76	2.261	2.63
GaAs	1.43	1.91	1.72
GaSb	0.72	1.05	0.76
InN	0.7–1.1	–	–
InN (ZB)	0.56	3.0[a]	5.8[a]
InP	1.35	2.21	2.05
InAs	0.359	1.37	1.07
InSb	0.17	1.63	0.93

[a] Theoretical

Higher-Lying Band Gaps

The important optical transition energies observed at energies higher than E_0 are labeled E_1 and E_2. We

Table 30.13 Bowing parameters used in the calculation of E_0, E_g^X and E_g^L for some III–V ternaries. In those cases where no value is listed, linear variation should be assumed. W = wurtzite; ZB = zinc blende

Ternary	Bowing parameter C_{A-B} (eV)		
	E_0	E_g^X	E_g^L
(Al,Ga)N (W)	−1.00	–	–
(Al,Ga)N (ZB)	0	−0.61	−0.80
(Al,In)N (W)	−3.70	–	–
(Al,In)N (ZB)			
(Ga,In)N (W)	−1.640	–	–
(Ga,In)N (ZB)	0		
(Al,Ga)P	0	0	
(Al,In)P	−0.40	0	
(Ga,In)P	−0.65	−0.200	−0.34
(Al,Ga)As	−0.37	−0.245	−0.055
(Al,In)As	−0.720		
(Ga,In)As	−0.580	−0.70	−0.50
(Al,Ga)Sb	−0.47	0	−0.55
(Al,In)Sb	−0.43		
(Ga,In)Sb	−0.415	−0.33	−0.40
Al(P,As)	−0.13	−0.40	−0.38
Al(P,Sb)	−2.13	−0.277	−0.756
Al(As,Sb)	−1.19	−0.250	−0.474
Ga(N,P) (ZB)	−3.9		
Ga(N,As) (ZB)	$-120.4 + 100x$		
Ga(P,As)	−0.19	−0.240	−0.16
Ga(P,Sb)	−2.70	−2.700	−2.70
Ga(As,Sb)	−1.25	−1.20	−1.20
In(N,P) (ZB)	−15		
In(N,As) (ZB)	−4.22		
In(P,As)	−0.145	−0.145	−0.145
In(P,Sb)	−1.60	−1.60	−1.60
In(As,Sb)	−0.600	−0.60	−0.60

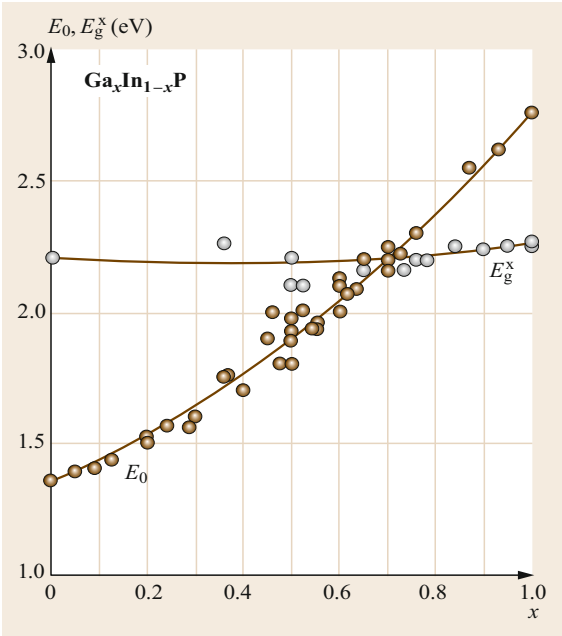


Fig. 30.4 Variation of the lowest direct gap (E_0) and lowest indirect gap energies (E_g^X) in the $Ga_xIn_{1-x}P$ ternary at $T = 300\,K$. The experimental data are gathered from various sources. The *solid lines* are calculated from (30.2) using the binary endpoint values and bowing parameters in Tables 30.12 and 30.13

summarize in Table 30.15 the higher-lying bandgap energies E_1 and E_2 for some III–V binaries. The corresponding bowing parameters for these gaps are listed in Table 30.16.

30.6.2 Carrier Effective Mass

Electron Effective Mass

Since the carrier effective mass is strongly connected with the carrier mobility, it is known to be one of the most important device parameters. Effective masses can be measured by a variety of techniques, such as the Shubnikov-de Haas effect, magnetophonon resonance, cyclotron resonance, and interband magneto-optical effects. We list in Table 30.17 the electron effective mass (m_e^Γ) at the Γ-conduction band and the density of states (m_e^α) and conductivity masses (m_c^α) at the X-conduction and L-conduction bands of some III–V binaries. We also list in Table 30.18 the bowing parameters used when calculating the electron effective mass m_e^Γ for some III–V ternaries from (30.2). Note that the density of states mass m_e^α for electrons in the conduction band minima $\alpha = \Gamma$, X, and L can be obtained from

$$m_e^\alpha = N^{2/3} m_{t\alpha}^{2/3} m_{l\alpha}^{1/3} , \qquad (30.16)$$

where N is the number of equivalent α minima ($N = 1$ for the Γ minimum, $N = 3$ for the X minima, and $N = 4$ for the L minima). The two masses m_l and m_t in (30.16) are called the longitudinal and transverse masses, respectively. The density of states effective mass m_e^α is used to calculate the density of states. The conductivity effective mass m_c^α, which can be used for calculating the conductivity (mobility), is also given by

$$m_c^\alpha = \frac{3 m_{t\alpha} m_{l\alpha}}{m_{t\alpha} + 2 m_{l\alpha}} . \qquad (30.17)$$

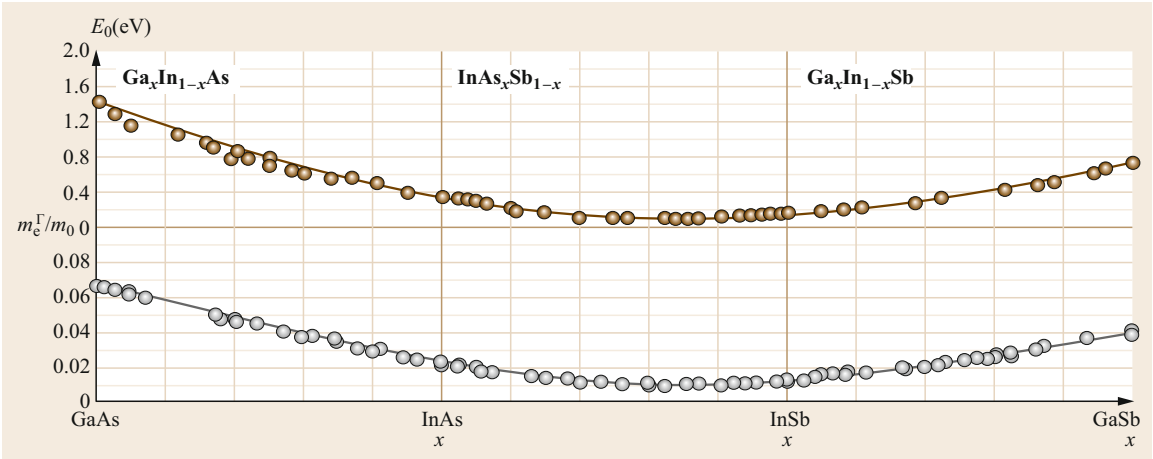


Fig. 30.5 Variation of the lowest direct gap energy E_0 ($T = 300\,K$) and electron effective mass m_e^Γ at the Γ-conduction bands of $Ga_xIn_{1-x}As$, $InAs_xSb_{1-x}$ and $Ga_xIn_{1-x}Sb$ ternaries. The experimental data are gathered from various sources. The *solid lines* are calculated from (30.2) using the binary endpoint values and bowing parameters in Tables 30.12 and 30.13 (E_0) and those in Tables 30.17 and 30.18 (m_e^Γ)

Table 30.14 Bandgap energies E_0 for some III–V quaternaries at 300 K

Quaternary	E_0 (eV)
$Ga_xIn_{1-x}P_yAs_{1-y}$/InP	$0.75 + 0.46y + 0.14y^2$
$Ga_xIn_{1-x}As_ySb_{1-y}$/GaSb	$0.28 - 0.16x + 0.60x^2$
$Ga_xIn_{1-x}As_ySb_{1-y}$/InAs	$0.359 - 0.415x + 0.718x^2$
$Al_xGa_yIn_{1-x-y}P$/GaAs[a]	$1.90 + 0.57x + 0.10x^2$
$Al_xGa_yIn_{1-x-y}As$/InP	$0.75 + 0.68x + 0.06x^2$
$InP_xAs_ySb_{1-x-y}$/InAs	$0.512 + 0.030y - 0.183y^2$

[a] The lowest indirect gap energy for this quaternary alloy can be obtained via $E_g^X = 2.25 + 0.09x - 0.01x^2$ eV

Table 30.15 Higher-lying bandgap energies, E_1 and E_2, for some III–V binaries at 300 K

Binary	E_1 (eV)	E_2 (eV)
AlN	7.76	8.79
AlP	4.30	4.63
AlAs	3.62–3.90	4.853, 4.89
AlSb	2.78–2.890	4.20–4.25
α-GaN	6.9	8.0
β-GaN	7.0	7.6
GaP	3.71	5.28
GaAs	2.89–2.97	4.960–5.45
GaSb	2.05	4.08–4.20
InN	5.0	7.6
InP	3.17	4.70 (E_0')
InAs	2.50	4.70
InSb	1.80	3.90

Since $m_{l\Gamma} = m_{l\Gamma}$ at the $\alpha = \Gamma$ minimum of cubic semiconductors, we have the relation $m_e^\Gamma = m_c^\Gamma$. In the case of wurtzite semiconductors, we have the relation $m_e^\Gamma \neq m_c^\Gamma$, but the difference is very small.

The composition dependence of the electron effective mass m_e^Γ at the Γ-conduction bands of $Ga_xIn_{1-x}As$, $InAs_xSb_{1-x}$ and $Ga_xIn_{1-x}Sb$ ternaries is plotted in Fig. 30.5. The solid lines are calculated from (30.2) using the binary endpoint values and bowing parameters in Tables 30.17 and 30.18. For conventional semiconductors, the values of the effective mass are known to decrease with decreasing bandgap energy (Fig. 30.5). This is in agreement with a trend predicted by the $\boldsymbol{k}\cdot\boldsymbol{p}$ theory [30.2]. In III–V-N alloys, the electron effective mass has been predicted to increase with increasing nitrogen composition in the low composition range [30.19]. This behavior is rather unusual, and in fact is opposite to what is seen in conventional semiconductors. However, a more recent study suggested that the effective electron mass in GaN_xAs_{1-x} decreases from $0.084m_0$ to $0.029m_0$ as x increases from 0 to 0.004 [30.20]. We also summarize in Table 30.19 the composition dependence of m_e^Γ, determined for $Ga_xIn_{1-x}P_yAs_{1-y}$ and $Al_xGa_yIn_{1-x-y}As$ quaternaries lattice-matched to InP.

Table 30.16 Bowing parameters used in the calculation of the higher-lying bandgap energies, E_1 and E_2, for some cubic III–V ternaries. In those cases where no value is listed, linear variation should be assumed

Ternary	C_{A-B} (eV)	
	E_1	E_2 (E_0')
(Ga,In)N	−1.11	−1.26
(Al,Ga)P	0	0
(Al,In)P	0	0
(Ga,In)P	−0.86	0
(Al,Ga)As	−0.39	0
(Al,In)As	−0.44	−0.24
(Ga,In)As	−0.51	−0.27
(Al,Ga)Sb	−0.31	−0.34
(Al,In)Sb	−0.30	−0.14
(Ga,In)Sb	−0.36	−0.15
Ga(N,P)	0	0
Ga(N,As)	0	0
Ga(P,As)	0	0
Ga(As,Sb)	−0.59	−0.19
In(P,As)	−0.17	−0.03
In(As,Sb)	≈ -0.8	≈ -1.4

Hole Effective Mass

The effective mass can only be clearly defined for an isotropic parabolic band. In the case of III–V materials, the valence bands are warped from spherical symmetry some distance away from the Brillouin zone center (Γ). Depending on the measurement or calculation technique employed, different values of hole masses are then possible experimentally or theoretically. Thus, it is always important to choose the correct definition of the effective hole mass which appropriate to the physical phenomenon considered.

We list in Table 30.20 the density of states heavy hole (m_{HH}^*), the averaged light hole (m_{LH}^*), and spin orbit splitoff effective hole masses (m_{SO}) in some cubic III–V semiconductors. These masses are, respectively, defined using Luttinger's valence band parameters γ_i by

$$m_{HH}^* = \frac{\left(1 + 0.05\gamma_h + 0.0164\gamma_h^2\right)^{2/3}}{\gamma_1 - \overline{\gamma}}\,, \tag{30.18}$$

$$m_{LH}^* = \frac{1}{\gamma_1 + \overline{\gamma}}\,, \tag{30.19}$$

$$m_{SO} = \frac{1}{\gamma_1} \tag{30.20}$$

Table 30.17 Electron effective mass at the Γ-conduction band (m_e^{Γ}) and density of states (m_e^{α}) and conductivity masses (m_c^{α}) at the X-conduction and L-conduction bands of some III–V binaries. ZB = zinc blende

Binary	m_e^{Γ}/m_0	Density of states mass		Conductivity mass	
		m_e^{X}/m_0	m_e^{L}/m_0	m_c^{X}/m_0	m_c^{L}/m_0
AlN	0.29–0.45	–	–	–	–
AlN (ZB)	0.26[a]	0.78[a]		0.37[a]	
AlP	0.220[a]	1.14[a]		0.31[a]	
AlAs	0.124	0.71	0.78	0.26[a]	0.21[a]
AlSb	0.14	0.84	1.05[a]	0.29	0.28[a]
α-GaN	0.21	–	–	–	–
β-GaN	0.15	0.78[a]		0.36[a]	
GaP	0.114	1.58	0.75[a]	0.37	0.21[a]
GaAs	0.067	0.85	0.56	0.32	0.11
GaSb	0.039	1.08[a]	0.54	0.44[a]	0.12
InN	0.044	–	–	–	–
InN (ZB)	0.03[a]				
InP	0.07927	1.09[a]	0.76[a]	0.45[a]	0.19[a]
InAs	0.024	0.98[a]	0.94[a]	0.38[a]	0.18[a]
InSb	0.013				

[a] Theoretical

Table 30.18 Bowing parameter used in the calculation of the electron effective mass m_e^{Γ} at the Γ-conduction bands of some III–V ternaries

Ternary	C_{A-B} (m_0)
(Ga,In)P	−0.01854
(Al,Ga)As	0
(Al,In)As	−0.012
(Ga,In)As	−0.008
(Ga,In)Sb	−0.010
Ga(P,As)	0
Ga(As,Sb)	−0.014
In(P,As)	0
In(As,Sb)	−0.027

with

$$\overline{\gamma} = \left(2\gamma_2^2 + 2\gamma_3^2\right)^{1/2}, \quad \gamma_h = \frac{6\left(\gamma_3^2 - \gamma_2^2\right)}{\overline{\gamma}(\gamma_1 - \overline{\gamma})}. \tag{30.21}$$

Only a few experimental studies have been performed on the effective hole masses in III–V alloys, e.g., the $Ga_xIn_{1-x}P_yAs_{1-y}$ quaternary [30.2]. While some data imply a bowing parameter, the large uncertainties in existing determinations make it difficult to conclusively state that such experimental values are preferable to a linear interpolation. The binary endpoint data listed in Table 30.20 enable us to estimate alloy values using the linear interpolation scheme.

30.6.3 Deformation Potential

The deformation potentials of the electronic states at the Brillouin zone centers of semiconductors play an important role in many physical phenomena. For example, the splitting of the heavy hole and light hole bands at the Γ point of the strained substance can be explained by the shear deformation potentials, b and d. The lattice mobilities of holes are also strongly affected by these potentials. Several experimental data have been

Table 30.19 Electron effective mass m_e^{Γ} at the Γ-conduction bands of some III–V quaternaries

Quaternary	m_e^{Γ}/m_0
$Ga_xIn_{1-x}P_yAs_{1-y}$/InP	$0.04084 + 0.03384y + 0.00459y^2$
$Al_xGa_yIn_{1-x-y}As$/InP	$0.043 + 0.031x$

Table 30.20 Density of states heavy hole (m_{HH}^*), averaged light hole (m_{LH}^*), and spin orbit splitoff effective hole masses (m_{SO}) in some cubic III–V semiconductors. ZB = zinc blende

Material	m_{HH}^*/m_0	m_{LH}^*/m_0	m_{SO}/m_0
AlN (ZB)	1.77[a]	0.35[a]	0.58[a]
AlP	0.63[a]	0.20[a]	0.29[a]
AlAs	0.81[a]	0.16[a]	0.30[a]
AlSb	0.9	0.13	0.317[a]
β-GaN	1.27[a]	0.21[a]	0.35[a]
GaP	0.52	0.17	0.34
GaAs	0.55	0.083	0.165
GaSb	0.37	0.043	0.12
InN (ZB)	1.959[a]	0.098[a]	0.186[a]
InP	0.69	0.11	0.21
InAs	0.36	0.026	0.14
InSb	0.38	0.014	0.10

[a] Theoretical

Table 30.21 Conduction-band (a_c) and valence-band deformation potentials (a_v, b, d) for some cubic III–V binaries. ZB = zinc blende

Binary	Conduction band	Valence band		
	a_c (eV)	a_v (eV)	b (eV)	d (eV)
AlN (ZB)	−11.7[a]	−5.9[a]	−1.7[a]	−4.4[a]
AlP	−5.54[a]	3.15[a]	−1.5[a]	
AlAs	−5.64[a]	−2.6[a]	−2.3[a]	
AlSb	−6.97[a]	1.38[a]	−1.35	−4.3
β-GaN	−21.3[a]	−13.33[a]	−2.09[a]	−1.75[a]
GaP	−7.14[a]	1.70[a]	−1.7	−4.4
GaAs	−11.0	−0.85	−1.85	−5.1
GaSb	−9	0.79[a]	−2.4	−5.4
InP	−11.4	−0.6	−1.7	−4.3
InAs	−10.2	1.00[a]	−1.8	−3.6
InSb	−15	0.36[a]	−2.0	−5.4

[a] Theoretical

Table 30.22 Conduction-band (D_i) and valence-band deformation potentials (C_i) for some wurtzite III–V binaries (in eV)

Binary	Conduction band		Valence band							
	D_1	D_2	C_1	D_1–C_1	C_2	D_2–C_2	C_3	C_4	C_5	C_6
AlN	−10.23[a]	−9.65[a]	−12.9[a]		−8.4[a]		4.5[a]	−2.2[a]	−2.6[a]	−4.1[a]
α-GaN	−9.47[a]	−7.17[a]	−41.4	−3.1	−33.3	−11.2	8.2	−4.1	−4.7	
InN				−4.05[a]		−6.67[a]	4.92[a]	−1.79[a]		

[a] Theoretical

reported on the deformation potential values for III–V alloys, $Al_xGa_{1-x}As$ [30.12], GaP_xAs_{1-x} [30.21] and $Al_xIn_{1-x}As$ [30.22]. Due to the large scatter in the experimental binary endpoint values, it is very difficult to establish any evolution of the deformation potentials with composition. We list in Table 30.21 the recommended values for the conduction band (a_c) and valence band deformation potentials (a_v, b, d) of some cubic III–V binaries. The deformation potentials for some wurtzite III–V semiconductors are also collected in Table 30.22. Until more precise data become available, we suggest employing the linear interpolation expressions in order to estimate the parameter values of these poorly explored properties.

30.7 Optical Properties

30.7.1 The Reststrahlen Region

It should be noted that in homopolar semiconductors like Si and Ge, the fundamental vibration has no dipole moment and is infrared inactive. In heteropolar semiconductors, such as GaAs and InP, the first-order dipole moment gives rise to a very strong absorption band associated with optical modes that have a $\boldsymbol{k}$ vector of essentially zero (i. e., long-wavelength optical phonons). This band is called the reststrahlen band. Below this band, the real part of the dielectric constant asymptotically approaches the static or low-frequency dielectric constant ε_s. The optical constant connecting the reststrahlen near-infrared spectral range is called the high-frequency or optical dielectric constant ε_∞. The value of ε_∞ is, therefore, measured for frequencies well above the long-wavelength LO phonon frequency but below the fundamental absorption edge.

The general properties of ε_s and ε_∞ for a specific family of compounds, namely III–V and II–VI compounds, suggest that the dielectric constants in alloy semiconductors could be deduced by using the linear interpolation method [30.26]. The simplest linear interpolation method is to use (30.1), (30.3) or (30.5). The linear interpolation scheme based on the Clausius–Mosotti relation can also be obtained from (30.6). In Fig. 30.6, we show the interpolated ε_∞ as a function of x for the $Al_xGa_{1-x}Sb$ ternary. The solid and dashed lines are, respectively, calculated from (30.1) and (30.6) (ternary). The experimental data are taken from *Lu-*

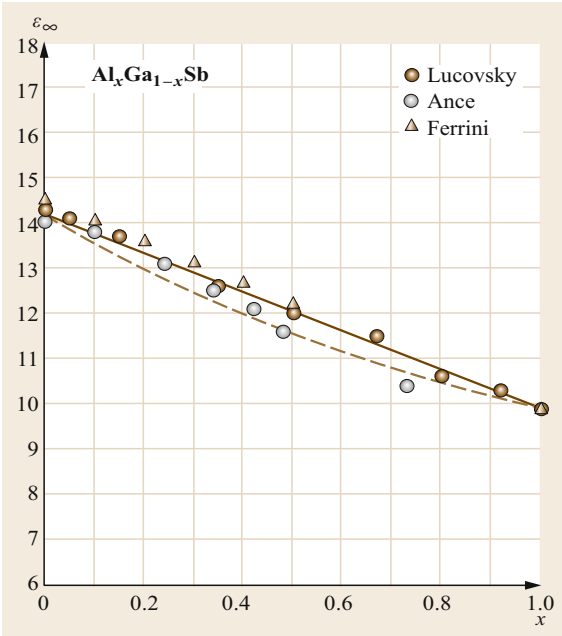


Fig. 30.6 High-frequency dielectric constant ε_∞ as a function of x for the $Al_xGa_{1-x}Sb$ ternary. The experimental data are taken from *Lucovsky* et al. [30.23] (*solid circles*), *Ance* and *Van Mau* [30.24] (*open circles*), and *Ferrini* et al. [30.25] (*open triangles*). The *solid* and *dashed lines* are, respectively, calculated from (30.1) and (30.6) (ternary) with the binary endpoint values in Table 30.23

Table 30.23 Static (ε_s) and high-frequency dielectric constants (ε_∞) for some cubic III–V binaries. ZB = zinc blende

Binary	ε_s	ε_∞
AlN (ZB)	8.07[a]	4.25
AlP	9.6	7.4
AlAs	10.06	8.16
AlSb	11.21	9.88
β-GaN	9.40[a]	5.35[a]
GaP	11.0	8.8
GaAs	12.90	10.86
GaSb	15.5	14.2
InN (ZB)	12.2[a]	7.92[a]
InP	12.9	9.9
InAs	14.3	11.6
InSb	17.2	15.3

[a] Calculated or estimated

covsky et al. [30.23], *Ance* and *Van Mau* [30.24], and *Ferrini* et al. [30.25]. The binary endpoint values used in the calculation are listed in Table 30.23. These two methods are found to provide almost the same interpolated values. Table 30.24 also lists the ε_s and ε_∞

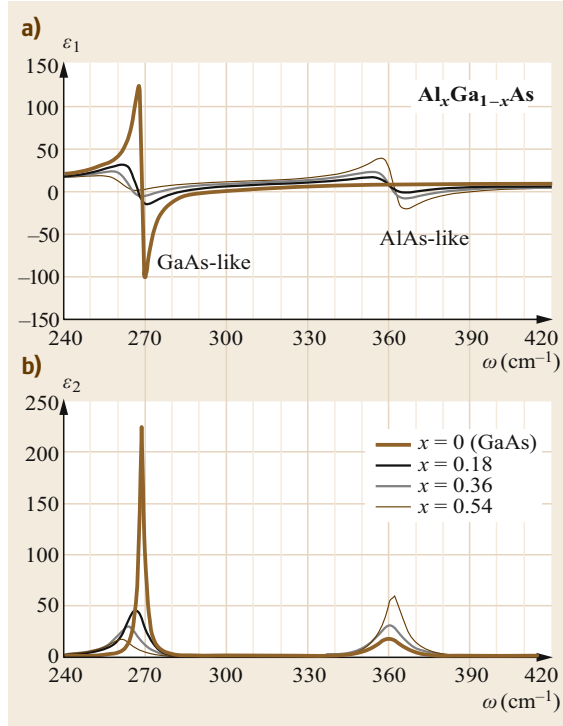


Fig. 30.7 $\varepsilon(\omega)$ spectra in the reststrahlen region of the $Al_xGa_{1-x}As$ ternary

Table 30.24 Static (ε_s) and high-frequency dielectric constants (ε_∞) for some wurtzite III–V binaries

Binary	$E \perp c$		$E \parallel c$	
	ε_s	ε_∞	ε_s	ε_∞
AlN	8.3	4.4	8.9	4.8
α-GaN	9.6	5.4	10.6	5.4
InN	10.6[a]	7.03[a]	12.3[a]	7.41[a]

[a] Estimated

values for some wurtzite III–V binary semiconductors.

The optical spectra observed in the reststrahlen region of alloy semiconductors can be explained by the following multioscillator model [30.12]

$$\varepsilon(\omega) = \varepsilon_\infty + \sum_j \frac{S_j \omega_{\mathrm{TO}j}^2}{\omega_{\mathrm{TO}j}^2 - \omega^2 - \mathrm{i}\omega\gamma_j}\,, \qquad (30.22)$$

where $S_j = \varepsilon_\infty\,(\omega_{\mathrm{LO}j}^2 - \omega_{\mathrm{TO}j}^2)$ is the oscillator strength, $\omega_{\mathrm{TO}j}$ ($\omega_{\mathrm{LO}j}$) is the TO (LO) phonon frequency, and γ_j is the damping constant of the j-th lattice oscillator. We show in Fig. 30.7, as an example, the optical spectra in the reststrahlen region of the $Al_xGa_{1-x}As$ ternary. As expected from the two-mode behavior of the long-wavelength optical phonons, the $\varepsilon(\omega)$ spec-

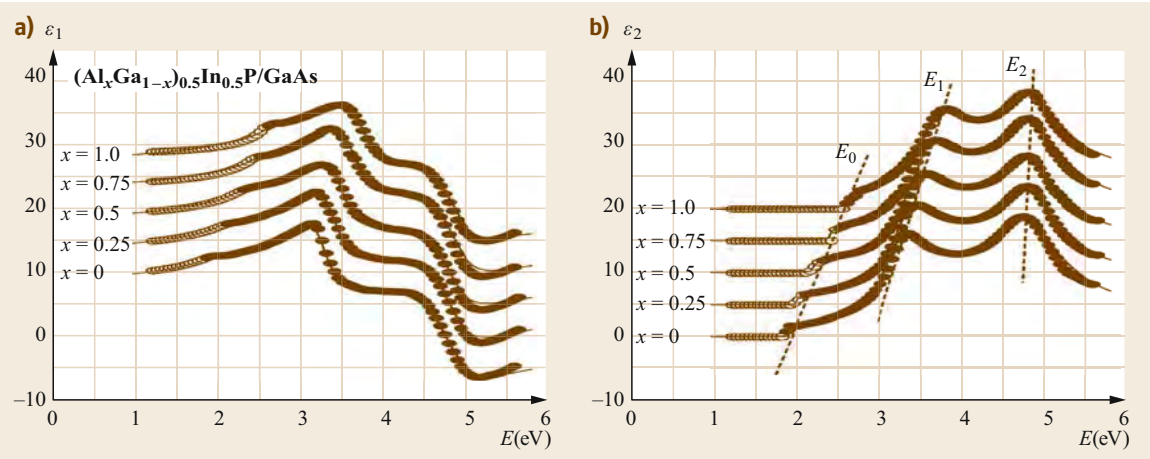


Fig. 30.8a,b $\varepsilon(E)$ spectra for $Al_xGa_yIn_{1-x-y}P/GaAs$ at room temperature. The experimental data are taken from *Adachi* [30.27]; *open* and *solid circles*. The *solid lines* represent the theoretical fits for the MDF calculation

tra of $Al_xGa_{1-x}As$ exhibit two main optical resonances: GaAs-like and AlAs-like.

30.7.2 The Interband Transition Region

The optical constants in the interband transition regions of semiconductors depend fundamentally on the electronic energy band structure of the semiconductors. The relation between the electronic energy band structure and $\varepsilon_2(E)$ is given by

$$\varepsilon_2(E) = \frac{4e^2\hbar^2}{\pi\mu^2E^2}\int \mathrm{d}\boldsymbol{k}\,|P_{\mathrm{cv}}(\boldsymbol{k})|^2\,\delta[E_{\mathrm{c}}(k)-E_{\mathrm{v}}(k)-E]\,, \tag{30.23}$$

where μ is the combined density of states mass, the Dirac δ function represents the spectral joint density of states between the valence-band $[E_{\mathrm{v}}(\boldsymbol{k})]$ and conduction-band states $[E_{\mathrm{c}}(\boldsymbol{k})]$, differing by the energy $E = \hbar\omega$ of the incident light, $P_{\mathrm{cv}}(\boldsymbol{k})$ is the momentum matrix element between the valence-band and conduction-band states, and the integration is performed over the first Brillouin zone. The Kramers–Kronig relations link $\varepsilon_2(E)$ and $\varepsilon_1(E)$ in a manner that means that $\varepsilon_1(E)$ can be calculated at each photon energy if $\varepsilon_2(E)$ is known explicitly over the entire photon energy range, and vice versa. The Kramers–Kronig relations are of fundamental importance in the analysis of optical spectra [30.9].

The refractive indices and absorption coefficients of semiconductors are the basis of many important applications of semiconductors, such as light-emitting diodes, laser diodes and photodetectors. The optical constants of III–V binaries and their ternary and quaternary alloys have been presented in tabular and graphical forms [30.27]. We plot in Fig. 30.8 the $\varepsilon(E)$ spectra for $Al_xGa_yIn_{1-x-y}P/GaAs$ taken from tabulation by *Adachi* ([30.27]; open and solid circles). The solid lines represent the theoretical fits of the model dielectric function (MDF) calculation [30.9]. The three major features of the spectra seen in Fig. 30.8 are the E_0, E_1 and E_2 structures at ≈ 2, ≈ 3.5 and ≈ 4.5 eV, respectively. It is found that the E_0 and E_1 structures move to higher energies with increasing x, while the E_2 structure does not do so to any perceptible degree. We can see that the MDF calculation enables us to calculate the optical spectra for optional compositions of alloy semiconductors with good accuracy.

30.8 Carrier Transport Properties

An accurate comparison between experimental mobility and theoretical calculation is of great importance for the determination of a variety of fundamental material parameters and carrier scattering mechanisms. There are various carrier scattering mechanisms in semiconductors, as schematically shown in Fig. 30.9. The effect of the individual scattering mechanisms on the total calculated carrier mobility can be visualized using Matthiessen's rule:

$$\frac{1}{\mu_{\mathrm{tot}}} = \sum_i \frac{1}{\mu_i}\,. \tag{30.24}$$

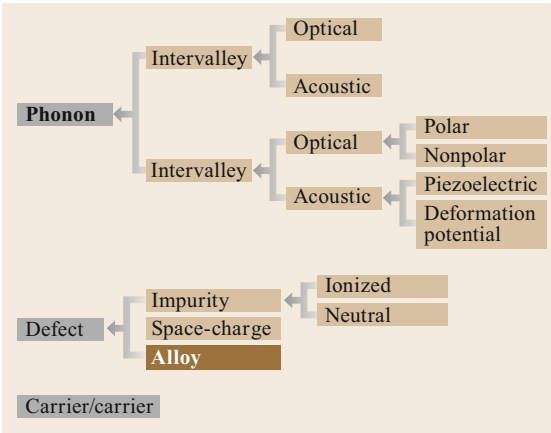


Fig. 30.9 Various possible carrier scattering mechanisms in semiconductor alloys

Table 30.25 Hall mobilities for electrons (μ_e) and holes (μ_h) obtained at 300 K for relatively pure samples of III–V binaries (in $cm^2/(V\,s)$)

Binary	μ_e	μ_h
AlN	125	14
AlP	80	450
AlAs	294	105
AlSb	200	420
α-GaN	1245	370
β-GaN	760	350
GaP	189	140
GaAs	9340	450
GaSb	12 040	1624
InN	3100	39
InP	6460	180
InAs	33 000	450
InSb	77 000	1100

The total carrier mobility μ_{tot} can then be obtained from the scattering-limited mobilities μ_i of each scattering mechanism. We note that in alloy semiconductors the charged carriers see potential fluctuations as a result of the composition disorder. This kind of scattering mechanism, so-called alloy scattering, is important in some III–V ternaries and quaternaries. The alloy scattering limited mobility in ternary alloys can be formulated as

$$\mu_{al} = \frac{\sqrt{2\pi}e\hbar^4 N_{al}\alpha}{3(m_c^*)^{5/2}(kT)^{1/2}x(1-x)(\Delta U)^2}\,, \qquad (30.25)$$

where N_{al} is the density of alloy sites, m_c^* is the electron or hole conductivity mass, x and $(1-x)$ are the mole fractions of the binary endpoint materials, and ΔU is the alloy scattering potential. The factor α is caused by the band degeneracy and is given by $\alpha = 1$ for electrons and by $\alpha = [(d^{5/2}+d^3)/(1+d^{3/2})^2]$ for holes with $d = m_{HH}/m_{LH}$, where m_{HH} and m_{LH} are the heavy hole and light hole band masses, respectively [30.12].

Let us simply express the total carrier mobility μ_{tot} in alloy $A_xB_{1-x}C$ as

$$\frac{1}{\mu_{tot}(x)} = \frac{1}{x\mu_{tot}(AC)+(1-x)\mu_{tot}(BC)} + \frac{1}{\mu_{al,0}/\frac{x}{1-x}}\,. \qquad (30.26)$$

The first term in (30.26) comes from the linear interpolation scheme and the second term accounts for the effects of alloying.

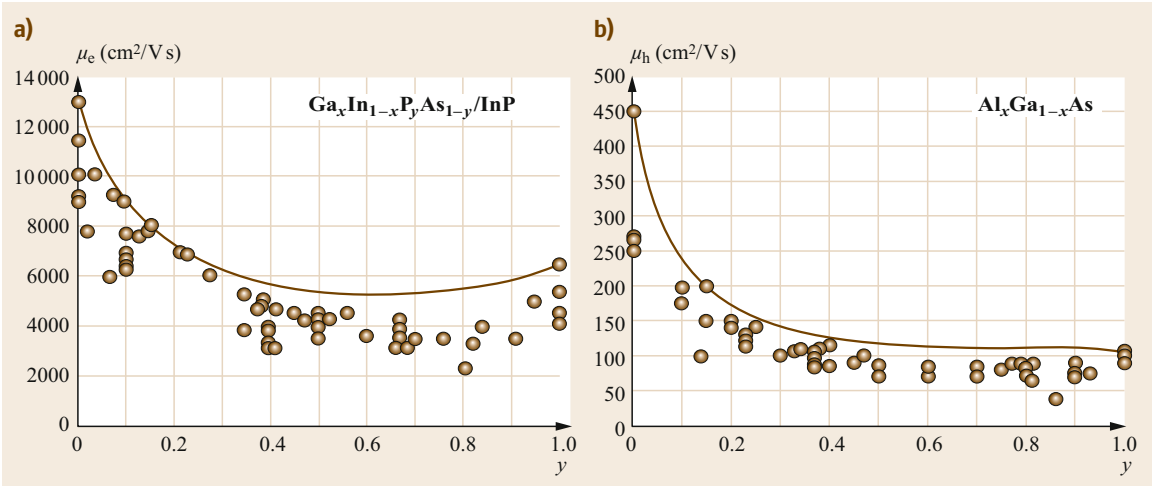


Fig. 30.10 **(a)** Electron Hall mobility μ_e in the $Ga_xIn_{1-x}P_y\,As_{1-y}$/InP quaternary and **(b)** the hole Hall mobility μ_h in the $Al_xGa_{1-x}As$ ternary, respectively. The experimental data correspond to those for relatively pure samples. The *solid lines* in **(a)** and **(b)** represent the results calculated using (30.26) with $\mu_{al,0} = 3000$ and $50\,cm^2/(V\,s)$, respectively

We plot in Figs. 30.10a and 30.10b the electron Hall mobility in the $Ga_xIn_{1-x}P_yAs_{1-y}/InP$ quaternary (μ_e) and the hole Hall mobility in the $Al_xGa_{1-x}As$ ternary, respectively. The experimental data correspond to those for relatively pure samples [30.28]. The solid lines in Figs. 30.10a,b represent the results calculated using (30.26) with $\mu_{al,0} = 3000$ and $50\,cm^2/(V\,s)$, respectively. The corresponding binary endpoint values for μ_{tot} are listed in Table 30.25. For $Ga_xIn_{1-x}P_yAs_{1-y}/InP$, we have considered the quaternary to be an alloy of the constituents $Ga_{0.53}In_{0.47}As$ ($y = 0$) and InP ($y = 1.0$) and we have used the value of μ_{tot} ($Ga_{0.47}In_{0.53}As$) $= 13\,000\,cm^2/(V\,s)$. It is clear that (30.26) can successfully explain the peculiar composition dependence of the carrier mobility in the semiconductor alloys.

References

30.1 A. Mascarenhas: *Spontaneous Ordering in Semiconductor Alloys* (Kluwer, New York 2002)
30.2 S. Adachi: *Physical Properties of III–V Semiconductor Compounds: InP, InAs, GaAs, GaP, InGaAs, and InGaAsP* (Wiley-Interscience, New York 1992)
30.3 S. Adachi: *Handbook on Physical Properties of Semiconductors*, III–V Compound Semiconductors, Vol. 2 (Springer, Berlin, Heidelberg 2004)
30.4 W. Martienssen (Ed.): *Landolt-Börnstein, Group III/41 Semiconductors, A1α Lattice Parameters* (Springer, Berlin, Heidelberg 2001)
30.5 D.Y. Watts, A.F.W. Willoughby: J. Appl. Phys. **56**, 1869 (1984)
30.6 D. Cáceres, I. Vergara, R. González, E. Monroy, F. Calle, E. Muñoz, F. Omnès: J. Appl. Phys. **86**, 6773 (1999)
30.7 M. Krieger, H. Sigg, N. Herres, K. Bachem, K. Köhler: Appl. Phys. Lett. **66**, 682 (1995)
30.8 W.E. Hoke, T.D. Kennedy, A. Torabi: Appl. Phys. Lett. **79**, 4160 (2001)
30.9 S. Adachi: *Optical Properties of Crystalline and Amorphous Semiconductors: Materials and Fundamental Principles* (Kluwer, Boston 1999)
30.10 C. Pickering: J. Electron. Mater. **15**, 51 (1986)
30.11 D.H. Jaw, Y.T. Cherng, G.B. Stringfellow: J. Appl. Phys. **66**, 1965 (1989)
30.12 S. Adachi: *GaAs and Related Materials: Bulk Semiconducting and Superlattice Properties* (World Scientific, Singapore 1994)
30.13 A.N.N. Sirota, A.M. Antyukhov, V.V. Novikov, V.A. Fedorov: Sov. Phys. Dokl. **26**, 701 (1981)
30.14 J.L. Pichardo, J.J. Alvarado-Gil, A. Cruz, J.G. Mendoza, G. Torres: J. Appl. Phys. **87**, 7740 (2000)
30.15 I. Kudman, R.J. Paff: J. Appl. Phys. **43**, 3760 (1972)
30.16 J. Bąk-Misiuk, H.G. Brühl, W. Paszkowicz, U. Pietsch: Phys. Status Solidi (a) **106**, 451 (1988)
30.17 S. Adachi: J. Appl. Phys. **54**, 1844 (1983)
30.18 I. Vurgaftman, J.R. Meyer, L.R. Ram-Mohan: J. Appl. Phys. **89**, 5815 (2001)
30.19 I.A. Buyanova, W.M. Chen, B. Monemar: MRS Internet J. Nitride Semicond. Res. **6**, 2 (2001)
30.20 D.L. Young, J.F. Geisz, T.J. Coutts: Appl. Phys. Lett. **82**, 1236 (2003)
30.21 Y. González, G. Armelles, L. González: J. Appl. Phys. **76**, 1951 (1994)
30.22 L. Pavesi, R. Houdré, P. Giannozzi: J. Appl. Phys. **78**, 470 (1995)
30.23 G. Lucovsky, K.Y. Cheng, G.L. Pearson: Phys. Rev. B **12**, 4135 (1975)
30.24 C. Ance, N. Van Mau: J. Phys. C **9**, 1565 (1976)
30.25 R. Ferrini, M. Galli, G. Guizzetti, M. Patrini, A. Bosacchi, S. Franchi, R. Magnanini: Phys. Rev. B **56**, 7549 (1997)
30.26 S. Adachi: J. Appl. Phys. **53**, 8775 (1982)
30.27 S. Adachi: *Optical Constants of Crystalline and Amorphous Semiconductors: Numerical Data and Graphical Information* (Kluwer, Boston 1999)
30.28 M. Sotoodeh, A.H. Khalid, A.A. Rezazadeh: J. Appl. Phys. **87**, 2890 (2000)

31. Group III Nitrides

Romualdo A. Ferreyra, Congyong Zhu, Ali Teke, Hadis Morkoç

Optical, electrical, mechanical, and thermal properties of group III nitrides, inclusive of AlN, GaN, InN and their ternary and quaternary alloys, are discussed. The driving force for group III nitride semiconductors is their important applications in optoelectronics, microwave amplifiers, and high voltage power switches. Owing to the aforementioned applications, the fundamental properties of each III nitride binary, as well as the alloys that have been acquired, are discussed in this chapter. In general, an appropriate assessment of the material properties for any material is not straightforward to begin with and the nitride family is no exception, particularly considering that group III nitrides are prepared on foreign substrates as low-cost native substrates are not yet available. Understandably, precise measurements of the mechanical, thermal, electrical and optical properties of the semiconductor nitride family are imperative for further advances. Notwithstanding the great progress that has already been made to further understand and exploit group III nitrides, especially GaN, reliable data for AlN and InN are still in the state of evolution, and naturally the subject of some controversy. This is, in part, a consequence of measurements having been performed on samples of widely varying quality. When possible, the spurious discrepancies have been disregarded. For some materials, too few measurements are available to yield a consensus, in which case the available data are simply reported. Defects in group III nitrides as well as GaN-based nanostructures are also discussed.

31.1 **Crystal Structures of Nitrides** 745
31.2 **Lattice Parameters of Nitrides** 746
31.3 **Mechanical Properties of Nitrides** 748
31.4 **Thermal Properties of Nitrides** 752
31.4.1 Thermal Expansion Coefficients 752
31.4.2 Thermal Conductivity 753
31.4.3 Specific Heat 756
31.5 **Electrical Properties of Nitrides** 757
31.5.1 Low Field Transport 758
31.5.2 High Field Transport 766
31.6 **Optical Properties of Nitrides** 769
31.6.1 Gallium Nitride 769
31.6.2 Aluminum Nitride 777
31.6.3 Indium Nitride 781
31.7 **Properties of Nitride Alloys** 782
31.8 **Doped GaN** 786
31.8.1 n-Type Doping 787
31.8.2 p-Type Doping 787
31.9 **Defects in GaN** 789
31.9.1 Points Defects 789
31.9.2 Extended Defects 792
31.10 **GaN-Based Nanostructures** 794
31.10.1 Quantum Wells 794
31.10.2 Quantum Dots 796
31.10.3 Vertical Cavities 796
31.10.4 Nitride Nanorods 799
31.11 **Summary and Conclusions** 801
References 802

During the last three decades, developments in the field of group III nitrides have been spectacular, with major breakthroughs taking place in the 1990s. They have been viewed as a highly promising material system for electronic and optoelectronic applications. As members of the group III nitrides family, AlN and GaN and their alloys with InN are all wide-bandgap materials (except for InGaN with very high In content) and can crystallize in both wurtzite and zincblende polytypes. The bandgaps of the wurtzite polytypes are direct and range from a possible value of ≈ 0.8 eV for InN, to 3.4 eV for GaN, and to 6.1 eV for AlN. GaN alloyed with AlN and InN may span a continuous range of direct-bandgap energies throughout much of the visible spectrum, well into ultraviolet (UV) wavelengths. This makes the nitride system attractive for optoelectronic applications, such as light-emitting diodes (LEDs), laser diodes (LDs), and UV detectors. Commercialization of bright

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_31

blue and green LEDs and the possibility of yellow LEDs paved the way for developing full-color displays. If the three primary-color LEDs, including red, produced by the InGaAlAs system are used in place of incandescent light bulbs in some form of a color-mixing scheme, they would provide not only compactness and longer lifetime, but also lower power consumption for the same luminous flux output. Additional possible applications include use in agriculture as light sources for accelerated photosynthesis, and in health care for diagnosis and treatment. Unlike display and lighting applications, digital information storage and reading require coherent light sources because the diffraction-limited optical storage density increases approximately quadratically with decreasing wavelength. The nitride material system, when adapted to semiconductor lasers in blue and UV wavelengths, offers increased data storage density, possibly as high as 50 Gb per disc with 25 Gb promised soon in the blu-ray system. Other equally attractive applications envisioned include printing and surgery. When used as UV sensors in jet engines, automobiles, and furnaces (boilers), the devices would allow optimal fuel efficiency and control of effluents for a cleaner environment. Moreover, visible-blind and solar-blind nitride-based photodetectors are also an ideal candidate for a number of applications including early missile-plume detection, UV astronomy, space-to-space communication, and biological effects. Another area gaining a lot of attention for group III–V nitrides is high-temperature/high-power electronic applications, such as radar, missiles, and satellites as well as in low-cost compact amplifiers for wireless base stations and high-voltage power switches, due to their excellent electron transport properties, including good mobility and high saturated drift velocity. The strongest feature of the group III nitrides compared to other wide-bandgap counterparts is the heterostructure technology that it can support. Quantum wells, modulation-doped heterointerfaces, and heterojunction structures can all be made in this system, giving access to new spectral regions for optical devices and new operational regimes for electronic devices. Other attractive properties of the nitrides include high mechanical and thermal stability, and large piezoelectric constants.

One of the main difficulties that has hindered group III nitride research is the lack of a lattice-matched and thermally compatible substrate material. A wide variety of materials have been studied for nitride epitaxy, including insulating metal oxides, metal nitrides, and other semiconductors. In practice, properties other than the lattice constants and thermal compatibility, including the crystal structure, surface finishing, composition, reactivity, and chemical and electrical properties, are also important in determining suitability as a substrate. The substrate employed determines the crystal orientation, polarity, polytype, surface morphology, strain, and the defect concentration of the epitaxial films. The most promising results on more conventional substrates so far have been obtained on sapphire, and SiC, however, accelerating progress has been made on Si substrates due to marketing issues, such as low-cost and compatibilities with Si processes. Also coming on the scene are thick freestanding GaN templates. Group III–V nitrides have been grown on Si, NaCl, GaP, InP, SiC, W, ZnO, $MgAl_2O_4$, TiO_2, and MgO. Other substrates have also been used for nitride growth, including Hf, $LiAlO_2$ and $LiGaO_2$. Lateral (lattice constant a) mismatched substrates lead to substantial densities of misfit and threading dislocations in broad-area epitaxially deposited GaN on foreign substrate, in the range 10^9–$10^{10}\,cm^{-2}$. An appropriate surface preparation such as nitridation, deposition of a low-temperature (LT) AlN or GaN buffer layer, selective epitaxy followed by a type of coalescence called lateral epitaxial overgrowth (LEO) or epitaxial lateral overgrowth (ELOG) can reduce dislocation densities down to $10^6\,cm^{-2}$. However, these numbers are still high compared to extended-defect densities of essentially zero for silicon homoepitaxy, and 10^2–$10^4\,cm^{-2}$ for gallium arsenide homoepitaxy. Vertical (lattice constant c) mismatch creates additional crystalline defects besetting the layers, including inversion domain boundaries and stacking faults. In addition, mismatch of thermal expansion coefficients between the epitaxial films and the substrate induces stress, which can cause crack formation in the film and substrate for thick films during cooling from the deposition temperature. A high density of defects, which increases the laser threshold current, causes reverse leakage currents in junctions, depletes sheet charge-carrier density in heterojunction field-effect transistors, reduces the charge-carrier mobility and thermal conductivity, and is detrimental to device applications and the achievement of their optimal performance. Thus, substrates capable of supporting better-quality epitaxial layers are always needed to realize the full potential of nitride-based devices. Nearly every major crystal-growth technique has been developed, including molecular beam epitaxy (MBE), hydride vapor-phase epitaxy (HVPE), and metalorganic chemical vapor deposition (MOCVD), in relation to nitride semiconductors. Several modifications to the conventional MBE method have been implemented for group III nitride growth: growth with ammonia or hydrazine (the latter is not attractive due to safety reasons and success of ammonia), plasma-assisted MBE (PAMBE), metalorganic MBE (MOMBE), pulsed laser deposition (PLD), and so on. Among other methods, radio-frequency (RF) and

electron-cyclotron resonance (ECR) plasma sources are the most commonly employed devices to activate the neutral nitrogen species in the MBE environment. Although all of these epitaxial methods contend with problems related to the lack of native GaN substrates, and difficulty with nitrogen incorporation, remarkable progress in the growth of high-quality epitaxial layers of group III nitrides by a variety of methods has been achieved. Additionally, growth of bulk GaN substrates by HVPE and ammonothermal methods have made tremendous progress in terms of dislocation density, impurities, growth speed, and so on.

Although many applications based on nitride semiconductors have emerged, and some of them are commercially available, as discussed throughout this chapter, there are many contradictions in identification of the basic physical properties of these materials. In this respect, they are not yet mature. Additionally, knowledge of the fundamental properties is crucial not only from the physics point of view but also when understanding and optimizing the device structures for better performance. In this chapter, therefore, we present the updated fundamental properties of GaN, AlN and InN, including structural, mechanical, thermal, electrical, and optical properties. The aim is to assist readers newly entering this field and other interested researchers in accessing the most recent available data. The reader is also urged to peruse the following publications for more detailed information on several aspects of ongoing research in group III nitrides. These consist of handbooks [31.1–3], and regular and review papers [31.4–31].

31.1 Crystal Structures of Nitrides

The group III nitrides share three crystal structures, which are wurtzite, zincblende, and rocksalt, where the thermodynamical stable phase is wurtzite for bulk AlN, GaN, and InN, and the cohesive energy per bond for the wurtzite phase is 2.88 eV (63.5 kcal/mol), 2.20 eV (48.5 kcal/mol), and 0.67 eV (15.5 kcal/mol) for AlN, GaN, and InN respectively [31.32]. The calculated energy difference between wurtzite and zincblende is small, which is -18.41 meV/atom for AlN, -9.88 meV/atom for GaN, and -11.44 meV/atom for InN; however, the wurtzite phase is energetically favorable for all three nitrides compared to zincblende. Being in the space group of $P6_3mc$, the wurtzite structure has a hexagonal unit cell with the a and c lattice parameters in the ratio of $c/a = \sqrt{8/3} = 1.633$. The wurtzite structure is composed of two interpenetrating hexagonal close-packed (hcp) sublattices, but displaced with respect to each other along the three-fold c-axis by an amount of $u = 3/8 = 0.375$, where each sublattice consists of four atoms per unit cell and every group III atom is surrounded tetrahedrally by four nitrogen atoms, or vice versa, these being located at the edges of a tetrahedron. The actual nitrides deviate from the above-mentioned ideal structure, which is signified by the c/a ratio or the u value [31.33]. These deviations come with the feature that the u parameter increases when the c/a ratio decrease while the four tetrahedral distances remain nearly constant through a distortion of the tetrahedral angles due to long-range polar interactions. These two slightly different bond lengths will be equal if the following equation holds

$$u = \left(\frac{1}{3}\right)\left(\frac{a^2}{c^2}\right) + \frac{1}{4}\,. \tag{31.1}$$

The c/a ratio also depends on the dispersion in the electronegativities of the two constituents in group III nitrides with larger differences resulting in larger departures from an ideal c/a ratio [31.34]. These two parameters can be measured by using the four-circle diffractometry technique. In the case of GaN, the c/a ratio and the u parameter are measured as 1.627 and 0.377, respectively, which are close to the ideal values [31.35], while the available AlN deviates significantly from ideal parameters: $c/a = 1.601$ and $u = 0.382$. Thus, the interatomic distance and angles differ by 0.01 Å and 3°, respectively. Due to the lack of suitable size single-crystal InN for single-crystal diffractometry measurement, reported values are rare and may not be reliable. One popularly cited c/a value is 1.611 [31.36], which is consistent with data reported later in InN crystalline films grown by MBE [31.37] and MOCVD [31.38].

A phase transition to the rocksalt (NaCl) structure in group III nitrides can take place at very high external pressures due to the reduction of the lattice dimensions, which causes inter-ionic Coulomb interaction to favor ionicity over covalent nature. The structural phase transition was observed to occur at the following pressure values: 22.9 GPa for AlN [31.39], 52.2 GPa for GaN [31.40], and 12.1 GPa for InN [31.41]; however, the rocksalt structure of group III nitrides is not stable during epitaxial growth. With a six-fold coordinated structure, the rocksalt type of structure belongs to space-group symmetry of $Fm3m$.

The zincblende structure can be stabilized only by heteroepitaxial growth on cubic structure substrates, such as cubic SiC [31.42], Si [31.43], MgO [31.44], and GaAs [31.45]; however, they are metastable and tend to

form the wurtzite phase due to the inherent tendency. Thus in the case of highly mismatched substrates, there can be a certain amount of zincblende phase of nitrides separated by crystallographic defects from the wurtzite phase. However, over the years with considerably improved preparation techniques, this polyphase issue has substantially diminished. The crystal structure of the zincblende structure is composed of two interpenetrating face-centered cubic (fcc) sublattices displaced by one quarter of a body diagonal, which belongs to the space group of $F\bar{4}3m$. The unit cell is composed of four atoms, where every atom of one type is tetrahedrally coordinated with four atoms of other type, and vice versa. The overall equivalent bond length for zincblende structures is about 1.623 Å.

Although zincblende and wurtzite structures share the same tetrahedral coordination, the main difference between these two structures comes from the stacking sequence of close-packed diatomic planes. The stacking sequence for the wurtzite structure in the ⟨0001⟩ direction is AaBbAaBb taking Ga and N pairs as an example. In contrast, the close-packed (111) planes along the ⟨111⟩ direction in the zincblende structure have a 60° rotation that causes a stacking order of AaBbCcAaBbCc, where the small and large letters stand for the two different kinds of constituents. Because none of the three available structures for group III nitrides possess inversion symmetry, the crystal structure exhibits crystallographic polarity. The (0001) planes are terminated by group III (Al, Ga, or In) atoms and denoted as Ga polarity (*c*-plane), while $(000\bar{1})$ planes are terminated by nitrogen and denoted as N polarity, and these two planes are referred to as polar planes. Many material properties strongly depend on the polarity, thus it is critical to choose certain planes for desired device properties. For example, GaN-based heterojunction field effect transistors (HFETs) are predominantly based on *c*-plane GaN films due to the large polarization charge. While for the purpose of LED applications, nonpolar (i. e., $(10\bar{1}0)$ plane or *m*-plane, and $(11\bar{2}0)$ plane or *a*-plane) or semipolar $(11\bar{2}2)$ planes are preferred due to the reduced polarization induced internal electric field under these planes [31.3], on the condition that all else are equal, which is not the case because of technological disparity. Considering the cost and maturity, *c*-plane wurtzite nitrides are the most commonly used planes for device applications, although active research is underway for other secondary planes and directions.

31.2 Lattice Parameters of Nitrides

The lattice parameters of nitride-based semiconductors, as with other semiconductors [31.46, 47], depend on the following factors [31.48]:

1. Free electron concentration, which expands the lattice proportionally to the deformation potential of the conduction-band minimum
2. The concentration of impurities and intrinsic defects, and the difference between their ionic radii and the substituted matrix ions
3. External strains (for example, those induced by substrate or heterostructures)
4. Temperature.

High-resolution x-ray diffraction (HRXRD) is the most common tool to measure the lattice parameters of any crystalline material, and the results are usually obtained at a standard temperature of 21 °C [31.49] by using the Bond method [31.50] for a set of symmetrical and asymmetrical reflections. Because the lattice parameters of ternary compounds depend on the composition, the measured lattice parameters can also be used to determine the composition of the compounds, where the strain must be taken into consideration for precise determination of the composition. For nitrides, an accuracy of about 0.1% or less can be achieved in determining the composition, equivalently down to a mole fraction of about 1%, if the elastic parameters of all nitrides and lattice parameters of AlN and InN are taken into consideration. Considering that different growth techniques and/or substrates are used, which inevitably induce different strain states, defect densities and free-carrier concentrations, there is a wide dispersion in the reported values. Table 31.1 shows a comparison of measured and calculated lattice parameters reported by several groups for AlN, GaN, and InN crystallized in the wurtzite structure. The AlN crystal has a molar mass of 20.495 g/mol when it crystallizes in the hexagonal wurtzite structure. The measured lattice parameters for AlN in either bulk or epitaxial films range from 3.110–3.112 Å for the *a* parameter and from 4.980–4.982 Å for the *c* parameter, which gives a c/a ratio in the range of 1.600 and 1.602. The deviation from that of the ideal wurtzite crystal is most likely due to lattice stability, iconicity, and defects. The influence of carbon dopant concentration on the lattice parameters of AlN substrates grown by hydride vapor phase epitaxy using AlN substrates prepared by the physical vapor transport technique has been reported [31.51]. For samples with concentrations of carbon in the HVPE substrates from $< 2\times10^{17}$ to $< 1\times10^{19}\,\mathrm{cm}^{-3}$, while the concentrations of oxygen and silicon remained below $5\times10^{17}\,\mathrm{cm}^{-3}$ for all substrates, the HVPE-AlN

Table 31.1 Measured and calculated lattice constants of wurtzite AlN, GaN and InN

Compound	Sample	a (Å)	c (Å)	Reference
AlN	PVT seeded grown bulk	3.1117(5)	4.98180(2)	[31.52]
	Free standing HVPE-AlN grown on PVT-AlN substrate	3.1111	4.9808	[31.53]
	Epitaxial layer on SiC	3.110	4.980	[31.54]
	First-principle – LDA	3.0698	4.9101	[31.55]
	First-principle – CGA	3.1095	4.9938	[31.55]
GaN	Undoped Bulk Crystal ($n = 1 \times 10^{16}$) – HVPE	3.18914(5)	5.18508(3)	[31.56]
	O doped bulk crystal ($n = 3 \times 10^{18}$) – HVPE	3.18921(5)	5.18574(3)	[31.56]
	Si doped bulk crystal ($n = 1 \times 10^{18}$) – HVPE	3.18868(5)	5.18574(3)	[31.56]
	Fe doped bulk crystal ($n = 2 \times 10^{13}$) – HVPE	3.18841(5)	5.18570(3)	[31.56]
	Bulk crystal ($n = \text{low}10^{17}$) – HVPE	3.1880(20)	5.1868(15)	[31.57]
	Nonpolar (1100) bulk ($n = 10^{19}$) – ammonothermal	3.18908(10)	5.18517(10)	[31.58]
	m-Plan bulk crystal – Ammonothermal	3.1897	5.1861	[31.59]
	GaN substrate grown using a lateral epitaxial overgrowth (LEO) technique	3.1896(2)	5.1855(2)	[31.60]
	Homoepitaxial layers (low free-electron concentration)	3.1885	5.1850	[31.61]
	Homoepitaxial layers (high free-electron concentration)	3.189	5.1864	[31.48]
	Relaxed layer on sapphire	3.1892	5.1850	[31.62]
	Relaxed layer on sapphire	3.1878	5.1854	[31.46]
	Powder	3.1893	5.1851	[31.63]
	First-principle – LDA	3.193	5.2174	[31.64]
	First-principle – CGA	3.245	5.2958	[31.64]
InN	Powder	3.538	5.703	[31.65]
	First-principle– LDA	3.544	5.762	[31.64]
	First-principle– CGA	3.614	5.883	[31.64]

substrates did not show significant deviation in the lattice parameters. The average lattice parameters are $a = 3.1110$ Å and $c = 4.9809$ Å with $3\sigma = 0.0002$ Å. There are several reports on the lattice parameters of the cubic form of AlN with a lattice parameter of $a = 4.38$ Å [31.66], which is consistent with the theoretically estimated value [31.67]. However, the zincblende polytype AlN is in a metastable state. While for the rocksalt phase of AlN, which is pressure induced, a lattice parameter of 4.043–4.045 Å was obtained at room temperature [31.68, 69].

The wurtzite (WZ) crystal structure of GaN has a molecular weight of 83.7267 g/mol with four atoms per cell. Its lattice parameters depend on a variety of factors including free-electron concentration, impurities, and growth techniques. It has been reported that free charge is the dominant factor responsible for expanding the lattice [31.48]. The lattice parameters were found to be proportionally dependent on the deformation potential of the conduction-band minimum and inversely proportional to the carrier density and bulk modulus. Impurities, intrinsic point defects and extended defects, such as threading dislocations, are also important factors determining the lattice parameters of WZ-GaN, especially because bulk or epitaxial GaN is generally obtained on foreign substrates at high temperature, and thus suffers from both the lattice constant and thermal conductivity mismatch. Depending on the growth techniques, the defect density and impurity levels vary, which can have a direct influence on the lattice parameters. For example, solution growth at high pressure (15–20 kbar) and high temperature (in the range of 1600 K) can achieve high structural quality GaN material with a dislocation density as low as 100 cm^{-2}. However, the high level of background impurities, for example oxygen, can often result in free-carrier concentration exceeding levels of 10^{19} cm^{-3}. The free-carrier density for undoped ammonothermally grown GaN substrates also reported a similar level of 10^{19} cm^{-3} [31.70–72] but can be as low as 2×10^{18} cm^{-3} [31.73]. The lattice parameters from ammonothermally grown nonpolar and semipolar GaN bulk substrates are around $a = 3.1890$–3.1900 Å and $c = 5.185$–5.186 Å [31.58, 59]. It was found that doping has a strong effect on the lattice parameters for HVPE-grown WZ-GaN [31.74]. Undoped HVPE-grown GaN has a free-carrier density of 1×10^{16} cm^{-3} and has $a = 3.18914$ Å, which increased to 3.18921 Å in oxygen-doped bulk GaN substrates with free-carrier concentrations of 3×10^{18} cm^{-3}. For Si- and Fe-doped bulk GaN substrates with free-carrier concentrations of 1×10^{18} cm^{-3} and 2×10^{13} cm^{-3} respectively, the a parameters decrease to 3.18868 and 3.18841 Å respectively. The trend for the change of lattice constant is generally consistent with the theoretical predictions based on the impurity size effect and deformation-potential effects as discussed in [31.75].

A precise analysis of the effect of doping on the lattice constant of GaN, however, requires a low-defect-density bulk material with controlled doping concentration. When and if the impurity levels can be controlled, attention then needs to be paid to the intrinsic defects. For zincblende GaN, the calculated lattice constant using the measured Ga–N bond distance in WZ-GaN is $a = 4.503$ Å. The measured lattice constant in epilayers ranges from 4.49–4.55 Å [31.42–44], where the fully relaxed lattice parameter was obtained to be 4.5036 ± 0.0004 Å [31.76]. The lattice parameter for freestanding cubic GaN grown by plasma-assisted molecular beam epitaxy on GaAs substrates, which subsequently is removed, demonstrated a lattice parameter of 4.510 ± 0.005 Å [31.77]. A high-pressure phase transition from the WZ to the rocksalt structure decreases the lattice constant down to $a_0 = 4.22$ Å in the rocksalt phase [31.78].

Due to the difficulties in the crystal growth of InN, reports on the physical properties of InN are limited. Nearly all the available data have been obtained from nonideal thin films. Indium nitride normally crystallizes in the wurtzite structure and has a molecular weight of 128.827 g/mol. The measured lattice parameters using a powder technique are in the range of $a = 3.530–3.548$ Å and $c = 5.960–5.704$ Å. By analyzing the strain in the InN film grown by MBE on sapphire substrates, the lattice parameters for strain-free InN are estimated to be in the range of $a = 3.535 \pm 0.005$ Å and $c = 5.699 \pm 0.004$ Å [31.37]. Zincblende (cubic) InN grown on sapphire (0001) using cubic indium oxide as a buffer layer generated an estimated lattice parameter of 4.9(8) Å [31.79], which is close to other reported values for epilayers grown on InAs (001) (5.04 Å) [31.80], on GaAs (001) (4.98 Å) [31.81], and on *r*-plane sapphire (4.986 Å) substrates [31.82].

31.3 Mechanical Properties of Nitrides

The mechanical properties of materials are usually characterized by parameters such as hardness, stiffness constant, Young's and bulk modulus and yield strength, and so on. For group III nitride semiconductor materials, the precise determination of the mechanical properties is hindered by various complexities, the most important of which is the lack of large size high-quality single crystals. Nevertheless, attempts have been made to estimate and measure the mechanical properties of both thin and thick epitaxial layers, and bulk crystals. The most precise technique to determine the elastic moduli of compound semiconductor materials is believed to be the ultrasonic measurement, which requires thick (about 1 cm-thick) single-crystalline samples to enable measurement of the timing of plane-wave acoustic pulses with sufficient resolution. Thus, this technique is almost inapplicable to the group III nitrides materials except wurtzite GaN where large size bulk substrates are commercially available. Brillouin scattering can determine the elastic constant and bulk moduli through the interaction of light with thermal excitation, in particular acoustic phonons, in the crystalline material. Various forms of x-ray diffraction, such as energy dispersive x-ray diffraction (EDX), angular dispersive x-ray diffraction (ADX) and x-ray absorption spectroscopy (XAS) can be used to determine the pressure dependence of the lattice parameters. By assuming that the bulk modulus has a linear dependence on the applied pressure to the crystal, it can be deduced as [31.83]

$$V = V_0 \left(1 + \frac{B'P}{B}\right)^{-\frac{1}{B'}} , \qquad (31.2)$$

where B and V_0 represent the bulk modulus and unit volume at ambient pressure respectively, and B' is the derivative of B with respect to pressure. X-ray diffraction can be used to determine the isothermal bulk modulus, and Brillouin scattering can determine the adiabatic bulk modulus. In the case of solids other than molecular solids, these two thermodynamic quantities bear no measurable difference. Predictive calculations of the structural and mechanical properties of group III nitrides can also be undertaken in addition to the above-mentioned experimental investigations. The most popular calculations are based on density functional theory within the local density approximation (LDA) using various types of exchange correlation functionals.

There are five independent elastic constants in hexagonal crystals, which are c_{11}, c_{33}, c_{12}, c_{13} and c_{44}. Among them, c_{11} and c_{33} correspond to longitudinal modes along the [1000] and [0001] directions respectively, while c_{44} and $c_{66} = (c_{11} - c_{12})/2$ can be determined from the speed of sound of the transverse modes propagating along the [0001] and [1000] directions respectively. c_{13} and the other four moduli relate to the velocity of the modes propagating in directions with lower symmetry, such as [0011]. The bulk modulus can be determined by the elastic constants through the following equation [31.84]

$$B = \frac{(c_{11} + c_{12})c_{33} - 2{c_{13}}^2}{c_{11} + c_{12} + 2c_{33} - 4c_{13}} .$$

If the material is isotropic, the Young's modulus E and shear modulus G can be evaluated by $E = 3B(1 - 2v)$

Table 31.2 Some mechanical properties of wurtzite AlN, GaN, and InN obtained by experimental measurements and theoretical calculations. The units are in GPa

Parameters	AlN (GPa)	GaN (GPa)	InN (GPa)
c_{11}	394 ± 1^{a}, 401.2 ± 0.5^{b}, 345^{c}, 396^{d}, 398^{e}	296^{r}, 390^{s}, 377^{t}, 370^{u}, 373^{an}, 367^{d}, 396^{e}	225 ± 7^{ae}, 223^{d}, 271^{e}
c_{12}	134 ± 1^{a}, 135.0 ± 0.5^{b}, 125^{c}, 137^{d}, 140^{e}	120^{r}, 145^{s}, 160^{t}, 145^{u}, 141^{an}, 135^{d}, 144^{e}	109 ± 8^{ae}, 115^{d}, 124^{e}
c_{13}	95 ± 1^{a}, 96.3 ± 22.1^{b}, 120^{c}, 108^{d}, 127^{e}	158^{r}, 106^{s}, 114^{t}, 106^{u}, 80^{an}, 103^{d}, 100^{e}	108 ± 8^{ae}, 92^{d}, 94^{e}
c_{33}	402 ± 1^{a}, 368.2 ± 27.9^{b}, 395^{c}, 373^{d}, 382^{e}	267^{r}, 398^{s}, 209^{t}, 398^{u}, 387^{an}, 405^{d}, 392^{e}	265 ± 3^{ae}, 224^{d}, 200^{e}
c_{44}	121 ± 1^{a}, 122.6 ± 0.2^{b}, 118^{c}, 116^{d}, 96^{e}	24^{r}, 105^{s}, 81^{t}, 105^{u}, 94^{an}, 95^{d}, 91^{e}	55 ± 3^{ae}, 48^{d}, 46^{e}
Poisson's ratio v	0.19^{f}, 0.287^{g}, 0.216^{g}	0.183 ± 0.003^{p}, 0.38^{v}, 0.372^{r}	0.21 ± 0.05^{ab}
Bulk modulus B	201^{c}, 210^{h}, 208^{i}, 160^{j}, 207^{d}, 218^{e}	187 ± 7 and 319 ± 10^{x}, 195^{r}, 210^{s}, 245^{am}, 237^{af}, 188^{ap}, 202^{d}, 207^{e}	152 ± 5^{ae}, 125^{af}, 141^{d}, 147^{e}
dB/dP	5.2^{j}, 6.3^{i}, 5.7^{al}, 3.74^{k}, 3.77^{l}	4^{am}, 4.3^{af}, 3.2^{ap}, 4.5^{e}, 2.9^{ak}	12.7^{af}, 3.4^{e}
Young's modulus E	308^{j}, 295^{c}, 374^{m}, 243.5^{n}	325.3 ± 5.4^{o}, 320^{q}, 305 ± 11^{w}, 295^{y}	149 ± 5^{ac}, 159^{ad}, 152.5 ± 3.9^{ag}
Shear modulus	154^{m}, 131^{j}, 117^{c}	157 ± 9.3^{o}, 152.6^{z}	43^{r}
Yield strength σ_Y	0.3 at 1000 °C^{m}	15^{y}, 0.1–0.2 at 900 °C^{ao}	
Hardness	17.7^{m}, 18.0^{m}, 16.2^{n},	28.5 ± 1.04^{o}, 19^{aa}, $18–20^{y}$	4.2 ± 0.1^{ag}, 8^{ah}, 3.6–9.1^{ai}, 11.2^{aj}

[a] Brillouin scattering measurement on PVT grown bulk AlN [31.85]; [b] Measured by ultrasonic micro-spectroscopy (UMS) technique on PVT grown bulk AlN [31.86]; [c] Ultrasonic measurement on thin film [31.87]; [d] Calculated using pseudopotential LDA[31.88]; [e] Calculated using FP-LMTO LDA [31.89]; [f] AlN thin film grown on SiC substrate [31.90]; [g] *c*-plan and *r*-plane calculated [31.91]; [h] Brillouin scattering on single crystal [31.92]; [i] ADX on single-crystal AlN [31.93]; [j] Ultrasonic measurement on sintered, isotropic, polycrystalline AlN ceramic [31.94]; [k] Calculated using plane-wave pseudopotential [31.95]; [l] Calculated using Keating-Harrison model [31.96]; [m] Hardness measurement on bulk single-crystal AlN by micro-hardness and nano-hardness, respectively [31.97]; [n] nano-indentation measurement on AlN thin films on *c*-plane sapphire substrate [31.98]; [o] nano-indentation measurement on HVPE grown bulk GaN crystal [31.99]; [p] Determined by HR-XRD on MOCVD grown *c*-plane GaN film [31.100]; [q] Microbeam bending test on suspended GaN-on-patterned-silicon technique [31.101]; [r] Elastic constants calculated from temperature-dependent x-ray diffraction on polycrystalline GaN or InN. Poisson's ratio estimated from elastic constants [31.102]; [s] Brillouin spectroscopy on bulk GaN [31.84]; [t] Resonance ultrasound method on GaN plate [31.103]; [u] Surface-acoustic-wave measurement on GaN grown on sapphire [31.104]; [v] Determined by Bond's x-ray method on heteroepitaxially grown *c*-plane GaN [31.62]; [w] Obtained from nanowires with diameters ≈ 84 nm [31.105]; [x] Results obtained from high-pressure energy-dispersive x-ray diffraction on both bulk and nanocrystalline GaN, respectively [31.106]; [y] Nano-indentation on bulk GaN [31.107]; [z] Nano-indentation examination on *c*-plane bulk GaN [31.108]; [aa] Nano-indentation measurement on lateral overgrown epitaxial GaN films [31.109]; [ab] N-polar InN films grown on GaN template by plasma-assisted MBE [31.110]; [ac] Measured by AFM microbending test on single-crystalline wurtzite InN thin films [31.111]; [ad] First-principle calculations with the plane-wave pseudopotential density functional theory method [31.112]; [ae] Determined by grazing incidence inelastic x-ray scattering on InN thin films deposited on GaN template [31.113]; [af] X-ray diffraction on bulk GaN [31.41]; [ag] Nano-indentation measurement on InN films grown on GaN template [31.114]; [ah] Measured on *c*-plane wurtzite InN films [31.115]; [ai] Measured by nano-indentation technique on single-crystalline InN films on Si(111) substrates, where the hardness values depend on the growth temperatures [31.116]; [aj] Measured by nano-indentation technique on InN grown on sapphire [31.117]; [ak] Calculated using plane-wave pseudopotential [31.118]; [al] EDXD on polycrystalline AlN [31.68]; [am] X-ray absorption spectroscopy on GaN [31.40]; [an] Brillouin spectroscopy on GaN substrate grown by LEO [31.60]; [ao] Hardness on single-crystal GaN [31.119]; [ap] EDX on bulk GaN [31.39].

and $G = E/2(1 + v)$ respectively, where the term v is the Poisson's ratio and is given by $v = c_{13}/(c_{11} + c_{12})$ [31.97].

The hardness of group III nitrides can be determined by micro- and nano-indentation methods over a wide range of size scales and temperatures. The measurements are usually conducted on the (0001) surface using a conventional pyramidal or spherical diamond tip, or alternatively, a sharp triangular indenter (Berhovich). Complete information on the hardness and pressure-induced phase transformation of semiconductor materials can be obtained via depth-sensing indentation measurements. Table 31.2 shows experimental as well as theoretical results of the mechanical parameters for wurtzite AlN, GaN and InN single crystals as well as polycrystallites reported by several groups.

Because the quality of the crystals varies, especially for InN, the experimental results are widely scattered

as demonstrated in Table 31.2. A clear example is the hardness data on InN, where the nanohardness of InN significantly depends on the growth temperature, which varies from 3.6 to 9.1 for an InN growth temperature from 440–500 °C [31.116]. The difference between the elastic moduli measured with the same technique (Brillouin scattering) for even GaN, which has a much better crystal quality than InN, is further evidence that the quality and nature (bulk single-crystal or epitaxial layer) of the samples is of significant importance. It has been demonstrated by *Moram* et al. [31.100] that the Poisson's ratio of GaN has a strong dependence on the strain state of the films. With the help of HR-XRD while inducing a series of different biaxial strain states in the same wafer, a Poisson's ratio of 0.183 ± 0.003 was estimated for a strain-free GaN film. Stress in GaN was also demonstrated to influence the Young's modulus. For high-quality suspended GaN microstructures fabricated from GaN-on-patterned-silicon techniques where the underlying silicon is removed, the stress in GaN decreases by 47%, which results in a 36% increase of the Young's modulus [31.101], nearly identical to the value deduced from the Brillouin scattering measurement on GaN films [31.120]. Except for InN, group III nitrides can be considered as hard and incompressible material family members with elastic and bulk moduli comparable to the magnitude of diamond. The hardness of a semiconductor relates to the bonding distance or shear modulus, which may explain why the InN is soft. InN has a smaller shear modulus and larger bonding distance (0.214 nm) compared to GaN (0.196 nm) and AlN (0.192 nm). Reports on the temperature dependence of the mechanical properties of group III nitrides are rare. It has been shown that the macroscopic dislocation motion and plastic deformation of GaN and AlN may start at around 1100 °C [31.121, 122]. The yield strength of bulk single-crystal GaN is found to be 100–300 MPa at 900 °C, while that for AlN was deduced to be $\approx$ 300 MPa at 1000 °C.

Many applications of group III nitrides utilize their high thermal conductivity, thus the precise knowledge of the vibrational modes in single crystals should be obtained in order to understand the fundamental thermal properties of nitrides. Infrared reflection and Raman spectroscopy are very useful techniques to derive zone-center and some zone-boundary phonon modes in nitrides. The A_1 and E_1 modes are each split into longitudinal optic (LO) and transverse optic (TO) components, giving a total of six Raman peaks. These A_1 and E_1 branches are both active for Raman and infrared techniques, while the E_2 branches are Raman-active only and B_1 branches are inactive. Table 31.3 gives a list of experimental as well as calculated values for the zone-center optical-phonon wave numbers of AlN, GaN, and InN.

The phonon dispersion spectrum of AlN has a total of twelve branches, including three acoustic and nine optical branches. The E_1 (LO), E_1 (TO), A_1 (TO), and E_2(high) peak positions were found to show a linear relationship with applied pressure up to 7 GPa [31.136]. Because the E_2 (high) peak position is reported to be sensitive to the in-plane strain due to its comparatively large strain coefficients, its stress-free values were obtained by linear fits to Raman frequencies of AlN layers under different biaxial strains and a stress-free value of $657.4 \pm 0.2\,\mathrm{cm}^{-1}$ is reported in AlN [31.146]. In contrast, the E_2 (low) phonon frequency is essentially constant under pressure, which indicates an approximate cancellation of the effects of central and noncentral forces on the pressure dependence of the phonon frequency [31.136]. The influence of impurities on E_2 (high) peak position showed controversial results. The E_2 (high) peak position showed no change for bulk AlN with carbon impurities up to $3 \times 10^{19}\,\mathrm{cm}^{-3}$. However, the principle Raman peaks are influenced by oxygen-related complexes [31.92], and the widths of principal Raman modes increase with increasing oxygen content, but at high oxygen concentrations ($\approx$ 5 at.%) the lowest values of these widths were obtained [31.103]. Through temperature-dependent Raman measurement, it was found out that both the lifetime of E_2 (high) and A_1 (LO) symmetry phonons of AlN are limited by two-phonon decay, where the E_2 (high) decays symmetrically while the A_1 (LO) asymmetrically decays into A_1 (TO) and LA phonons [31.125, 127].

The phonon frequencies in GaN also spread out due to different strain and defects states in different materials. Besides, the high phonon frequencies due to the strong bond in GaN and the light N atoms limit the range of observable impurity-related local vibrational modes to even lighter elements at higher frequencies. The E_1-TO and the E_2-(high) phonon modes in GaN possess comparatively large strain coefficients, thus their values are especially sensitive to the strain in the crystal. Therefore, care must be taken when measuring temperature-dependent phonon frequencies as different substrates have different thermal expansion coefficients and hence induce different strains at different temperatures [31.147]. Heretofore, few reports have appeared for the infrared and Raman modes of impurities, dopants, and hydrogen complexes [31.148, 149], and most of the reports focused on the hydrostatic pressure dependence of the zone-center phonon modes in bulk GaN [31.134, 136, 138, 150], and mode pressure coefficients up to a hydrostatic pressure of 6 GPa have been determined by Raman measurements [31.136].

Table 31.3 Optical phonon frequencies of wurtzite AlN, GaN, and InN at the center of the Brillouin zone in the units of cm^{-1}

Symmetry	AlN (cm^{-1})	GaN (cm^{-1})	InN (cm^{-1})
A_1-TO	613.64[b], 607.3[c], 609[e], 610[f], 609[g], 612[h], 601[i]	531.0[j], 531.4[l], 533.54[m], 531.2[n], 531.7[o], 540[p]	440[t], 446[u], 440 [31.123], 480[w], 445[x], 440[x]
E_1-TO	671.41[b], 666.5[c], 668[e], 669.6[f], 668[g] 679[h], 650[i]	558.0[j], 558.4[k], 559.99[m], 558.2[o], 568[p], 558.4[q]	477.9[v], 476[w], 472[x], 472[x]
A_1-LO	883.6[a], 891.80[b], 884.5[c], 891[d], 895[e], 888[f]	736.5[o], 748[p], 735 [31.124], 733[y], 737[r]	585.4[s], 592[t], 590[u], 590 [31.123], 580[w], 588[x]
E_1-LO	919.09[b], 911[e], 912.6[f], 911[g]	739.9[n], 742[o], 757[p], 743 [31.124], 740[y], 745[r]	570[w]
E_2-(low)/E_2^1	247.8[a], 249.57[b], 249[c], 246[e], 248[f], 246[g], 247[h], 228[i]	144.1[o], 142[p], 144 [31.124], 144[y], 146[i]	89[u], 88 [31.123], 87[w], 104[x]
E_2-(high)/E_2^2	653.6[a], 658.51[b], 653.6[c], 659[c], 655[e], 656.6[f], 655[g], 672[h], 638[i]	566.6[j], 567.6[k], 567.5[l], 568.28[m], 567.0[o], 576[p], 566.9[q]	490.1[s], 491[t], 491[u], 491.1[v], 490 [31.123], 488[w], 483[x]
B_1-(low)	636[h], 534[i]	337[p], 526[h], 335[i]	192[u], 200[w], 270[x]
B_1-(high)	645[h], 703[i]	713[p], 584[h], 697[i]	540[w], 530[x]

[a] Seeded grown of AlN boules on PVT-grown *c*-plane AlN [31.52]; [b] Bulk wurtzite AlN crystals grown by PVT method [31.125]; [c] 0.8 μm-thick AlN layer under a biaxial tensile stress of 0.6 GPa grown on Si(111) by MBE [31.126]; [d] Freestanding bulk AlN grown by sublimation sandwich technique on SiC seed [31.127]; [e] AlN bulk crystal grown by PVT technique [31.128]; [f] Self-nucleated AlN single crystal with facets [31.129]; [g] Bulk AlN grown by PVT technique [31.130]; [h] Calculated using first-principle total energy [31.131]; [i] Calculated using pseudopotential LDA [31.132]; [j] Nonpolar $(1\bar{1}00)$ bulk GaN grown by ammonothermal method [31.58]; [k] *c*-Plan bulk GaN grown by HVPE [31.57]; [l] *m*-Plane GaN substrate grown by HVPE [31.133]; [m] Nonpolar *a*-plane GaN grown on *r*-plane sapphire substrate [31.134]; [n] Strain-free frequencies in a high-quality bulk GaN [31.135]; [o] 50 μm thick hexagonal crystal of GaN grown on 6H-SiC by HVPE [31.136]; [p] Ab initio calculation using a pseudopotential-plane-wave method [31.137]; [q] Bulk-like GaN grown by HVPE and with removed substrate by laser liftoff [31.138]; [r] Raman study on high-quality freestanding GaN templates grown by HVPE [31.139]; [s] Strain-free values obtained by Raman measurements on a freestanding InN film grown by MBE [31.140]; [t] Raman measurements on wurtzite InN film deposited on sapphire substrate by MOVPE [31.141]; [u] Raman measurements on hexagonal InN thin films grown by MOVPE [31.142]; [v] Strain-free value obtained by Rama measurement on InN films grown on sapphire by MBE [31.143]; [w] Raman study on InN grown on sapphire and calculation based on the pairwise interatomic potentials and rigid-ion Coulomb interaction [31.144]; [x] Raman study on polycrystalline and faceted platelets of InN and calculation using FP-LMTO LDA [31.145]; [y] Brillouin spectroscopy on GaN substrate grown by LEO [31.60].

The E_2 (low) phonon mode in GaN exhibits a weak softening, which is qualitatively similar to that of the zone-boundary transverse acoustic modes of zincblende III–V semiconductors. An increase of the LO-TO splitting has been observed with hydrostatic pressure. Splitting of the GaN E_1 (TO), E_1 (LO), and E_2 phonons under anisotropic strain in the basal plane were also investigated, and their phonon deformation potentials were determined [31.135]. Phonon modes on virtually strain-free freestanding bulk GaN have been reported as well [31.138].

Strain-free Raman frequencies of the E_2 (high) and A_1 (LO) modes of hexagonal InN have been determined to be 490.1 ± 0.2 and $585.4 \pm 0.4\,cm^{-1}$ by Raman measurements on freestanding InN film grown by MBE [31.110]. The strain-free Raman frequencies were further verified by measuring biaxial strain-dependent Raman frequencies, where the strain-free values were extrapolated by a linear fit of the curve. The slope of the linear fit gave the Raman linear biaxial stress coefficients for the E_2 (high) and A_1 (LO) modes of InN to be 9.0 ± 0.8 and $8.4 \pm 0.8\,cm^{-1}\,GPa^{-1}$ respectively. Another set of pressure-dependent Raman frequencies demonstrated the phonon modes to be 440, 491, and $592\,cm^{-1}$ for A_1 (TO), E_2 (high), and A_1 (LO) respectively. Smaller stress coefficients of 5.81, 5.56, and $5.96\,cm^{-1}\,GPa^{-1}$ were also demonstrated for the above three phonon modes [31.151]. The wurtzite to rocksalt phase transition was also evidenced to be at a pressure of 13.5 ± 0.5 GPa as analyzed by means of Raman spectroscopy measurements conducted under high pressure (up to 50 GPa) [31.141]. Temperature-dependent Raman analysis of A_1 (LO) and E_2 phonon lifetimes from 80–700 K indicated that among the various possible decay channels, the A_1 (LO) phonon decays asymmetrically into a high energy and a low energy phonon, whereas the E_2 phonon predominantly decays into three phonons [31.152]. Possible decay channels of A_1 (LO) phonon may involve combinations of transverse optical and acoustic phonons.

31.4 Thermal Properties of Nitrides

31.4.1 Thermal Expansion Coefficients

The lattice parameters of semiconductors are temperature dependent, and the change of the lattice parameters are defined by thermal expansion coefficients (TEC), which are defined as $\Delta a/a$ or α_a and $\Delta c/c$ or α_c, for the *in* and *out of plane* orientations respectively. The importance of such parameters is owed to the expansion-caused strain during material growth, which is particularly important for group III nitrides as many applications of group III nitrides rely on successful growth of heterostructures, for example AlGaN/GaN HFETs and InGaN LEDs. What matters most is the lack of easily available substrates for group III nitride growth, thus the difference of TEC will cause strain and even cracking during the growth or cooling-down process for tensile strain, which is the case when grown for example on Si and SiC substrates. TEC is dependent on the stoichiometry and defects as well as the free-carrier concentration. A large scatter in the published data exists for TEC, partially due to the varying foreign substrates used for the growth.

The TEC was reported to be 4.03×10^{-6} in the temperature range of 298–473 K by assuming linear expansion in hot pressed aluminum nitride [31.155]. Ceramic AlN has been demonstrated to show a negative TEC at a temperature range of from 0 K to around 100 K [31.156]. Another TEC investigation on AlN powder showed that experimental data indicate no minimum for the a parameter and a possible shallow minimum for the c parameter, however, the magnitude of experimental errors does not allow confirmation of the existence/absence of negative thermal expansion without additional experimental efforts [31.157].

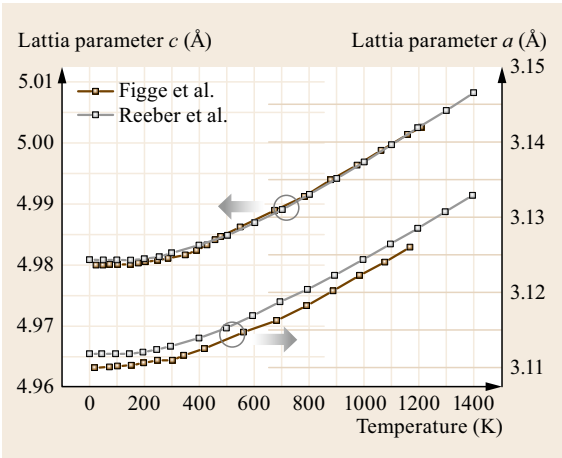


Fig. 31.1 Temperature-dependent lattice parameters of AlN in the c-direction and the a-direction. (After [31.153, 154])

Using x-ray techniques across a broad temperature range (77–1269 K), it has been noted that AlN demonstrated an isotropic thermal expansion with a room-temperature value of $2.56\times10^{-6}\,\mathrm{K}^{-1}$ [31.158]. Mean values of 4.20×10^{-6} and $5.30\times10^{-6}\,\mathrm{K}^{-1}$ were reported for $\Delta a/a$ and $\Delta c/c$, respectively, in the temperature range of 293–1073 K [31.159].

The thermal expansion coefficient of AlN bulk crystals were also investigated in the temperature range of 20–1300 K with the measured lattice parameters from both (002) and (006) reflections [31.153], and the lattice parameters were plotted at different temperature as shown in Fig. 31.1. In both lattice directions almost no thermal expansion is observed at low temperatures and an almost linear expansion at temperatures above 750 K, which is consistent with models based on the intrinsic phonon energy of a solid system. By fitting the experimental data to models based on Debye- and Einstein-like phonon dispersion, the lattice parameter c with its temperature dependence can be described as

$$c(T) = c(0\,\mathrm{K})\left[1+\alpha_\infty\,\Theta f\left(\frac{\Theta}{T}\right)\right], \tag{31.3}$$

where α_∞ is the lattice expansion coefficient in the high temperature limit, Θ is a characteristic temperature, and $f(x)$ is either the Debye or Einstein function defined as below

$$f_\mathrm{D}(x) = 3\int_0^1 \frac{t^3}{\mathrm{e}^{tx}-1}\mathrm{d}t$$

$$f_\mathrm{E}(x) = \frac{1}{\mathrm{e}^{x}-1}\,. \tag{31.4}$$

The thermal expansion coefficients were reported to be 3.38×10^{-6} and $2.68\times10^{-6}\,\mathrm{K}^{-1}$ for $\Delta a/a$ and $\Delta c/c$ at room temperature, respectively.

Thermal expansion of single-crystal wurtzitic GaN has been investigated in the temperature range of 300–900 K [31.160], with a mean thermal expansion coefficient of $5.59\times10^{-6}\,\mathrm{K}^{-1}$ for $\Delta a/a$.

A superlinear dependence on temperature was also shown for the c parameter with mean thermal coefficients of 3.17×10^{-6} and $7.75\times10^{-6}\,\mathrm{K}^{-1}$, over the temperature ranges of 300–700 and 700–900 K respectively. Other values were also reported for α_a of 3.1×10^{-6} and $6.2\times10^{-6}\,\mathrm{K}^{-1}$ for the temperature ranges of 300–350 and 700–750 K respec-

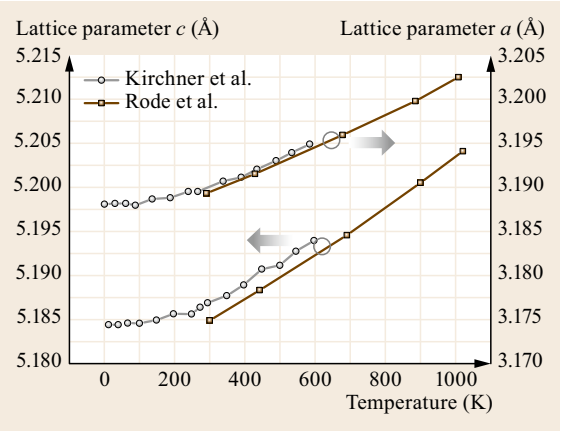


Fig. 31.2 Temperature-dependent lattice parameters of GaN in *c*-direction and *a*-directions. (After [31.162, 163])

tively [31.161]. The α_c values of 2.8×10^{-6} and $6.1\times10^{-6}\,K^{-1}$ were also reported in the same report. Thermal expansion coefficients should be obtained from freestanding bulk GaN substrates with minimal dislocations, defects and impurities, so that the influence by the substrate and defect states can be minimized. Thermal expansion coefficients for freestanding GaN bulk substrates with dislocation densities below $10^5\,cm^{-2}$ and room-temperature free-electron concentrations below $2\times10^{19}\,cm^{-3}$ were investigated in the temperature ranges of 0–600 K [31.162] and 300–1000 K [31.163] respectively. The temperature dependence of the lattice parameters are shown in Fig. 31.2. The thermal expansion coefficients at the very high temperature limits were reported to be $\alpha_c = 5.7\times10^{-6}\,K^{-1}$ and $\alpha_a = 6.2\times10^{-6}\,K^{-1}$, comparable to those proposed by *Reeber* and *Wang* [31.164]. The experimental results were also fitted to both Debye and Einstein models, and the results were demonstrated and compared with AlN as tabulated in Table 31.4. By comparing the lattice parameters of AlN and GaN at different temperatures, it was found that the maximum thermal mismatch between these two materials occurs at around 700 K [31.153]. Because the typical growth temperatures for AlN/GaN systems are much higher, the heterostructures suffer a larger strain during cool-down than during growth.

There is limited information available regarding the thermal expansion of InN. The first published experimental data in the temperature range of 190–560 K demonstrated that the TECs increase with temperature: α_a from 3.4×10^{-6} to $5.7\times10^{-6}\,K^{-1}$, α_c from 2.7×10^{-6} to $3.7\times10^{-6}\,K^{-1}$ [31.165]. More recent experimental data in the 100–673 K range based on diffraction data demonstrated average TECs of $\alpha_a = 3.6(2)\times10^{-6}\,K^{-1}$ and α_a from $\alpha_c = 2.6(3)\times10^{-6}\,K^{-1}$ [31.65]. However, due to the scatter of experimental data points obtained with the use of conventional equipment, there is lack of detectable TEC variation in the investigated temperature range. Lattice parameters and TECs were also calculated in the temperature range of 50–800 K based on a semi-empirical model [31.166]. More recently, results of Rietveld refinement for InN data collected in the temperature range of 105–295 K were published [31.167], wherein acicular microcrystals of InN were prepared by reaction liquid indium with nitrogen plasma and lattice parameters were measured by XRD. The TECs derived from linearly approximated lattice-parameter dependencies are $\alpha_a = 3.09(14)\times10^{-6}\,K^{-1}$ and $\alpha_c = 2.79(16)\times10^{-6}\,K^{-1}$. The temperature-dependent lattice parameters of InN are presented in Fig. 31.3a,b.

31.4.2 Thermal Conductivity

One key issue that enables group III nitrides to be useful for high-power/high-temperature electronic and optoelectronic devices is the high thermal conductivity (k), which is determined by the contributions from the vibrational, rotational, and electronic degrees of freedom, and as such is related to the mechanical properties of the material. The electronic thermal conductivity contribution is negligible for carrier concentrations $\leq 10^{19}\,cm^{-3}$. Heat transport is predominantly determined by phonon-phonon Umklapp scattering as well as phonon scattering by point and extended defects, such as vacancies, impurities, isotope fluctuations, grain boundaries, and dislocations. For the case when the effect of defects is trivial, phonon-phonon scattering, which is ideally proportional to T^{-1} above the Debye temperature, is the limiting process. However, due to the defective nature of group III nitride

Table 31.4 Thermal expansion parameters determined by XRD for AlN and GaN. (After [31.153] and [31.163] respectively)

Direction	Material	Debye model		Einstein model	
		α_∞ ($10^{-6}\,K^{-1}$)	Θ (K)	α_∞ ($10^{-6}\,K^{-1}$)	Θ (K)
out of plane	AlN	5.8 ± 0.1	1317 ± 25	5.6 ± 0.1	937 ± 25
	GaN	5.73 ± 0.44	898 ± 24	5.71 ± 0.43	662 ± 18
in-plane	AlN	7.1 ± 0.3	1455 ± 25	6.9 ± 0.3	1025 ± 25
	GaN	6.24 ± 0.41	868 ± 20	6.21 ± 0.35	636 ± 13

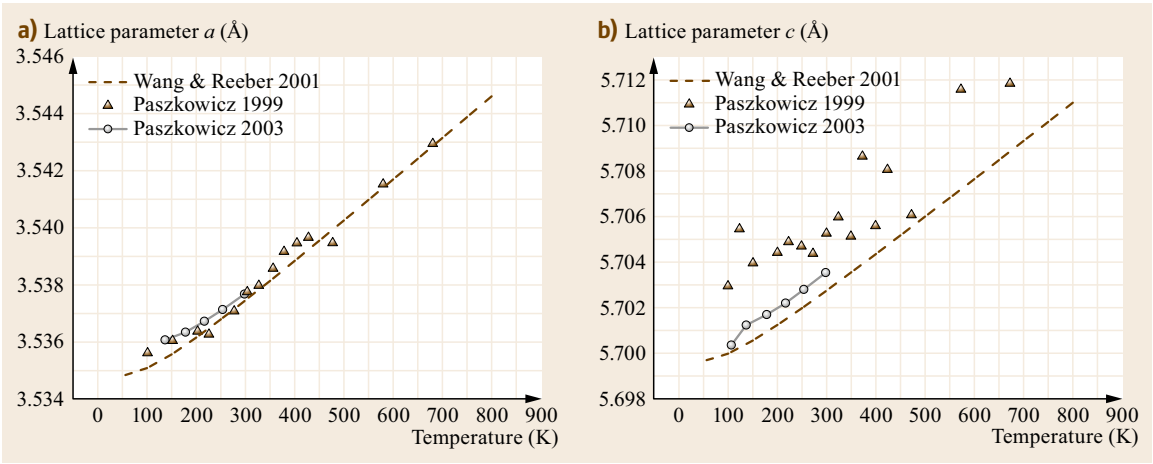


Fig. 31.3a,b Temperature dependence of lattice parameters of InN. (After [31.65, 166, 167])

materials, point defects play a significant role for single crystals during heat transport.

The thermal conductivity of AlN at room temperature was theoretically estimated to be $3.19\,\mathrm{W\,cm^{-1}\,K^{-1}}$ for pure AlN single crystals [31.168], and experimental values of 2.5 [31.169] and $2.85\,\mathrm{W\,cm^{-1}\,K^{-1}}$ [31.158] were reported at 300 K for AlN single crystals obtained by the sublimation technique. However, a more recent calculation predicted thermal conductivity of $5.4\,\mathrm{W\,cm^{-1}\,K^{-1}}$ for a pure AlN single crystal, which is much larger than commonly reported experimental data [31.170]. One type of impurity that impacts the thermal conductivity of AlN is identified to be oxygen, and the thermal conductivities of AlN with various oxygen concentrations were demonstrated in Fig. 31.4. Although it was found that higher oxygen concentration reduced the thermal conductivity, it was believed that the reduction of the thermal conductivity was due to the oxygen-induced Al vacancies [31.171], which is supported by another report showing that AlN crystals with comparable oxygen vacancies ($\approx 1\times10^{19}\,\mathrm{cm^{-3}}$) exhibited decreasing thermal conductivity with increasing Al vacancies [31.172]. As shown in Fig. 31.4, the reduction of thermal conductivity due to oxygen vacancies only occurs below room temperature. For high temperatures, the role of defects in affecting the thermal conductivity vanishes and the role of anharmonicity becomes dominant [31.173]. Recent AlN single-crystals have demonstrated higher room-temperature thermal conductivity values in the range of $3.0–3.3\,\mathrm{W\,cm^{-1}\,K^{-1}}$ for freestanding [31.174] and 300–800 μm-thick AlN samples grown originally on Si (111) substrate by HVPE with a $10^8\,\mathrm{cm^{-2}}$ dislocation density [31.175]. The thermal conductivity of polycrystalline or ceramic AlN has a much smaller value than that of single-crystalline AlN due to the larger amount of defects and grain boundaries [31.176], and the reported maximum value is $2.2\,\mathrm{W\,cm^{-1}\,K^{-1}}$ [31.177].

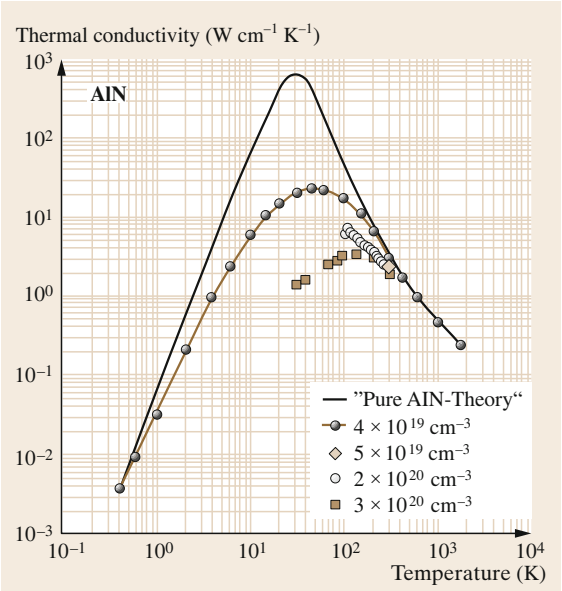


Fig. 31.4 Thermal conductivity of single-crystal AlN. The solid line represents the calculated data for pure AlN, while others represent AlN with various concentrations of oxygen impurities. (After [31.171])

Although a theoretical thermal conductivity of $k \approx 4.10\,\mathrm{W\,cm^{-1}\,K^{-1}}$ at room temperature was predicted for pure GaN [31.170], experimental results yield much smaller values. The thermal conductivity of GaN layers grown on sapphire substrate by HVPE were measured as a function of temperature (25–60 K) using a heat-flow method, and a room-temperature thermal

Table 31.5 Thermal conductivity of high-quality GaN

Reference	Growth condition	Dislocation density (cm^{-2})	Impurity concentration (cm^{-3})	Carrier density (cm^{-3})	$k_{(300\,K)}$ ($W\,cm^{-1}\,K^{-1}$)	$k_{(max)}$ ($W\,cm^{-1}\,K^{-1}$)	Power n at 300 K
[31.178]	HVPE GaN boule	6×10^5	$\approx 6\times10^{16}$	1×10^{16}	2.94 ± 0.44		
[31.171]	HVPE		$\approx 2\times10^{16}$		2.27	20	-1.22
[31.179]	High-pressure-grown GaN		$\approx 10^{20}$	5×10^{19}	2.3	16	≈ -1.43
[31.180]	HVPE	$\approx 5\times10^6$	$\approx 2\times10^{17}$		2.53		
[31.181]	HVPE	$\approx 6.2\times10^6$	$\approx 3\times10^{16}$		2.25		≈ -1

conductivity of $\approx 1.3\,W\,cm^{-1}\,K^{-1}$ was recorded in the early days of GaN technology [31.182]. More recently, the room-temperature thermal conductivity of GaN was calculated to be $2.27\,W\,cm^{-1}\,K^{-1}$ assuming no isotope scattering [31.171]. This prediction is close to the generally reported value for high-quality bulk GaN as shown in Table 31.5. However, as also demonstrated in Table 31.5, the highest thermal conductivity is reported as $2.94\pm0.44\,W\,cm^{-1}\,K^{-1}$ for HVPE-grown GaN boule with very low dislocation density ($6\times10^5\,cm^{-2}$) and impurity concentrations $\approx 6\times10^{16}\,cm^{-2}$) [31.178], which in turn demands for more accurate predictions of thermal conductivity for high-quality GaN. From theoretical calculations, the highest possible thermal conductivity of bulk GaN can only be realized when point impurities such as oxygen and silicon are in small concentrations ($\approx 10^{16}\,cm^{-3}$ or less) and other defects are either absent or present in very small concentrations [31.183]. For GaN with relatively high quality, the measured thermal conductivity of GaN in the temperature range of 80–300 K has a $T^{-1.22}$ temperature dependence as demonstrated in Fig. 31.5, whose slope is typical of pure adamantine crystals below the Debye temperature, indicating acoustic phonon transport. Thus, the phonon-phonon scattering is a combination of acoustic–acoustic and acoustic–optic interactions, which means that thermal conductivity is mainly limited by the intrinsic phonon-phonon scattering instead of phonon-impurity scattering [31.171]. When the dislocation density (ρ_{TD}) exceeds $10^7\,cm^{-2}$, phonon-dislocation scattering will become nonnegligible compared to phonon-phonon scattering at room temperature [31.184], and an empirical relation indicated that thermal conductivity would be inversely proportional to $\rho_{TD}^{0.12}$ as ρ_{TD} goes to infinity. Another limiting factor of the thermal conductivity is the extent of dopants, and thermal conductivity decreased linearly with log n, where n is the electron concentration, the variation being about a factor of two decrease in k for every decade increase in n [31.185, 186].

Thermal conductivity measurements on InN are rare, partially due to the difficulty of growing high-quality InN films and single crystals. The first report of thermal conductivity was based on ceramic InN, which demonstrated a room-temperature value of $k = 0.45\,W\,cm^{-1}\,K^{-1}$ measured using the laser-flash method [31.187]. A more recent investigation is on single-crystalline InN films grown by MBE on c-sapphire substrate with a GaN buffer layer [31.188]. With background electron concentrations between 8×10^{17} and $3\times10^{18}\,cm^{-3}$ and mobilities between 1500 and $1045\,cm^2/Vs$, a room-temperature thermal conductivity of $1.2\,W\,cm^{-1}\,K^{-1}$ was obtained, which is closer to the theoretical thermal conductivity of $1.8\,W\,cm^{-1}\,K^{-1}$ estimated from $k_{theoretical} = C_\upsilon l\upsilon/3$, where c_υ is the molar specific heat, υ is the Debye-averaged acoustic phonon velocity, and $l = a/(\alpha_\upsilon\gamma T)$ is the phonon mean free path [31.188]. The effect of He^{2+} implantation on the thermal conductivity of InN

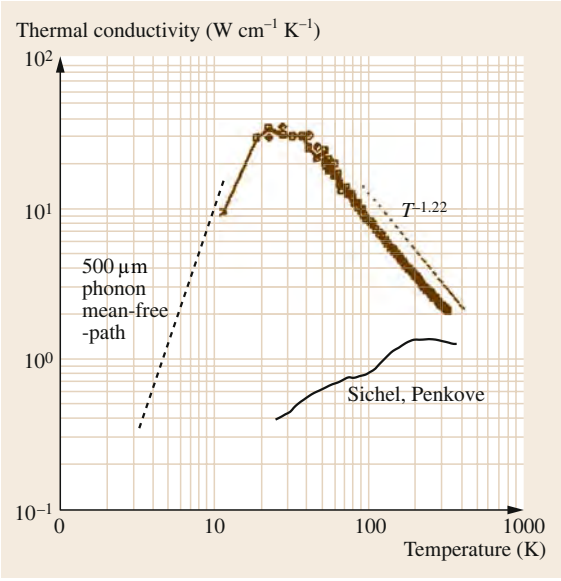


Fig. 31.5 Thermal conductivity of a 200 µm-thick free-standing GaN sample as a function of temperature. The *dashed line* indicates a calculation using the boundary scattering limit for a phonon mean free path of 500 µm. Also shown is the $T^{-1.22}$ dependence in the range of 80–300 K, and earlier results from [31.182] measured using a 400 µm HVPE sample. (After [31.171])

was also investigated, where the thermal conductivity reduced to around 0.1 W cm^{-1} K^{-1} when the implantation dose increased to 10^{16} cm^{-2}.

31.4.3 Specific Heat

The specific heat of a semiconductor can have various contributors, such as those by lattice vibrations, free carriers, point defects, and extended defects. For an ideal case where the semi-insulating crystals have good quality, the specific heat is only determined by the lattice vibrations. Due to the defective nature of group III nitrides, the specific heat is affected by contributions from free carriers and defects, especially at low temperatures. The temperature dependence of the specific heat can be expressed by the Debye expression as

$$C_\mathrm{p} = 18R\left(\frac{T}{\Theta_\mathrm{D}}\right)^3 \int_0^{x_\mathrm{D}} \frac{x^4 \mathrm{e}^x}{(\mathrm{e}^x - 1)^2}\mathrm{d}x\,, \tag{31.5}$$

where $x_\mathrm{D} \equiv \Theta_\mathrm{D}/T$, and $R = 8.3144\,\mathrm{J/(molK)}$ is the molar gas constant. The coefficient in front of the term R has been multiplied by 2 to take into account the two constituents making up the group III nitrides. The Debye temperature Θ_D can be obtained by fitting the measured temperature-dependent capacity to the Debye expression.

The constant pressure specific heat C_p of AlN in the temperature range of 298–1800 K was approximated using the following expression [31.189]

$$C_\mathrm{p} = 45.94 + 3.347\times10^{-3}T - 14.98\times10^{-5}T^2\,\mathrm{J/(mol\,K)}\,. \tag{31.6}$$

By using the specific heat of $C_\mathrm{p} = 51.5\,\mathrm{J/mol\,K}$ at $T = 1800\,\mathrm{K}$ and the estimated value of $C_\mathrm{p} = 58.6\,\mathrm{(mol\,K)}$ at $T = 2700\,\mathrm{K}$, the specific heat at higher temperature range of 1800–2700 K, can be approximated as [31.190]

$$C_\mathrm{p} = 37.34 + 7.86\times10^{-3}T\,\mathrm{J/(mol\,K)}\,. \tag{31.7}$$

The above equations show the general and simplified trends for the specific heat of AlN. However, the exact values vary from sample to sample because the specific heat is a function of free electrons (very effective at low temperatures), impurities, defects (including point defects), and lattice vibrations.

The experimental temperature-dependent specific heat results are shown in Fig. 31.6. Also shown are the calculated specific heat results using the Debye equation for Debye temperature values of 800–1100 K with 50 K increments. The best fit between the experimental data and the Debye specific-heat expression reveals

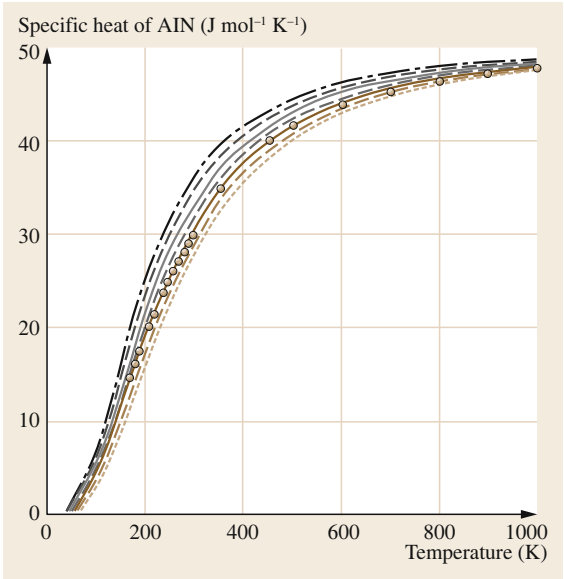


Fig. 31.6 Temperature-dependent molar specific heat of AlN at constant pressure C_p. *Brown dots* represent experimental data. The *lines* are for calculated data based on the Debye model for Debye temperature Θ_D in the range of 800–1100 K with 50 K increments. The best fit is obtained for a Debye temperature of 1000 K, which is close to the value of 950 K reported. (After [31.191])

a Debye temperature of 1000 K, which is close to the value of 950 K reported in reference [31.191].

The temperature-dependent specific heat for powdered GaN was obtained in a temperature range of 200–1400 K [31.192], which is in good agreement with the heat content data reported by *Itagaki* and *Yamaguchi* [31.193]. The combined data can be approximated by the following equation

$$C_\mathrm{p} = (49.552 \pm 2.279) + (5.440 \pm 2.936)\times10^{-3}\,T - (2.190 \pm 0.288)\times10^{6}\,T^{-2} + (2.460 \pm 0.459)\times10^{8}\,T^{-3}\,\mathrm{J/(mol\,K)}\,. \tag{31.8}$$

The specific heat for single-crystal GaN was also investigated in the temperature range of 20–1400 K by *Kremer* et al. [31.195] as demonstrated in Fig. 31.7, where the data below 300 K are in very good agreement with a later report by *Danilchenko* et al. [31.194]. Measured data in the range of 300–1400 K from *Kremer* can be best described with a Debye temperature of 863 K, which is consistent with the value calculated by *Nipko* et al. [31.196] from inelastic neutron scattering data. The dependence of the heat capacities of GaN, as a canonical binary material, on each of the Ga and N masses was also investigated, and it was shown that Ga

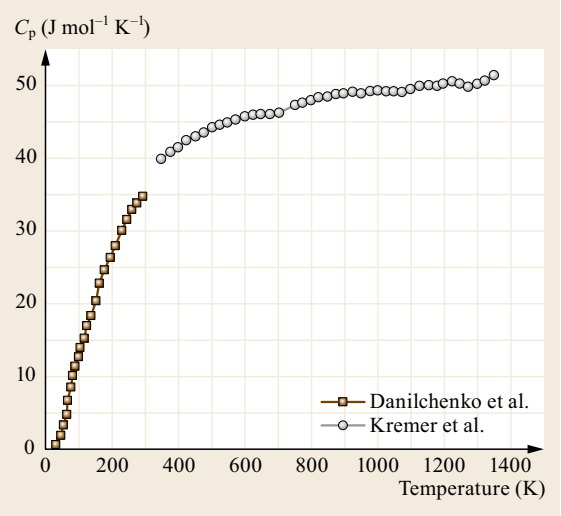


Fig. 31.7 Dependence of measured heat capacity on temperature measured in the range of 5–300 K [31.194] for single-crystal GaN, which overlaps with the data obtained from [31.195] in the same temperature range. The open circles represent measurement results by *Kremer* et al. in the temperature range of 300 to 1400 K

mass affects mainly the acoustic while N affects the optic phonons [31.195]. The temperature dependence of specific heat on bulk hexagonal GaN was also investigated [31.194], and it was shown that, for the studied specimens, the electron contribution to the C_p is negligible due to the fact that lattice C_p exhibits T^3 behavior and thus a C_p/T versus T^2 plot results in a straight line through the origin. From the measurement results in the temperature range of 20–300 K, a Debye temperature of 365 K was obtained.

A Debye temperature of 660 K was obtained for single-crystal InN based on fitting the measurement results in the temperature range of 150–300 K to the Debye expression [31.187]. However, due to the narrow temperature range of these measurements, it is difficult to compare these results to the Debye curve. The heat capacity and heat content of powdered InN were also measured by Calvet calorimetry (305–390 K) and by drop calorimetry (427–774 K), and the temperature dependence of heat capacity can be fitted as the following equation [31.197]

$$C_p = 43.886 + 8.194\times10^{-3}T - 1.007\times10^{6}T^{-2} + 8.353\times10^{7}T^{-3}\,\mathrm{J/(mol\,K)}\,. \tag{31.9}$$

Similar measurements were also conducted for powdered InN with 8.4% indium impurities in the temperature range of 314–978 K [31.198], and the temperature dependence of the heat capacity was presented in the following form

$$C_p = 41.400 + 0.499\times10^{-3}\,T - 1.355\times10^{5}\,T^{-2} - 2.617\times10^{7}\,T^{-3}\,\mathrm{J/(mol\,K)}\,. \tag{31.10}$$

For both cases, liquid droplets of pure indium were observed on the surface of InN after experiments at high temperature, thus a recalculation had to be conducted to determine only the heat capacity of indium nitride. Also due to the difficulty of growing high-quality pure InN crystals, the heat capacity of InN samples may have significant contributions from nonvibrational modes, which limits the credibility of the data presented above.

31.5 Electrical Properties of Nitrides

GaN and related nitrides, being direct and large bandgap materials, lend themselves to a variety of electronic and optoelectronic applications. Advantages associated with a large bandgap include relatively high breakdown voltages, the ability to sustain large electric fields, low noise generation, and high temperature and high power operation. Reasonable low-field mobility, large satellite energy separation, and high phonon frequency are among the other attributes. A high thermal conductivity, large electrical breakdown fields, and resistance to hostile environments also support the group III nitrides as true materials of choice for device applications. The electron transport in semiconductors, including nitrides, can be considered at low and high electric field conditions. At sufficiently low electric fields, the energy gained by the electrons from the applied electric field is small compared to the thermal energy of electrons, and therefore, the energy distribution of electrons is unaffected by such a low electric field. Since the scattering rates determining the electron mobility depend on the electron distribution function, electron mobility remains independent of the applied electric field, and Ohm's law is obeyed. When the electric field is increased to a point where the energy gained by electrons from the external field is no longer negligible compared to the thermal energy of the electron, the electron distribution function changes significantly from its equilibrium value. These electrons become hot electrons characterized by an electron temperature larger than the lattice temperature. Furthermore, as the

dimensions of the device are shrunk to the submicron regime, transient transport occurs where there is minimal or no energy loss to the lattice. The transient transport is characterized by the onset of ballistic or velocity overshoot phenomena, not that they have been demonstrated in the nitride semiconductor family. Since the transient electron drift velocity is higher than its steady-state value one can design a device operating at higher frequency, which is yet to be demonstrated for GaN based devices as of the time of this writing.

31.5.1 Low Field Transport

The Hall effect is the most widely used technique to measure the transport properties and assess the quality of epitaxial layers. For semiconductor materials, it yields the carrier concentration, its type, and carrier mobility. More specifically, temperature-dependent Hall measurements over a wide range (4.2−300 K) give information on impurities, imperfections, uniformity, scattering mechanism, and so on. The Hall coefficient and resistivity are experimentally determined and then related to the electrical parameters through $R_H = r_H/ne$ and $\mu_H = R_H/\rho$, where n is the free-carrier concentration, e is the unit of electrical charge, μ_H is the Hall mobility, and r_H is the Hall scattering factor. The drift mobility, μ, is the average velocity per unit electric field in the limit of zero electric field and is related to the Hall mobility through the Hall scattering factor by $\mu_H = r_H\mu$. The Hall scattering factor depends on the details of the scattering mechanisms in play, which limits the drift velocity. As the carriers travel through a semiconductor, they encounter various scattering mechanisms that govern the carrier mobility in the electronic system. The parameter for characterizing the various scattering mechanisms is the relaxation time τ, which determines the rate of change in electron momentum as it moves about in the semiconductor crystal. Mobility is related to the scattering time by

$$\mu = \frac{q\langle\tau\rangle}{m^*} , \tag{31.11}$$

where m^* is the electron effective mass, q is the electronic charge, and $\langle\tau\rangle$ is the relaxation time averaged over the energy distribution of electrons. The total relaxation time, τ_T, when various scattering mechanisms are operative is given by Matthiessen's rule

$$\frac{1}{\tau_T} = \sum_i \frac{1}{\tau_i} , \tag{31.12}$$

where i represents each scattering process. The major scattering mechanism that generally governs the electron transport in III-V semiconductors is also valid for group III nitrides. They are briefly listed as follows:

1. Ionized impurity scattering is due to deflection of free carriers by the long-range Coulomb potential of the charged centers caused by defects or dopants/impurities. This can be thought of as a local perturbation of the band edge, which affects the electron motion.
2. Polar longitudinal-optical (LO) phonon scattering is caused by the interaction of a moving charge with an electric field induced by electric polarization associated with lattice vibration due to the ionic nature of the bonds in polar semiconductors such as nitrides.
3. Acoustic phonon scattering through deformation potential arising from the energy change of the band edge induced by strain associated with acoustic phonons, where the scattering rate increases with the wave vector of the phonons.
4. Piezoelectric scattering arises from the electric fields that are produced by the strain associated with phonons in a crystal without inversion symmetry, such as wide bandgap nitrides.
5. Because of the high density of dislocations and native defects induced by nitrogen vacancies in GaN, dislocation scattering and scattering through nitrogen vacancies have also been considered as a possible scattering mechanism. Dislocation scattering is due to the fact that acceptor centers are introduced along the dislocation lines, which capture electrons from the conduction band in an n-type semiconductor. The dislocation lines become negatively charged and a space charge region is formed around them, which scatters electrons traveling across the dislocations, thus reducing the mobility.

Gallium Nitride

Electron mobility in GaN is one of the most important parameters associated with the material with great impact on devices. It has been the subject of intensive studies in recent years from both experimental and theoretical points of view. Experimental investigation of the temperature-dependent carrier mobility and concentration can be used to determine the fundamental material parameters and understand the carrier scattering mechanisms along with an accurate comparison with theory [31.200, 201]. Despite the many attributes of GaN, the lack of high-quality material, until very recently, prevented detailed investigations of carrier transport. The earlier transport investigations had to cope with poor crystal quality and low carrier mobility, well below predictions [31.202, 203]. Early MBE layers exhibited mobilities as high as 580 $cm^2/(Vs)$ on SiC

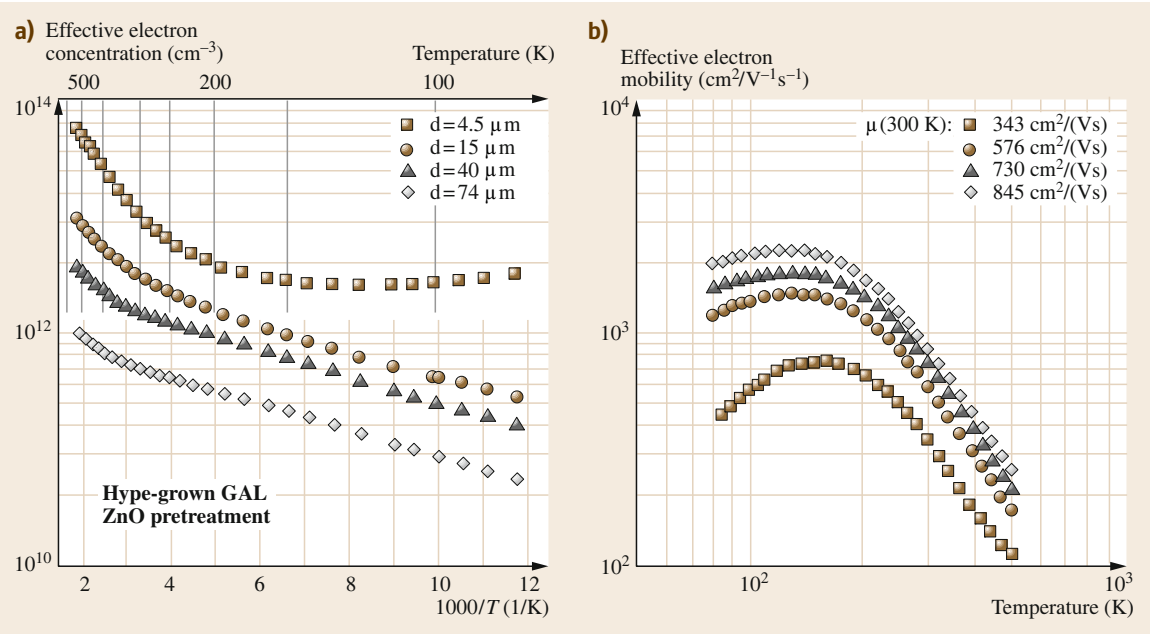


Fig. 31.8a,b Effective electron concentration versus reciprocal temperature in unintentionally doped n-type GaN films grown by HVPE on sapphire substrates **(a)** and effective electron mobility versus temperature **(b)** as determined from Hall-effect measurements under the assumption of uniform film properties. (After [31.199])

substrates, which at that time were not as commonly used as in recent times [31.204]. Typically however, the MBE-grown films produce much lower mobility values of 100–300 cm^2/(Vs) [31.205]. Different models were used to explain the observed low electron mobilities in GaN, especially at low temperatures. The scattering of electrons at charged dislocation lines [31.205–209] and scattering through elevated levels of point defects [31.210, 211], such as nitrogen vacancies [31.212, 213] were considered as possible mechanisms responsible for these observations. These scattering mechanisms were investigated by studying the temperature dependence of the carrier concentration and electron mobility. It has been argued that mobility is related to the dislocation density (N_{dis}) and free-carrier concentration (n) via a $\mu_{dis} \propto \sqrt{n}/N_{dis}$ relationship [31.207]. At low carrier concentrations ($< 5\times10^{17}$ cm^{-3}), the mobility decreases due to charged dislocation scattering, while at higher carrier concentrations ionized impurity is the dominant mechanism determining the mobility. The temperature dependence of mobility for samples where dislocations play a dominant role shows that the mobility increases monotonically with temperature following a $T^{2/3}$ dependence. Moreover, electron mobility limited by nitrogen vacancy scattering was taken into account in n-type GaN grown by MOVPE by *Zhu* and *Sawaki*, [31.212] and *Chen* et al. [31.213]. A good fitting was obtained between the calculated and experimental results. The estimated mobility shows $T^{-1/2}$ temperature dependence and it was argued that the measured mobility is dominated by ionized impurity and dislocation scattering at low temperatures, but by polar optical phonon and nitrogen vacancy scattering at high temperatures.

Hall mobility and electron concentration in undoped GaN were investigated as a function of thickness of the buffer layers and epilayers. *Nakamura* [31.215] reported Hall effect data for undoped GaN layers grown by MOVPE on GaN buffer/sapphire substrates. As the thickness of the initial buffer layer increased from 100 Å, the mobility also increased up to a thickness of 200 Å. For larger thicknesses the mobility began to decrease. The value of the mobility was 600 cm^2/(V s) at room temperature for a 200 Å-thick initial buffer layer. The electron concentration was minimal at 520 nm buffer thickness and increased monotonically as buffer thickness increased. The mechanism responsible for this observation is not clearly established. *Götz* et al. [31.199] have studied the effect of the layer thickness on the Hall mobility and electron concentration in unintentionally doped n-type GaN films grown by HVPE on sapphire substrates, which were pretreated with either ZnO or GaCl. They found that the mobility increased and the carrier concentration decreased as the thickness of the epilayers increased as shown in Fig. 31.8.

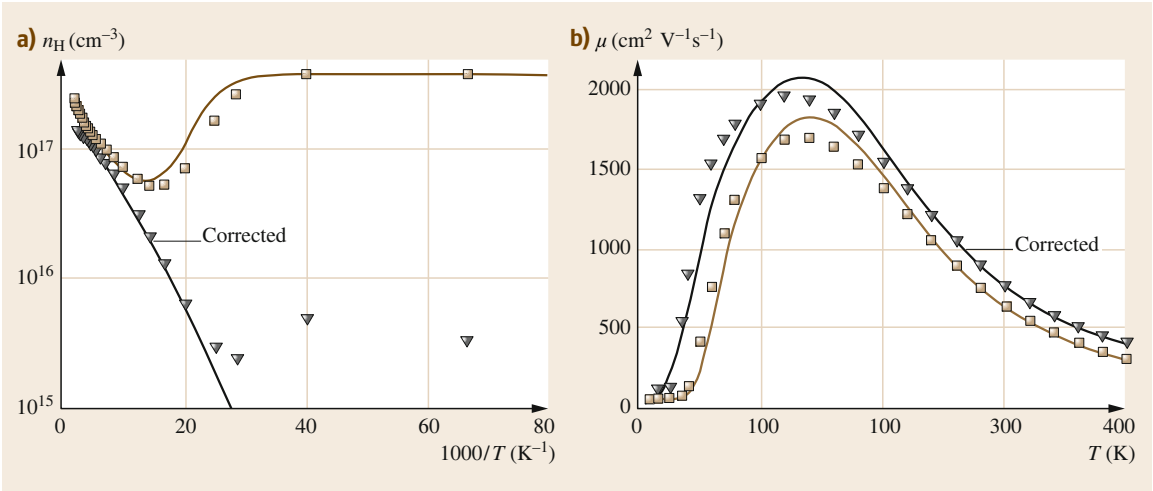


Fig. 31.9a,b GaN on sapphire: **(a)** Uncorrected Hall concentration data (*squares*) and fit (*solid line*), and corrected data (*triangles*) and fit (*dashed line*) versus inverse temperature. **(b)** Uncorrected Hall mobility data (*squares*) and fit (*solid line*), and corrected data (*triangles*) and fit (*dashed line*) versus temperature. (After [31.214])

Götz et al. related this variation to a nonuniform distribution of electrically active defects through the film thickness for both types of film. For GaCl-pretreated sapphire the presence of a highly conductive, 200 nm-thick, near-interface layer was assumed to account for the observed phenomena. For ZnO-pretreated sapphire, the Hall-effect data indicated a continuous reduction of the defect density with increasing film thickness. Since GaN layers are often grown on foreign substrates with very different properties, a degenerate layer forms at the interface caused by extended defects and impurities. Experiments show that, even for thick GaN grown by HVPE, the degenerate interfacial layer has an important contribution to the Hall conductivity.

Look and *Molnar* [31.214] investigated the Hall effect in the temperature range of 10–400 K in HVPE-grown layers on sapphire substrates by assuming a thin, degenerate n-type region at the GaN/sapphire interface. This degenerate interfacial region dominates the electrical properties below 30 K, but also significantly affects those properties even at 400 K, and can cause a second, deeper donor to falsely appear in the analysis. The mobility versus temperature curve is also affected in that the whole curve is shifted downward from the true, bulk curve. A model consisting of two layers was constructed to interpret these observations and the result is shown in Fig. 31.9.

The dependence of Hall mobilities on electron concentrations was also investigated in the electron concentration range of $\approx 10^{16}$–$10^{19}\,\mathrm{cm}^{-3}$ for undoped and intentionally doped (commonly used Si and Ge, which substitute for Ga, are shallow donors with almost identical binding energies) GaN layers grown

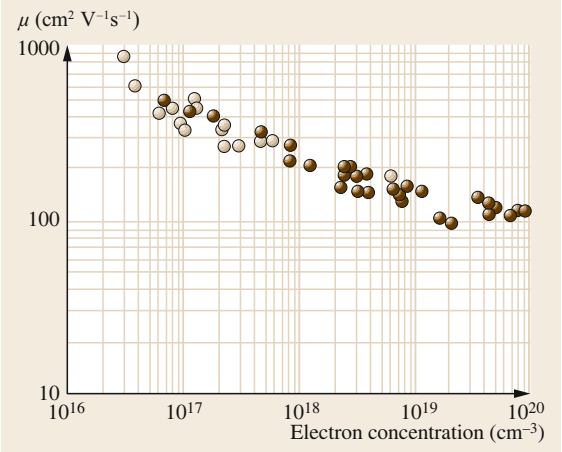


Fig. 31.10 The 300 K Hall mobility versus free-electron concentration for GaN from various groups using both MOVPE and MBE. The *open circles* are from unintentionally doped samples and the *solid circles* are for samples doped with either Si or Ge. (After [31.216])

by different growth techniques, and the results are displayed in Fig. 31.10. The most clearly observed trend is that the mobility shows no signs of leveling, down to the lowest carrier concentration reported, and the mobility values are practically the same irrespective of growth techniques and dopants used, which reflects transport properties inherent to GaN, not to extrinsic effects. Returning to the data of *Nakamura* et al. [31.218] the room-temperature mobility of $600\,\mathrm{cm}^2/(\mathrm{V\,s})$ for an electron concentration of $\approx 3\times10^{16}\,\mathrm{cm}^{-3}$ decreased slowly with increas-

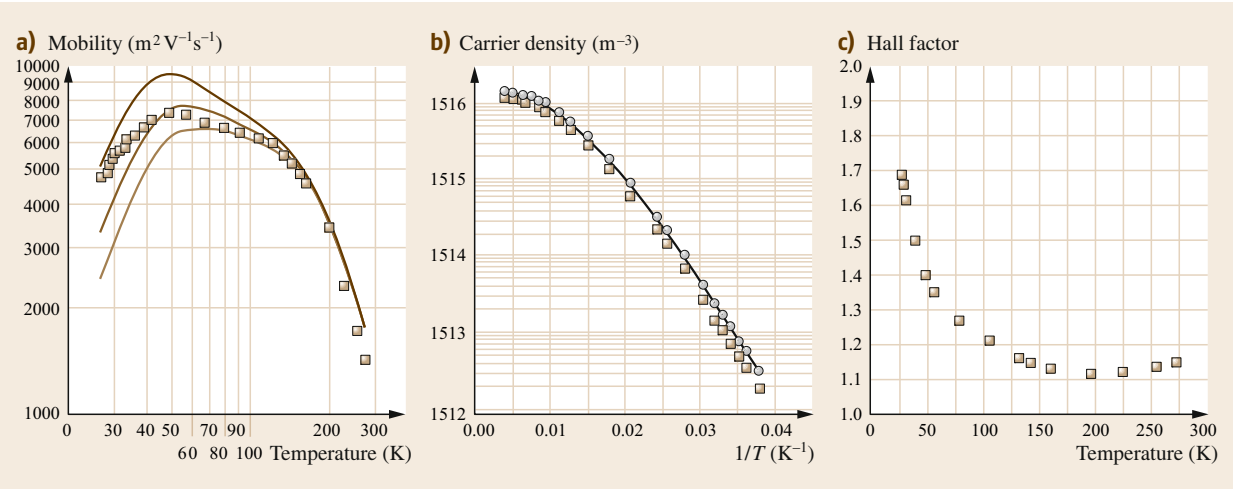


Fig. 31.11 **(a)** The measured Hall mobility data (*solid squares*) from the GaN template grown by HVPE as a function of temperature. The *solid line* is the calculated result using $N_a = 2.4\times10^{10}\,\mathrm{cm}^{-3}$, representing the best fit to the measured results. The *upper* and *lower dotted lines* are the calculated results using $N_a = 1.4\times10^{15}$ and $3.4\times10^{15}\,\mathrm{cm}^{-3}$. **(b)** The measured Hall densities n_H (*solid squares*) as a function of reciprocal temperature from the GaN template grown by HVPE. The *open circles* represent the carrier density corrected by the Hall factor, $n = n_H r_H$. The *solid line* is the fit to the theoretical expression of charge balance with hole and neutral acceptor densities neglected. **(c)** Temperature dependence of the calculated Hall factor, r_H. (After [31.217])

ing carrier concentration, dropping down to a value of about $100\,\mathrm{cm}^2/(\mathrm{V\,s})$ at a carrier concentration of $3\times10^{18}\,\mathrm{cm}^{-3}$. For an electron concentration of $1\times10^{18}\,\mathrm{cm}^{-3}$, the mobility was $250\,\mathrm{cm}^2/(\mathrm{V\,s})$. A higher room-temperature mobility of $845\,\mathrm{cm}^2/(\mathrm{V\,s})$ at an electron concentration of $\approx 6\times10^{16}\,\mathrm{cm}^{-3}$ was achieved by *Götz* et al. [31.199]. In later publications, due to the advent of the high-quality samples grown by several growth techniques, there was a significant improvement in reported room-temperature mobility values. *Nakamura* et al. [31.219] and *Binari* and *Dietrich* [31.220] were able to obtain $900\,\mathrm{cm}^2/(\mathrm{V\,s})$ room-temperature mobility at an electron concentration of 3×10^{16} and $5\times10^{16}\,\mathrm{cm}^{-3}$ respectively. Later on even higher electron mobilities of $1100\,\mathrm{cm}^2/(\mathrm{V\,s})$ at room temperature and $1425\,\mathrm{cm}^2/(\mathrm{V\,s})$ at 273 K were reported by *Yun* et al. [31.221] and *Huang* et al. [31.217] respectively, for a 200 μm thick freestanding n-type GaN template grown by HVPE. This achievement was attributed to the excellent crystalline structure of the GaN sample with low levels of compensation and the defect-related scattering. A quantitative comparison with theoretical calculations demonstrated that the one-layer and one-donor conductance model is sufficient to account for the measured data in the entire temperature range without considering any dislocation scattering and any adjustable parameters other than the acceptor concentration. The measured temperature-dependent Hall mobility, carrier concentration, and Hall scattering factor are shown in Fig. 31.11 along with the best-fit theoretical calculation based on an iterative solution of the Boltzmann equation.

As shown, a quantitative agreement with the measured mobility in the entire temperature range was obtained to within about 30%. *Heying* et al. [31.222] investigated both the morphology and electrical properties of homoepitaxial GaN layers grown by MBE as a function of Ga/N ratio during growth. GaN films grown with higher Ga/N ratios (intermediate regime) showed fewer pits with areas of atomically flat surface, which gives the highest mobility $\approx 1191\,\mathrm{cm}^2/(\mathrm{V\,s})$ reported at the time at room temperature. *Koleske* et al. [31.223] investigated the effect of the AlN nucleation layer on transport properties of GaN films grown on 6H- and 4H-SiC substrates. Room-temperature electron mobilities of 876, 884, and $932\,\mathrm{cm}^2/(\mathrm{V\,s})$ were obtained on 6H-SiC, 4H-SiC, and 3.5° off-axis 6H-SiC substrates respectively. The authors attributed the observed high electron mobilities to the improved AlN morphology and reduction in screw dislocation density near the AlN/GaN interface.

The thermal activation energy of free carriers has also been extracted from the temperature-dependent Hall concentration and mobility measurements by fitting simple exponential dependence of the carrier concentration on inverse temperature, by a two-band model and by other theoretical fitting techniques. It was concluded that the activation energy of n-type GaN free carriers lies in the range 14–36 meV depending on the extent of screening [31.217, 218, 224–226]. In the di-

lute limit, the values are close to 30 meV. In order to analyze the mobility data, one must understand the scattering processes that dominates mobility at different temperatures. Monte Carlo simulations of the electron velocity in GaN as a function of electric field at different doping concentrations and temperatures predicted a peak drift velocity of 2×10^7 cm/s at an electric field of $\approx 1.4\times10^5$ V/cm for an electron concentration of $10^{17}\,\mathrm{cm}^{-3}$ [31.227, 228]. These values show that an electron mobility as high as $900\,\mathrm{cm}^2/(\mathrm{V\,s})$ could be achieved in the case of uncompensated GaN at room temperature with $\approx 10^{17}\,\mathrm{cm}^{-3}$ doping concentration. As discussed above, in the case of high-quality samples with very low compensation, a mobility of even more than $900\,\mathrm{cm}^2/(\mathrm{V\,s})$ at room temperature with a similar doping concentration has been reported.

Albrecht et al. [31.229] calculated the electron mobility for different concentrations of the ionized impurities and at different temperatures by using a Monte Carlo simulation technique based on empirical pseudopotential band structure calculations. For practical use they have also derived an analytical expression describing the dependence of the mobility on temperature and ionized impurity concentration as

$$\frac{1}{\mu_\mathrm{e}} = a\left(\frac{N_\mathrm{I}}{10^{17}\,\mathrm{cm}^{-3}}\right)\ln\left(1+\beta_\mathrm{cw}^2\right)\left(\frac{T}{300\,\mathrm{K}}\right)^{-1.5} + b\left(\frac{T}{300\,\mathrm{K}}\right)^{1.5} + c\frac{1}{\exp\left(\frac{\Theta}{T}\right)-1}\,, \tag{31.13}$$

where

$$\Theta = \frac{\hbar\omega_\mathrm{LO}}{k_\mathrm{B}} = 1065\,\mathrm{k}\,,$$

$$\beta_\mathrm{WC}^2 = 3.00\left(\frac{T}{300\,\mathrm{K}}\right)^2\left(\frac{N_\mathrm{I}}{10^{17}\,\mathrm{cm}^{-3}}\right)\,,$$

$$N_\mathrm{I} = \frac{(1+k_\mathrm{c})}{N_\mathrm{D}}\,,$$

and

$$a = 2.61\times10^{-4}\,(\mathrm{V\,s})/\mathrm{cm}^2\,,$$

$$b = 2.90\times10^{-4}\,(\mathrm{V\,s}/\mathrm{cm}^2)\,,$$

$$c = 1.70\times10^{-2}\,(\mathrm{V\,s})/\mathrm{cm}^2\,.$$

Here N_D is the ionized donor concentration in cm^{-3} and $k_\mathrm{c} = N_\mathrm{A}/N_\mathrm{D}$ is the compensation ratio. Comparing the mobility calculated using this analytical expression and that calculated by Monte Carlo simulations, a reasonable agreement with a maximum error of 6% is realized between the temperature range of 300–600 K, and ionized donor concentration from 10^{16}–$10^{18}\,\mathrm{cm}^{-3}$.

The calculation of low field electron mobility in GaN has also been carried out by using different techniques. *Chin* et al. [31.202] have used the variational principle to calculate low field electron mobilities as a function of temperature for carrier concentrations of 10^{16}, 10^{17} and $10^{18}\,\mathrm{cm}^{-3}$ with the compensation ratio as a parameter. GaN exhibits maximum mobilities between 100 and 200 K, depending on the electron density and compensation ratio, with lower electron density peaking at lower temperature. This behavior is related to the interplay of piezoelectric acoustic phonon scattering at low carrier concentrations and ionized impurity scattering at higher carrier concentrations. The polar mode optical phonon scattering-limited maximum room-temperature mobility in GaN is found to be about $1000\,\mathrm{cm}^2/(\mathrm{V\,s})$. Although a degree of correlation is achieved with the experimental data, there is a disparity, which is attributed to the structural imperfections and an overestimated compensation ratio (compensation reduces electron mobility in GaN for a given electron concentration). Typical compensation ratios observed for MOCVD- and MBE-grown films are about 0.3, though a lower ratio of ≈ 0.24 was reported for HVPE-grown crystals.

Rode and *Gaskill* [31.203] have used an iterative technique, which takes into account all major scattering mechanisms for low field electron mobility in GaN, to investigate the dependence of mobility on electron concentration, but not on temperature. The result was applied to the Hall mobility data published by *Kim* et al. [31.231] and a good fitting between theory and experiment was demonstrated within 2.5% error for the lowest doped samples with free-electron concentrations of 7.24×10^{17} and $1.74\times10^{17}\,\mathrm{cm}^{-3}$. However, there is a significant disagreement for the more heavily doped samples having much higher free-electron concentrations, where the Born approximation applied for the ionized impurity scattering might be the reason for poor fitting at high electron concentrations. By assuming uncompensated material and carrier freeze-out onto donors, the ionization energy was theoretically determined to be about 45 and 57 meV for the best-fitted samples with free-electron concentrations of 7.24×10^{17} and $1.74\times10^{17}\,\mathrm{cm}^{-3}$ respectively. Recent developments in major growth techniques such as HVPE, MOCVD, and MBE for GaN and other group III nitride semiconductors have led to the growth of high-quality epitaxial layers [31.232–234], which allowed the comparison of the Hall data with theory at a finer scale. *Dhar* and *Gosh* [31.230] have calculated the temperature and doping dependencies of the electron mobility using an iterative technique, in which the scattering mechanisms have been treated beyond the Born approximation. The compensation ratio was used

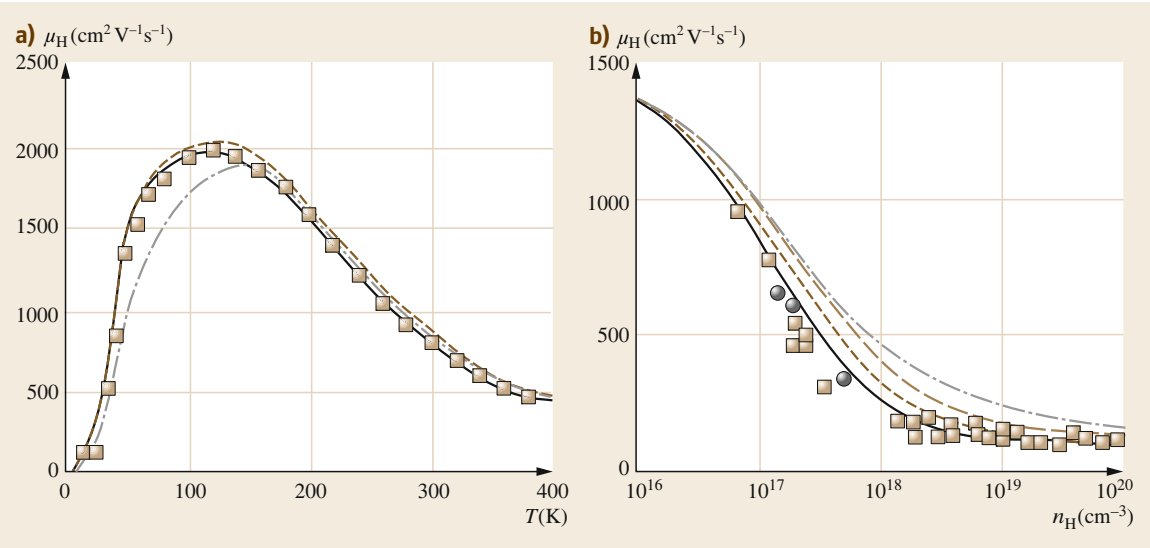


Fig. 31.12 **(a)** Temperature dependence of the calculated electron Hall mobility. The *dashed curve* shows the calculated mobility using the BH technique. The *dotted curve* shows the calculated mobility using phase-shift analysis with multi-ion screening correction. The *solid curves* show the calculated mobility using phase-shift analysis with multi-ion screening correction and electron-plasmon scattering. **(b)** Electron concentration dependence of the calculated electron Hall mobility for a compensation ratio $r = 0.2$. The *dashed curve* shows the calculated mobility using the BH technique. The *dashed-dotted curve* shows the calculated mobility using phase-shift analysis without multi-ion screening correction. The *dotted curve* shows the calculated mobility using phase-shift analysis with multi-ion screening correction. The solid curve shows the calculated mobility using phase-shift analysis with multi-ion screening correction and electron-plasmon scattering. Experimental data are taken from different references. (After [31.230])

as a parameter with a realistic charge neutrality condition. They tested their model with the experimental Hall data taken from literature [31.199, 202, 214, 218, 219, 225, 235]. Reasonable agreement was achieved between these data points and the calculation over a wide range of temperatures and electron concentrations, as shown in Fig. 31.12. They also investigated the effect of the degenerate layer at the GaN/substrate interface to extract reliable experimental values of the bulk electron mobility and concentration.

GaN samples from Kyma Technologies, Inc. (grown by hydride vapor phase epitaxy – HVPE) with a thickness of 40 µm were characterized by *Wang* et al. [31.236]. Undoped and Si-doped samples exhibited carrier densities of 4.58×10^{15} and $2.41\times10^{18}\,\mathrm{cm}^{-3}$ with mobilities of 1050 and $286\,\mathrm{cm}^2/(\mathrm{Vs})$ respectively. The decreased mobility in Si-doped GaN samples is, according to the authors, due to the increase of ionized impurity scattering. GaN grown by HVPE on an MOCVD-grown GaN/sapphire template with a photolithographically patterned Ti mask and ammonothermally grown GaN crystals (Am-GaN) was compared by *Sochacki* et al. [31.237]. Carrier concentration and carrier mobility were $2\times10^{16}\,\mathrm{cm}^{-3}$ and $1107\,\mathrm{cm}^2/(\mathrm{Vs})$, and $4\times10^{16}\,\mathrm{cm}^{-3}$ and $883\,\mathrm{cm}^2/(\mathrm{Vs})$ for GaN grown on MOCVD-GaN/sapphire and Am-GaN (at 800 mbar), respectively. The resistivity was in the range of 0.1–0.3 Ω cm in both cases. Higher mobility and lower carrier concentration suggests that HVPE-GaN grown on MOCVD-GaN/sapphire templates incorporated fewer unintentional impurities, which was also consistent with the low-temperature PL spectra (not sown here). It is believed that the better electrical characteristic of HVPE-GaN grown on MOCVD-GaN/sapphire template is the result of a higher growth rate, which fosters a higher purity of the material.

Besides the numerical simulation techniques mentioned above, recently, *Mnatsakanov* et al. [31.238] derived a simple analytical approximation to describe the temperature and carrier concentration dependencies of the low-field mobility in wide temperature $50 \le T \le 1000\,\mathrm{K}$ and carrier concentration ($10^{14} \le N \le 10^{19}\,\mathrm{cm}^{-3}$ ranges. As the first step of this model, an adequate approximation of the doping level dependence of the mobility at room temperature is used on the basis of the Goughey–Thomas approximation [31.239]

$$\mu_i(N) = \mu_{\min,i} + \frac{\mu_{\max,i} - \mu_{\min,i}}{1 + \left(\frac{N}{N_{g,i}}\right)^{\gamma_i}}\,, \qquad (31.14)$$

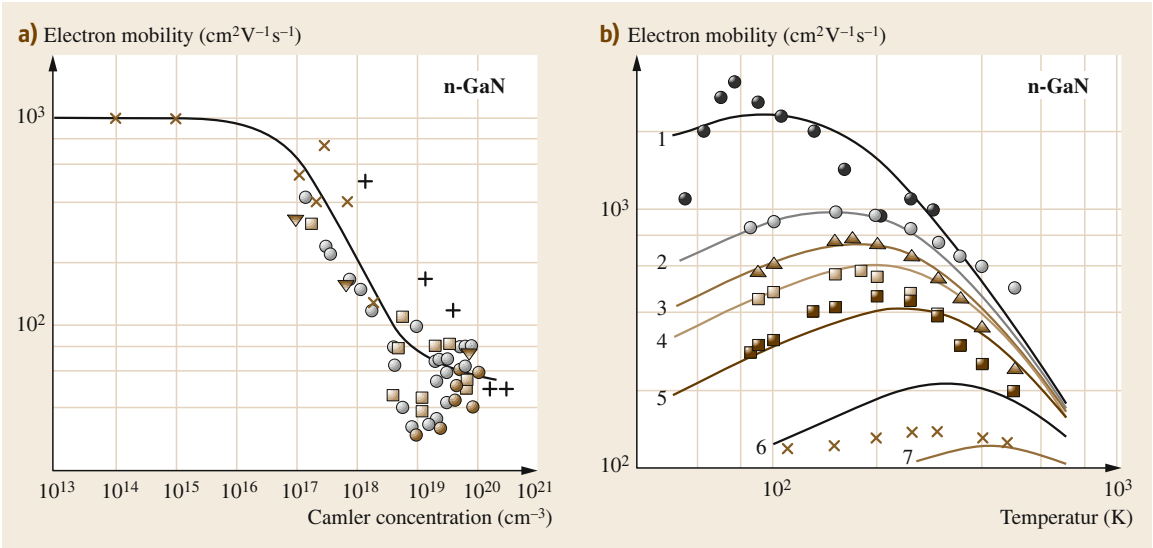


Fig. 31.13 **(a)** Low-field electron mobility as a function of doping concentration in GaN at room temperature. The *curve* represents the best approximate equation. The symbols (*crosses*, *open circles*, and *pluses*) are experimental data taken from different references for wurtzite GaN and zincblende. **(b)** Temperature dependencies of low-field electron mobility in wurtzite GaN at different values of doping concentration. Curves represent the best approximate equations. 1: $N = 3\times 10^{16}$ cm^{-3}, 2: $N = 10^{17}$ cm^{-3}, 3: $N = 1.5\times 10^{17}$ cm^{-3}, 4: $N = 2\times 10^{17}$ cm^{-3}, 5: $N = 3.5\times 10^{17}$ cm^{-3}, 6: $N = 10^{18}$ cm^{-3}, 7: $N = 3\times 10^{18}$ cm^{-3}. (After [31.230])

where i= n, p for electrons and holes respectively, $\mu_{\min,i}$, $\mu_{\max,i}$, $N_{g,i}$ and γ_i are the model parameters dependent on the type of semiconductor materials, and N is the doping concentration. Figure 31.13a shows the comparison between the calculated low-field electron mobility as a function of doping level and some experimental data on the room-temperature electron mobility in GaN.

The proposed approximation provides a rather good agreement with the experimental data. For the temperature-dependent mobility calculation, the authors derived the following equation by taking into account the main scattering mechanisms

$$\mu_i(N,T) = \mu_{\max,i}(T_0)\frac{B_i(N)\left(\frac{T}{T_0}\right)^{\beta_i}}{1+B_i(N)\left(\frac{T}{T_0}\right)^{\alpha_i+\beta_i}}\,, \tag{31.15}$$

where

$$B_i(N) = \left(\frac{\mu_{\min,i}+\mu_{\max,i}\left(\frac{N_{g,i}}{N}\right)^{\gamma_i}}{\mu_{\max,i}-\mu_{\min,i}}\right)\Bigg|_{T=T_0}\,.$$

Figure 31.13b presents the calculated temperature-dependent electron mobility and experimental data reported in the literature and a good agreement is realized. Although ionized impurity scattering is the dominant mechanism at low temperatures, where the mobility increases with increasing temperature, at high temperatures mobility is limited by polar optical phonon scattering where mobility decreases with increasing temperature. Although a good agreement was achieved between theoretical calculations and experimental data, many material and physical parameters of GaN were not available for some of the previous simulations where those parameters were treated as adjustable parameters. This underscores the need for reliable parameters in the calculation of the electron mobility and in the interpretation of experimental results to gain more accuracy.

Aluminum Nitride

Due to the low intrinsic carrier concentration, and the deep native defect and impurity energy levels (owing to the wide bandgap of AlN ≈ 6.2 eV at 300 K), the electrical transport properties of AlN have not been studied extensively and have usually been limited to resistivity measurements. Resistivities in the range of $\rho = 10^7$–$10^{13}\,\Omega$ cm have been reported for unintentionally doped AlN single crystals [31.240, 241], a value consistent with other reports [31.242–244]. The conductivity exhibited an Arrhenius behavior for all crystals and the activation energies were reported to be 1.4 eV for temperatures between 300 and 400 K and 0.5 eV between 300 and 330 K. Intentional doping of AlN has resulted

in both n- and p-type AlN by introducing Hg and Se respectively (caution is warranted here). *Gorbatov* and *Kamyshon* [31.245] obtained n-type conductivity of polycrystalline AlN with the incorporation of Si. Unintentionally doped n-AlN films grown by a modified physical transport technique by *Rutz* [31.246] had a resistivity as low as 400 Ω cm. Although the source of the electrons has not been determined, *Rutz* et al. [31.247] observed an interesting transition in their AlN films in which the resistivity abruptly decreased by two orders of magnitude with an increase in the applied bias. This observation found applications to switchable resistive memory elements that are operated at 20 MHz. Based on an ab initio density functional calculation, *Fara* et al. [31.248] concluded that energy levels are deep for acceptors, donors, and native defects in AlN, due to the semi-insulating characteristic of AlN. For example, acceptors are deeper (Be ≈ 0.25, Mg ≈ 0.45) and less soluble in AlN than that in GaN, and both the extrinsic donors Si_{Al} and C_{Al}, and the native donor V_N are found to be deep (about 1 to 3 eV below the conduction band). Consequently, doped AlN may in general turn out to be semi-insulating, when attained using Al-rich conditions, in agreement with the known doping difficulties of high Al-content AlGaN alloys.

As far as the mobility of AlN is concerned, in early studies, *Edwards* et al. [31.240] and *Kawabe* et al. [31.241] carried out Hall measurements in p-type AlN, which produced a very rough estimate of the hole mobility; $\mu_p = 14\,cm^2/(V\,s)$ at 290 K. Later work on carbon-doped materials has resulted in highly conductive p-type AlN with hole mobilities in the range of $60\,cm^2/(V\,s)$ for carbon doping of $10^{18}\,cm^{-3}$ (*Wongchotigul* et al. [31.249] – again caution should be exercised here). The theoretical calculations have only been made to estimate the mobility in AlN by *Chin* et al. [31.202]. Since AlN is a semi-insulator with extremely low carrier concentration, only phonon-limited scattering processes were considered in the calculation of the temperature-dependent mobility as shown in Fig. 31.14. The mobility was found to decline rapidly at high temperatures, with a value of about $2000\,cm^2/(V\,s)$ at 77 K and dropping to $300\,cm^2/(V\,s)$ at 300 K for optical phonon-limited mobility.

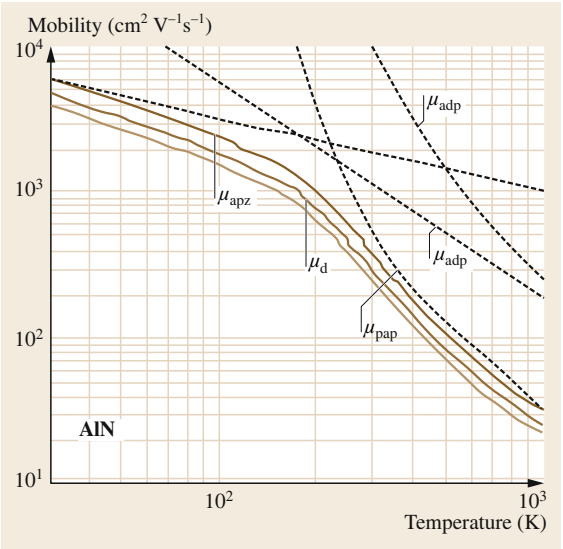


Fig. 31.14 The phonon-limited electron drift mobility in AlN plotted as a function of temperature for $m^* = 0.48$. The *upper* and *lower solid curves* are for the phonon-limited electron drift mobility calculated using $m^* = 0.42$ and 0.52 respectively. (After [31.202])

Indium Nitride

InN suffers from the lack of a suitable substrate material as well, causing a high concentration of extended defects in the films. A sapphire substrate is usually used for InN growth, but the difference in lattice constants between InN and sapphire is as large as 25% and even more than 19% with the other nitrides. A large disparity of the atomic radii of In and N is an additional contributing factor to the difficulty of obtaining InN of good quality. Because of all these factors, the electron mobilities obtained from various films have varied significantly. Electrical properties also vary substantially depending on the growth techniques. A range of film deposition techniques, such as reactive ion sputtering, reactive RF magneton sputtering, metal organic vapor phase epitaxy and MBE, have been used. Table 31.6 gives values of electron mobility and concentration in InN films grown by various growth techniques.

The transport properties reported in the literature are mostly based on Hall effect measurements, assuming the Hall scattering factor to be unity. In the majority of the samples, electron mobilities often remain relatively poor, despite predicted values as high as $3000\,cm^2/(V\,s)$ at room temperature for InN [31.268]. It is widely believed that nitrogen vacancies lead to large background electron concentrations, which are responsible for the observed low electron mobility. An empirical linear relationship between the electron mobility and electron concentration can be deduced from Table 31.6 for a series of InN films, although not all films exhibit this type of behavior. A systematic study carried out by *Tansley* et al. indicates that the electron concentration decreases as the nitrogen density is increased in the plasma in reactive ion sputtering. Although the reported time is relatively old, the maximum mobility of $2700\,cm^2/(V\,s)$ at an electron concentration of $5\times10^{16}\,cm^{-3}$ was reported for RF reactive ion sputtered growth of InN. Early studies of the electron

Table 31.6 A compilation of electron mobilities obtained n wide-bandgap InN on different substrates and for various deposition conditions in InN

Growth method	Carrier concentration (cm^{-3})	Electron mobility ($cm^2\,V^{-1}\,s^{-1}$)	Reference
Reactive ion sputtering	7.0×10^{18}	250	[31.250]
Reactive ion sputtering	2.1×10^{17}	470	[31.251]
Reactive ion sputtering	8.0×10^{16}	1300	[31.251]
Reactive ion sputtering	5.5×10^{16}	2700	[31.251]
Reactive ion sputtering	2.0×10^{20}	9	[31.252]
RF magnetron sputtering	–	44	[31.253]
Reactive ion sputter	$\approx 10^{20}$	60	[31.254]
Plasma assisted MBE	$\approx 10^{20}$	229	[31.255]
ECR-assisted MOMBE	2.0×10^{20}	100	[31.256]
ECR-assisted reactive ion sputtering	–	80	[31.257]
Reactive sputter	6.0×10^{18}	363	[31.258]
MOVPE	5.0×10^{19}	700	[31.233]
Migration-enhanced epitaxy	3.0×10^{18}	542	[31.259]
RF MBE	3.0×10^{19}	760	[31.260]
MOMBE	8.8×10^{18}	500	[31.261]
MBE	$2-3\times10^{18}$	800	[31.262]
Reactive ion sputtering	$\approx 10^{19}$	306	[31.263]
RF MBE	1.0×10^{19}	830	[31.264]
Plasma-assisted MBE	1.6×10^{18}	1180	[31.265]
MBE	4×10^{17}	2100	[31.266]
Plasma-assisted MBE	1.4×10^{18}	1420	[31.267]

mobility of InN as a function of the growth temperature indicate that the mobility of ultrahigh vacuum-electron cyclotron resonance-radio-frequency magnetron sputtering (UHV-ECR-RMS)-grown InN can be as much as four times the mobility of conventionally grown (vacuum deposition) InN [31.257]. However, more recent work indicates a progressive improvement in electrical properties of InN films grown by MBE and MOVPE. The value of electron mobility as high as $1420\,cm^2/(V\,s)$ at an electron concentration of $1.4\times10^{18}\,cm^{-3}$ was reported for the InN layers grown by plasma-assisted MBE using a low-temperature-grown GaN intermediate layer and a low-temperature-grown InN buffer layer. Hall measurements in InN films grown on AlN buffer layers, which are in turn grown on sapphire, indicated an electron mobility to be $2100\,cm^2/(V\,s)$ with a relatively low electron concentration ($4\times10^{17}\,cm^{-3}$) at room temperature in material grown by MBE. Very high inadvertent donor concentrations, $>10^{18}\,cm^{-3}$, seem to be one of the major problems for further progress of device applications of InN. O_N and Si_{In} and possible interstitial H have been proposed to be the likely dominant defects responsible for high electron concentration in the state-of-the-art MBE-grown InN due to their low formation energies [31.269, 270].

Electron mobility in InN has been calculated using the variational principle for a range of temperatures, carrier concentrations, and compensation ratios [31.202]. Figure 31.15 shows the theoretical results with experimental mobility values taken from [31.251]. The calculated peak mobilities are found to be 25 000, 12 000, and $8000\,cm^2/(V\,s)$ for carrier concentrations of 10^{16}, 10^{17} and $10^{18}\,cm^{-3}$ respectively, at different temperatures (between 100–200 K) depending on the electron density and compensation ratio. This is due to the interplay of piezoelectric acoustic phonon scattering at low concentrations and ionized impurity scattering at high temperatures. These two mechanisms are the dominant scattering mechanisms below 200 K, while polar-mode optical phonon scattering is the most significant process above this temperature. The low-concentration limit for room-temperature mobility in uncompensated InN is estimated to be $4400\,cm^2/(V\,s)$.

31.5.2 High Field Transport

Ensemble Monte Carlo simulations have been the popular tool to theoretically investigate the steady-state electron transport in nitrides. In particular, the steady-state velocity-field characteristics have been determined for AlN [31.271, 272], GaN [31.229, 271, 273–276], and InN [31.277, 278]. These reports show that the drift velocity initially increases with the applied electric field to reach a maximum and decreases with further increase in the field strength. The intervalley electron

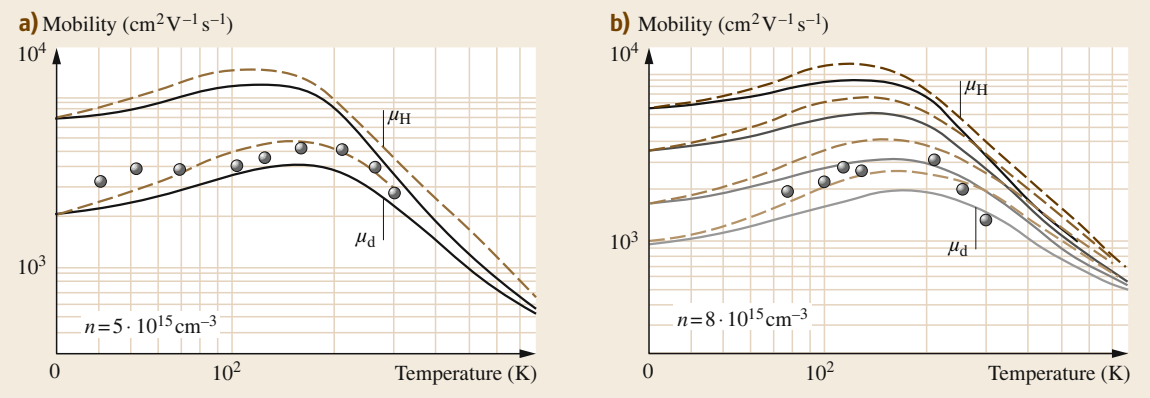


Fig. 31.15a,b The electron drift and Hall mobilities in InN as a function of temperature for **(a)** $n = 5 \times 10^{16}\,\mathrm{cm}^{-3}$ for compensation ratios of 0.00 and 0.60; and **(b)** $8 \times 10^{16}\,\mathrm{cm}^{-3}$ for compensation ratios of 0.00, 0.30, 0.60, and 0.75. (After [31.202])

transfer plays a dominant role at high electric fields leading to a strongly inverted electron distribution and to a large negative differential resistance (NDR). Note that credible experimental confirmation of intervalley scattering is still pending. The reduction in the drift velocities was attributed to transfer of electrons from the high-mobility Γ-valley to the low-mobility satellite X-valley. The onset electric field and peak drift velocities, however, show some disparity among the reported calculations due to the variety of the degree of approximation and used physical constants of the materials. A typical velocity-field characteristic for bulk group III nitrides at room temperature is shown in Fig. 31.16 along with the well-studied GaAs data used to test the author's Monte Carlo model. For the doping concentration set to $10^{17}\,\mathrm{cm}^{-3}$, InN has the highest steady-state peak drift velocity; $4.2 \times 10^7\,\mathrm{cm/s}$ at an electric field of 65 kV/cm. In the case of GaN and AlN, steady-state peak drift velocities are rather low and occur at larger electric fields; $2.9 \times 10^7\,\mathrm{cm/s}$ at 140 kV/cm for GaN and $1.7 \times 10^7\,\mathrm{cm/s}$ at 450 kV/cm for AlN.

Another interesting aspect of electron transport is its transient behavior, which is relevant to short channel devices with dimensions smaller than 0.2 μm where a significant overshoot is expected to occur in the electron velocity over the steady-state drift velocity. Transient electron transport and velocity overshoot in both wurtzite and zincblende GaN, InN, and AlN have been studied theoretically by a number of groups. *Foutz* et al. [31.279] employed both Monte Carlo simulations and one-dimensional energy-momentum balance techniques. They used a three-valley model for the conduction band by taking the main scattering mechanisms, such as ionized impurity, polar optical phonon, acoustic phonon through deformation potential and

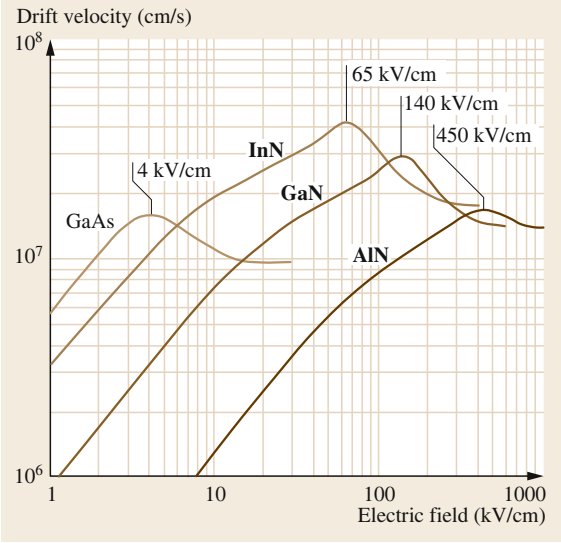


Fig. 31.16 The velocity-field characteristics associated with wurtzite GaN, InN, AlN, and zincblende GaAs. In all cases, the temperature was set to 300 K and the doping concentration was set to $10^{17}\,\mathrm{cm}^{-3}$. The critical fields at which the peak drift velocity is achieved for each velocity-field characteristic are clearly marked: 140 kV/cm for GaN, 65 kV/cm for InN, 450 kV/cm for AlN, and 4 kV/cm for GaAs. (After [31.279])

piezoelectric, and intervalley scatterings, into account. In particular, they examined how electrons, initially in equilibrium, respond to the instant application of a constant electric field. Figure 31.17 shows the average velocity of the electrons in AlN, GaN, and InN as a function of distance.

According to their calculation, electron velocity overshoot only occurs when the electric field exceeds

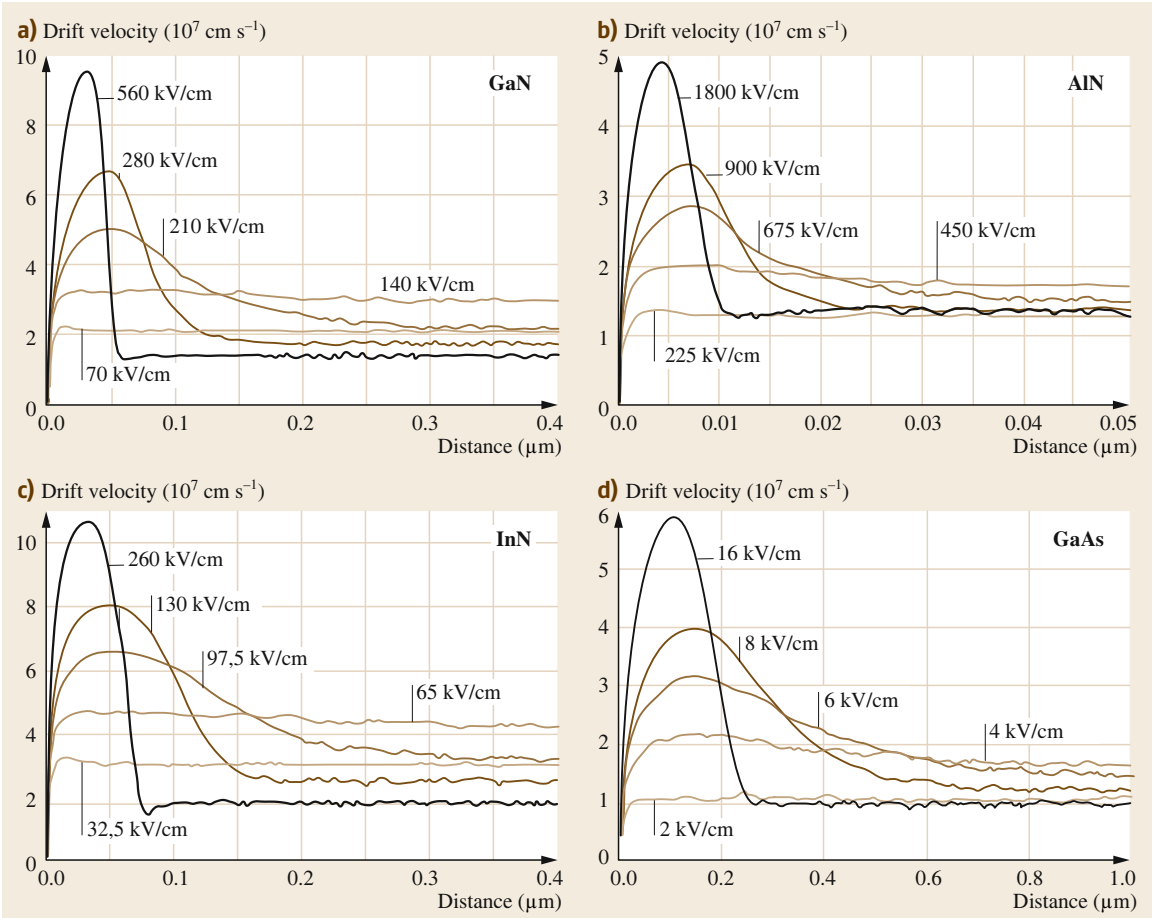


Fig. 31.17a–d The average electron velocity as a function of the displacement for various applied fields for the cases of **(a)** GaN, **(b)** InN, **(c)** AlN, and **(d)** GaAs. In all cases, we have assumed an initial zero-field electron distribution, a crystal temperature of 300 K, and a doping concentration of 10^{17} cm^{-3}. (After [31.279])

a certain critical value unique to each material and it lasts over a certain distance dependent on applied field. These critical fields are points where the highest steady-state peak drift velocities are achieved and being reported as 65, 140, and 450 kV/cm with corresponding peak velocities of 2.9×10^7, 1.7×10^7, and 1.6×10^7 cm/s for InN, GaN, and AlN respectively. Among them, InN exhibits the highest peak overshoot velocity on the order of 10^8 cm/s at 260 kV/cm and the longest overshoot relaxation distance on the order of 0.8 μm at 65 kV/cm. The variation of the velocity with respect to the field, not just the peak velocity, must be considered in our attempts to minimize the transit time. The upper bound for the cutoff frequency of InN- and GaN-based HFETs benefits from larger applied fields and accompanying large velocity overshoot when the gate length is less than 0.3 μm in GaN- and 0.6 μm in InN-based devices. However, all measured cutoff frequencies are gate-length-dependent and well below these expectations, indicating that devices operate in the steady-state regime and other effects, such as real-space transfer, should also be considered. On the other hand, there is some controversy in the reports related to the onset of velocity overshoot in nitride semiconductors. For example, *Rodrigues* et al. [31.280] reported overshoot onsets at 10 kV/cm in InN, 20 kV/cm in GaN, and 60 kV/cm in AlN by using a theoretical model based on a nonlinear quantum kinetic theory, which compares the relation between the carriers' relaxation rate of momentum and energy. Experimental investigations of the transient transport in group III nitrides are very limited and only a few results are reported by using different techniques. *Wraback* et al. [31.281] employed a femtosecond time-resolved electroabsorption technique to study the transient electron velocity overshoot for trans-

port in the AlGaN/GaN heterojunction p-i-n photodiode structures. It has been reported that electron velocity overshoot can be observable at electric fields as low as 105 kV/cm. The velocity overshoot increases with electric fields up to ≈ 320 kV/cm with a peak velocity of 7.25×10^7 cm/s relaxing within the first 0.2 ps after photoexcitation. The increase in electron transit time across the device and the decrease in peak velocity overshoot with increasing field beyond 320 kV/cm is attributed to a negative differential resistivity region of the steady-state velocity-field characteristic in this high field range. *Collazo* et al. [31.282] used another experimental technique based on the measurement of the energy distribution of electrons, which were extracted into a vacuum through a semitransparent Au electrode, after their transport through intrinsic AlN heteroepitaxial films using an electron spectrometer. They observed electron velocity overshoot as high as five times the saturation velocity and a transient length of less than 80 nm at the field of 510 kV/cm. In order to design an electronic device that is expected to operate at high power and high frequency, one could consider harnessing velocity overshoot in group III nitride semiconductor heterojunctions. However, the strong coupling between hot electrons and LO phonons appear to limit the velocity attainable by electrons, being dependent on the LO phonon decay time. A systematic investigation of InN, GaN, AlN and their alloys as a function of various parameters in dynamics mode would be very beneficial for the development of higher performance, next generation electronic and optoelectronic devices.

31.6 Optical Properties of Nitrides

The optical properties of a semiconductor have vital importance due to the wide range of useful information that can be obtained in our attempts to understand the semiconductor properties, especially the direct bandgap material for optoelectronic applications. For ideal defect-free semiconductors, the optical transitions occur between conduction band and valence band, including excitonic effects due to the Coulomb interaction, where excitons can be classified into free and bound varieties. For high-quality samples with low defect concentrations, the free excitons can also exhibit excited states in addition to their ground state. Bound excitons are those that are bound to dopants or defects, which usually create a discrete energy below the bandgap energy, and therefore influence the optical behavior of the excitons as well as optical absorption and emission. Excitons could be bound to neutral or charged donors and acceptors. It is generally assumed that the dominant coupling is of the like particles for principal bound exciton (BEs) states associating with neutral donors and acceptors [31.283]. For the case of shallow neutral donor-bound excitons (DBEs), the two electrons in the BE state are assumed to pair off into a two-electron state with zero spin, where the additional hole is assumed to be weakly bound in the net hole-attractive Coulomb potential set up by this bound two-electron aggregate. A similar scenario applies to acceptor-bound excitons (ABEs). The DBEs and ABEs are by far the most important cases for bound excitons in direct-bandgap materials, and thus attract intensive investigation. Various techniques can be used to investigate the optical properties of group III nitrides, and the most commonly used ones are optical absorption, transmission, photoreflection, photoluminescence, time-resolved photoluminescence, cathodeluminescence, calorimetric spectroscopy, pump-probe spectroscopy, and so on. This section will only summarize some important optical properties of group III nitride materials.

31.6.1 Gallium Nitride

Free Exciton in GaN

The conduction band (Γ_7^c) of wurtzite GaN is mainly constructed from the s state of gallium, while the valence band is mainly constructed from the p state of nitrogen. Due to the crystal-field and spin-orbit interactions, the valence band splits into three bands, namely the highest Γ_9^v (heavy hole), upper Γ_7^v (light hole), and lower Γ_7^v (crystal-field split band). The optical transitions between the conduction and valence bands are usually denoted as $\mathrm{A} \equiv \Gamma_7^c \leftrightarrow \Gamma_9^v$, $\mathrm{B} \equiv \Gamma_7^c \leftrightarrow \Gamma_7^v$ (the upper one), and $\mathrm{C} \equiv \Gamma_7^c \leftrightarrow \Gamma_7^v$ (the lower one), which have almost the same binding energy of around 25–26 meV [31.284]. For optical one-phonon processes, these three excitons of the wurtzite structure obey the following selection rules: for σ polarization ($E \| c$ and $k \perp c$ axis), C all excitons are allowed although C exciton is quite weak. For π polarization ($E \| c$ and $k \perp c$ axis), C exciton is strongly allowed and B exciton is weakly observable while A exciton is forbidden in this geometry. In the α polarization ($E \perp c$ and $k \| c$), all three transitions are clearly observable [31.285]. The A exciton consists of a two-fold degenerate spin-singlet state with Γ_5 symmetry and a two-fold degenerate spin-triplet state with Γ_6 symmetry. Both the B and C

excitons consist of a two-fold degenerate spin-singlet state with Γ_5 symmetry, a spin-triplet state with Γ_2 symmetry, and a mixed singlet-triplet state with Γ_1 symmetry. Among all the exciton states, only those with Γ_5 and Γ_1 symmetry are dipole-allowed and can be seen in reflectance or photoluminescence spectra, where the Γ_5 states are allowed for $E \perp c$ polarization while the Γ_1 states are allowed for $E \| c$ polarization. In order to reveal the fine structure of the exciton states, which have splitting on the order of 1–2 meV [31.286], the spectroscopic line width in optical spectra (such as photoluminescence) needs to be a fraction of an meV. The line width is determined by microscopic electric field and strain fields, which in practice requires single crystals with a low residual doping ($< 10^{16}\,\mathrm{cm}^{-3}$) and a low structural defect density ($\leq 10^{6}\,\mathrm{cm}^{-2}$). These requirements can only be achieved for nominally undoped freestanding GaN with a thickness of at least several hundred μm, which is not possible for most commonly available GaN grown on foreign substrates, such as sapphire, SiC or silicon. The commonly employed techniques for investigating the exciton properties are photoluminescence (PL), absorption, and/or derivative techniques like photoreflectance (PR) and calorimetric absorption and reflection techniques [31.287, 288], which are powerful techniques to determine the exciton binding energy, exciton Bohr radii, dielectric constant, and spin-orbit and crystal-field parameters with the aid of the quasicubic model.

There have been intensive studies on the emission properties of GaN during the last decade. The energy position, relative intensities, and linewidth of the emission peaks have been found to strongly depend on the residual strain, crystal line defects, and impurity incorporation, thus making the interpretation of the spectra in heteroepitaxial GaN layers rather difficult and even leading to confusing conclusions. With the availability of high-quality freestanding and homoepitaxial GaN layers, correct identifications of the emission peaks becomes possible. There are reports in the literature on the reflectance study for thick GaN epilayers [31.289–293] as well as homoepitaxial layers [31.294–298] and bulk GaN [31.58, 294, 299–303]. Studies by Monemar et al. on thin and thick GaN layers on various substrates, including homoepitaxial layers on GaN substrates, concludes that the A, B, and C excitons in GaN, with an accuracy of ±2 meV, are at 3.478, 3.484, and 3.502 eV respectively, at 2 K. These values are consistent with results obtained from freestanding or bulk GaN samples published in the literature as demonstrated in Table 31.7. An example of a reflectance spectrum obtained in the α polarization for a 2 μm-thick GaN epilayer on sapphire is displayed in Fig. 31.18, where the corresponding excitonic transition energies were

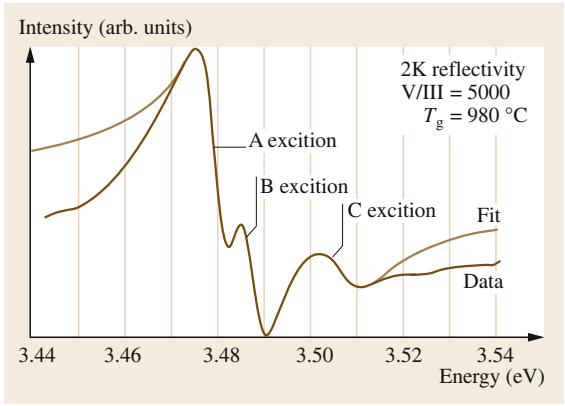


Fig. 31.18 Reflectance spectrum taken at 2 K for a 2 μm-thick GaN film grown on a sapphire substrate. The experimental data was fitted using a damped-oscillator theoretical model. (After [31.304])

evaluated using a classical damped-oscillator transition model [31.304].

Another powerful technique for studying the exciton structure is photoluminescence (PL), and Fig. 31.19 displays a typical PL spectrum taken at 2 K, which is dominated by donor-bound excitons, for an almost strain-free thick GaN epilayer grown on sapphire by HVPE. The free excitons appear at relatively higher energies, at their ground states as well as the $n = 2$ excited states of the A exciton [31.307]. The splitting between the A, B, and C exciton ground states, Δ_{AB} and Δ_{AC}, are (5.5 ± 0.1) and (22.0 ± 0.1) meV, which are close to recently reported values for quasibulk GaN samples grown by HVPE [31.303]. Based on the exciton energies, the crystal-field splitting and spin-orbit splitting of the top three valance bands were estimated to be $\Delta_{cr} = 20 \pm 2$ meV and $\Delta_{so} = 10 \pm 2$ meV respectively [31.288]. Other reported values are $\Delta_{cr} \approx \Delta_{so} \approx 16 \pm 2$ meV [31.240], and $\Delta_{cr} \approx 12.3 \pm 0.1$ meV and $\Delta_{so} \approx 18.5 \pm 0.1$ meV.

Another interesting feature of exciton lines is the fine structure near the band edge of GaN due to various splitting mechanisms. As mentioned earlier, the A, B, and C excitons have spin-related degenerate states. Based on experimental results as well as theoretical calculation, the following splittings have been reported [31.314]: $\Gamma_5(\mathrm{A}) - \Gamma_6(\mathrm{A}) = 0.53$ meV, $\Gamma_5(\mathrm{B}) - \Gamma_2(\mathrm{B}) = 0.4$ meV, $\Gamma_1(\mathrm{B}) - \Gamma_2(\mathrm{B}) = 0.41$ meV, $\Gamma_5(\mathrm{C}) - \Gamma_2(\mathrm{C}) = 0.23$ meV and $\Gamma_1(\mathrm{C}) - \Gamma_2(\mathrm{C}) = 0.75$ meV, where we can find out that the Γ_1 and Γ_1 states of the B exciton are almost degenerate with only 0.1 meV difference. For wavevector $k \perp c$, the long-range exchange interaction splits the spin-singlet states Γ_5 into a transverse and a longitudinal state. The longitudinal-transverse splitting of $\Gamma_5(\mathrm{A})$,

Table 31.7 Reported free-exciton (FX) energies in GaN using different experimental techniques

FX (A)	FX (B)	FX (C)	Orientation	Substrate	Strain	Growth	Technique	Reference
3.4776	3.4827	3.5015	*m*-Plane	Bulk GaN	Unstrained	Ammonothermal	PL	[31.294]
3.4786	3.4834		*c*-Plane	Free standing	Unstrained	HVPE	PL	[31.301]
3.4779	3.4831		*c*-Plane	Bulk GaN	Unstrained	Ammonothermal	PL	[31.295]
3.477	3.494	3.511	*a*-Plane	Sapphire	Compressive	ELO	Reflectivity	[31.305]
3.478	3.483	3.501	*c*-Plane	GaN template	N/A	PA-MBE	Reflectivity	[31.306]
3.4760	3.4817	3.4991	*c*-Plane	Bulk GaN	Unstrained	MOCVD	PR	[31.296]
3.483	3.491	3.512	*c*-Plane	Sapphire	Compressive	MOCVD	PR	[31.296]
3.474	3.481	3.501	*c*-Plane	Sapphire	Compressive	VPE	Reflectance	[31.283]
3.480	3.486	3.503	*c*-Plane	Sapphire	Compressive	HVPE	PL	[31.307]
3.485	3.493	3.518	*c*-Plane	Sapphire	Compressive	MOCVD	Reflectance	[31.308]
3.4903	3.4996	3.525	*c*-Plane	Sapphire	Compressive	MBE	Reflectance	[31.293]
3.476	3.489	3.511	*c*-Plane	ZnO	Compressive	RAMBE	Reflectance	[31.309]
3.470	3.474	3.491	*c*-Plane	6H-SiC	Tensile	MOCVD	Reflectance	[31.310]
3.478	3.484	3.502	*c*-Plane	GaN	Unstrained	MOCVD	Reflectance	[31.311]
3.4772	3.4830	3.4998	*c*-Plane	Bulk GaN	Unstrained	Na-Ga melt	Reflectance	[31.299]
3.490	3.500	3.520	*c*-Plane	Freestanding	Unstrained	HVPE	Contactless electroreflectance	[31.312]

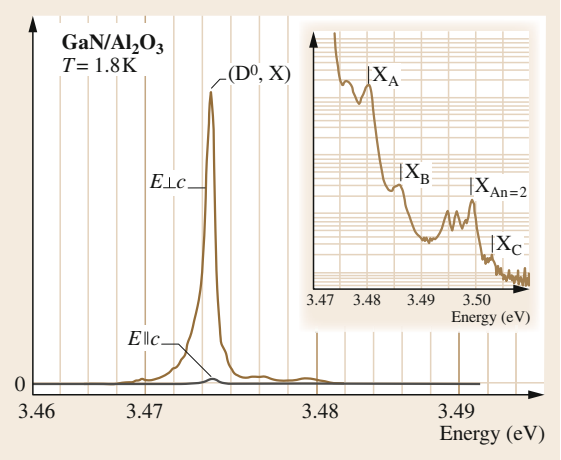


Fig. 31.19 Photoluminescence spectra at 1.8 K for a 400 µm-thick GaN grown on a sapphire substrate, where the spectrum is dominated by the donor-bound exciton transition. The intrinsic exciton states are resolvable after magnification as demonstrated in the *inset*. (After [31.313])

Γ_5(B) and Γ_5(C) states has been found to be 1.02, 0.85 and 0.34 meV respectively, in a homoepitaxial GaN layer [31.298].

Discrepancies on the reported results of free-exciton energies can be noted as indicated in Table 31.7, most of which is due to the in-plane strains induced by the foreign substrates. For most of the earlier studies of free excitons in GaN, the substrates used were either sapphire or SiC, thus inducing compressive or tensile strains respectively. When the strain is isotropic the crystal symmetry and correspondingly the symmetry of the exciton states are not changed. Instead, the strain will mix the states with different symmetry, modifying their energies and oscillator strengths [31.303]. For compressive strains, caused by for example sapphire, the bandgap, the three free-exciton energies, and the splittings, are blueshifted compared to that of unstrained GaN layers. The A and B free-exciton energies have been found to increase by as much as 20 meV at 2 K, while the C free-exciton energy increased about 50 meV [31.291]. An opposite trend was observed for tensile-strained GaN, which decreases the bandgap, overall exciton energies, and the splittings among the three exciton energies. The dependence of the free-exciton energies on the in-plane strains of GaN films is displayed in Fig. 31.20. A frequently cited set of free-exciton energies is 3.4776, 3.4827, and 3.5015 eV for the A, B, and C excitons respectively, as measured on ammonothermally grown bulk GaN [31.294].

The temperature dependence of free-exciton energies was also found to be dependent on the particular sample and strain as demonstrated in Fig. 31.21, where comparisons of temperature dependence were based on three samples: relaxed (bulk GaN), compressed (grown on sapphire), and under tension (grown on SiC). The temperature dependence of the excitonic resonance, in the absence of localization, can be described by the Varshni empirical relation

$$E(T) = E_0(0) - \frac{\alpha T^2}{(\beta + T)} , \qquad (31.16)$$

where $E_0(0)$ is the transition energy at 0 K, and α and β are the temperature coefficients.

Similar to other direct-bandgap semiconductors, the dipole-allowed exciton states in GaN encounter strong coupling with photons, which results in formation of a mixed-mode state (so called exciton-polariton) with

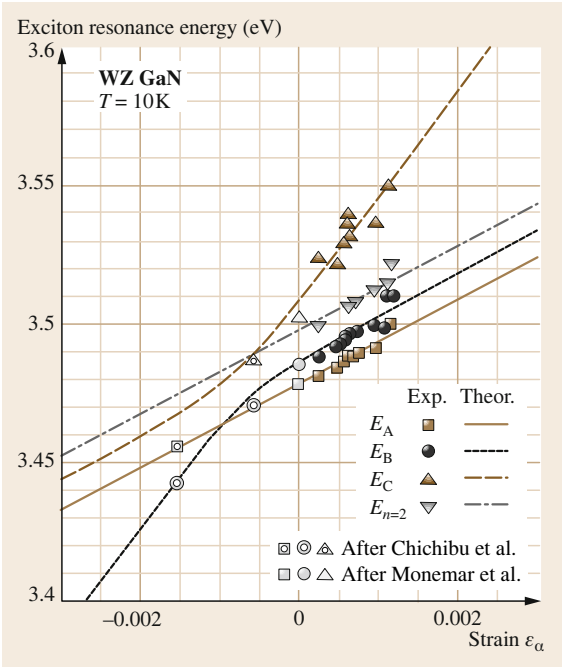


Fig. 31.20 Dependence of free-exciton energies on strain in wurtzite GaN grown on a sapphire substrate, where the strain values were obtained from measured lattice parameters. Also shown is the theoretical modeling of the strain dependence of exciton energies. (After [31.315])

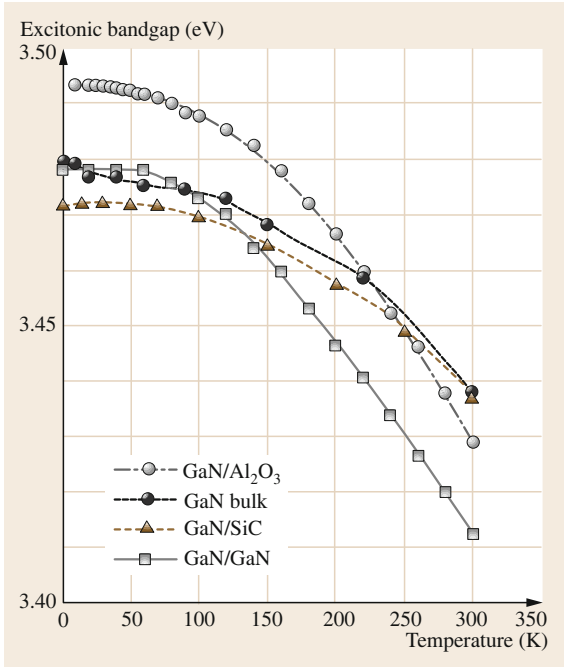


Fig. 31.21 Temperature dependence of the A free exciton energy in samples grown on different substrates. (After [31.297, 316])

a two-branch dispersion curve in the energy region near the intersection of the original exciton and photon dispersion curves [31.317]. The existence of the exciton-polariton has a profound effect on the optical emission in GaN, making the exciton recombination a propagation of polariton waves to the crystal surface and subsequent transmission to the outside as photons instead of transition from the exciton state to the photon state and subsequent photon propagation [31.318]. Thus, the emission line width does not reflect the inherent lifetime broadening of the free excitons. Instead, the emission peaks are usually quite broad and their line shapes are not exactly symmetric and are determined by the spectral dependence of the density of states, group velocity, and transmission coefficient of the different polariton branches [31.301]. Although it is difficult to study the emission of exciton-polaritons, the experimentally observed exciton-polariton spectrum is usually a doublet peak with a dip around the energy of the longitudinal exciton as demonstrated in heteroepitaxial [31.319], homoepitaxial [31.320], and freestanding GaN layers [31.321], where the doublet peaks were interpreted as an emission from the lower and upper polariton branches. An example of excitonic spectra taken at 4.2 K for homoepitaxial GaN layers is illustrated in Fig. 31.22, where the A exciton line splits into two components about 2 meV apart with the lowest component at about 3.477 eV.

Free-exciton transitions in wide-bandgap materials have a characteristic of coupling to phonons, which results in homogeneous broadening of the free-exciton states. At low temperatures, the coupling is dominated by exciton-acoustic phonon scattering, while at high temperatures the longitudinal-optical (LO) phonon Fröhlich scattering is dominant. The exciton line width is composed of three parts as [31.322]

$$\Gamma(T) = \Gamma_0 + \gamma_{\mathrm{ph}} T + \frac{\Gamma_{\mathrm{LO}}}{\left[\exp\left(\frac{E_{\mathrm{LO}}}{k_{\mathrm{B}}T}\right) - 1\right]} ,$$

where Γ_0 is the inhomogeneous line width due to the scattering by defects and impurities, and exciton-exciton interaction, γ_{ph} is the exciton-acoustic-phonon coupling strength, Γ_{LO} is the exciton-LO-phonon coupling strength, and E_{LO} is the energy of the LO-phonons involved in the scattering event (≈ 92 meV for GaN [31.323]). The temperature dependence of the line width has been investigated by many groups with various methods including absorption, PL, reflectance, spectroscopy ellipsometry, and four-wave mixing; however, the reported values have a large spread (varies between 10 and 25 meV) as summarized by *Zhang*

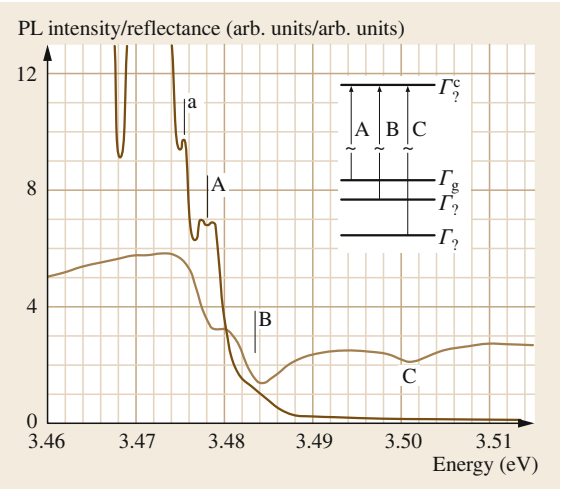


Fig. 31.22 Doublet structure of the A exciton as measured by photoluminescence (*darker line*) at 4.2 K for a homoepitaxial GaN layer. Also shown is the corresponding reflectance spectrum for the same sample (*lighter line*). (After [31.311])

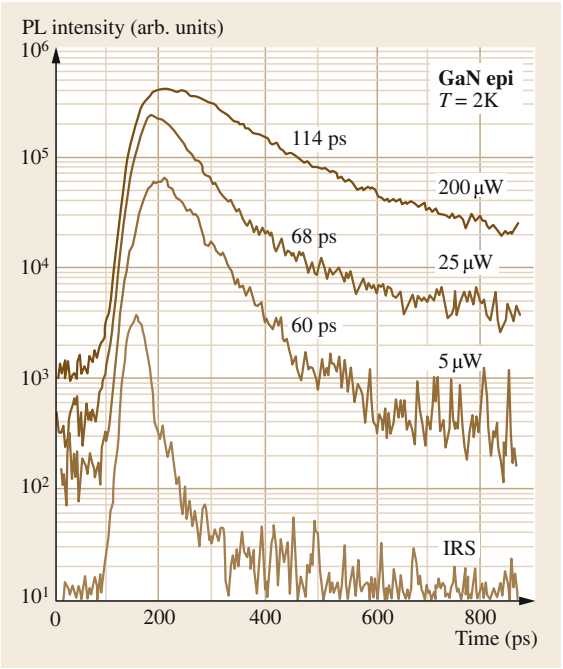


Fig. 31.23 Photoluminescence transients at 2 K for a GaN epilayer grown on sapphire, measured at different excitation power intensities. IRS denotes the instrumental response in the experiment. (After [31.288])

et al. [31.324]. The exciton-LO-phonon coupling is strongest for the first two replicas, and the 3-LO and 4-LO replicas are also clearly observable. As predicted by theory, the LO-phonon replicas are linearly proportional to the temperature, and this behavior is experimentally confirmed for the A exciton for temperatures below 100 K [31.325]. The line shape of the 1-LO and 2-LO replicas can be fitted to a good accuracy with $\approx E^{3/2}\exp(-E/k_BT)$ and $\approx E^{1/2}\exp(-E/k_BT)$ respectively, where E is the exciton kinetic energy and the exponential terms represent the equilibrium Maxwell distribution [31.326, p. 177].

The binding energies of the excitons determine the bandgap of the materials, which also denote the thermal stability of the excitons. There is large discrepancy on the binding energies of excitons in GaN, ranging from 18 to 26 meV as reviewed by [31.316], which was explained due to the strain in the studied layers or misinterpretations of some features in the optical spectra. The exciton binding energies can be obtained indirectly using temperature dependence of the free-exciton transition in PL [31.327] or the position of the excited states if the positions of the excited states of the excitons are known [31.328–330]. The temperature dependence of the excitons energies showed a monotonic shift to lower energies, which can be satisfactorily fitted by Varshni or Bose–Einstein formulas [31.331]. If exciton energies shift to higher energies, it is indicative of the transition from free-exciton to a band-to-band recombination. The survival of the free excitons at room temperature in GaN is supported by the exciton binding energies of around ≈ 25 meV for the A, B, and C excitons [31.298, 332]. This value is comparable to the thermal energy at room temperature ($k_BT = 25.8$ meV). Another fact that supports the presence of excitons at room temperature is that the exciton Bohr radius is very small (≈ 3.5 nm) and a high carrier density ($< 10^{18}$ cm^{-3}) is needed to screen the Coulomb interaction [31.303].

The transient behavior of free-exciton recombination in GaN can be investigated by time-resolved PL spectroscopy, while noting that PL transient of the free excitons is very complicated, partially due to the presence of exciton polaritons (EPs). The intrinsic lifetime of this EP was estimated to be about 300 ps (based on acoustic phonon scattering only) [31.333]. The much shorter lifetimes observed in experiments are most likely due to the nonradiative recombination in GaN. Typical transients are shown in Fig. 31.23, which were obtained for different excitation intensities for a GaN epilayer grown on sapphire. The decay time at 2 K was found to vary between 60 and 115 ps, which strongly indicates a contribution from defect-related nonradiative processes. Thus, the decay in luminescence is described by a combination of fast processes followed by a slower processes with a decay time of about 300 ps, although the free-exciton lifetime was estimated to be about 2 ns at low temperatures, deduced from the free-exciton

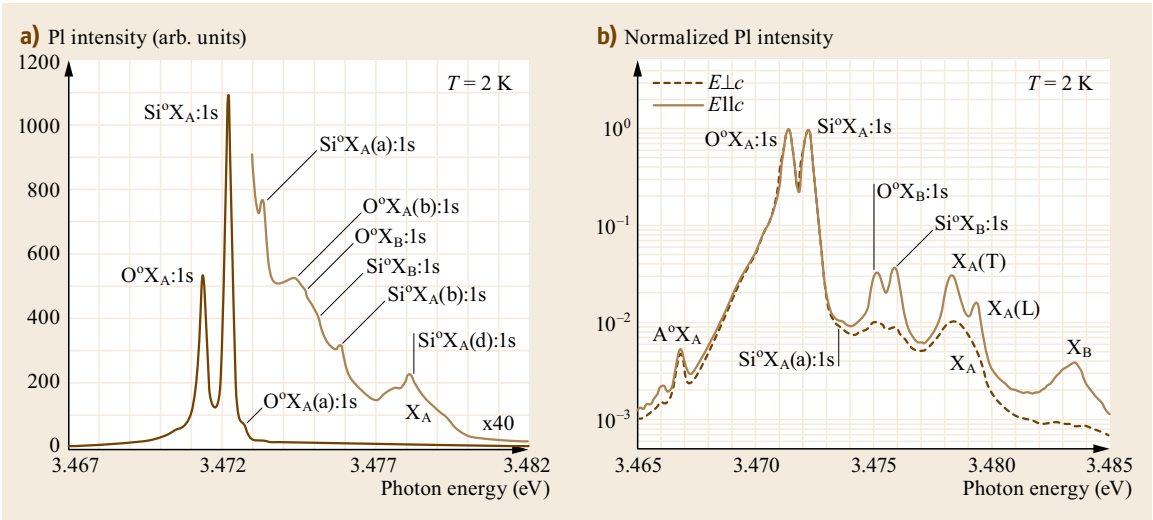


Fig. 31.24a,b PL spectra measured at 2 K for Ga-face bulk GaN for the following configurations: **(a)** α polarization ($k\|c, E\perp c$); **(b)** $\sigma(k\perp c, E\perp c)$ and $\pi(k\perp c, E\|c)$ polarizations. (After [31.303], courtesy of Dr. B. Monemar)

lifetime with two-photo spectroscopy on freestanding HVPE samples [31.334].

Bound Excitons in GaN

Due to the abundance of impurities in GaN, the neutral shallow donor-bound exciton is often dominant in optical spectra. For samples containing acceptors, the acceptor-bound exciton is also observable. Each donor, acceptor, or neutral defect typically is associated with a bound exciton. The recombination of bound excitons typically gives rise to sharp lines, and can easily be identified for each BE. The neutral donor BE is a four-particle complex with two electrons and one hole located around the positive impurity ion. For excited BEs, an electron will be excited into a series of exciton states, where the first excited state shifts away from the ground state by about three-quarters of the donor binding energy. If the hole in the neutral DBE is excited instead of an electron, this state is called a rotational states, which is quite weak in PL spectra, and is difficult to resolve unless selective excitation is used [31.335]. Oxygen and silicon are the two most dominant residual shallow donors in GaN, thus induce a series of oxygen- and silicon-related DBE states in PL spectra. An example of PL spectra for a bulk GaN sample featuring oxygen and silicon-related DBE states is displayed in Fig. 31.24.

The two dominant PL lines at 3.4714 and 3.4723 eV presented in Fig. 31.24 are believed to be due to the principal recombination of an A exciton bound to a neutral O donor on the N site and a neutral Si donor on the Ga site respectively [31.303]. Emission peaks at very similar energies of 3.4718 and 3.4755 eV were also reported in a high-quality MBE-grown GaN layer on a GaN substrate, where the two peaks are also believed to be associated with Si- and O-related shallow donor states [31.336]. Weak excited rotational states of the O and Si DBEs were observed as well at 3.4727 eV (O^0X_A(a):1s) and 3.4735 eV (Si^0X_A(a):1s) respectively. The energy separation between these lines and the correspondingly principal DBE lines are found to be consistent with previously reported values [31.335, 337]. For even the higher energy DBE region, the DBEs related to the B valence band come into the picture, where the major peaks at 3.4755 eV [31.336] or 3.4751 and 3.4759 eV [31.303, 335] were observed. The peaks at 3.4751 and 3.4759 eV were believed to correspond to O^0X_B:1s and Si^0X_B:1s respectively. These two peaks are more pronounced in the PL spectra measured in π ($k\perp c, E\|c$) polarization as demonstrated in Fig. 31.24. For a complete review of the commonly observed and reported peak positions and localization energies of DBE in GaN grown by various growth techniques on different substrates, readers can refer to *Viswanath*'s review [31.338, p. 63]. In addition to the emission lines discussed above, there also exist the so-called two-electron transitions (TETs) of the shallow DBEs in the PL spectra with energies ranging from 3.43 to 3.46 eV [31.303]. The origin of these PL lines is attributed to the neutral donor in the final state of the DBE recombination that remains in an excited donor state [31.339]. Strong O- and Si-related TETs pave the way for determining the associated donor binding energies, which are 33.2 ± 0.4 and 30.4 ± 0.4 meV for O and Si donors respectively. These values are quite close to results reported by *Reshchikov* et al. with binding energies of 32.6 and 28.8 meV for O and Si

donors respectively, as measured in MBE-grown GaN on a freestanding GaN template [31.340]. Similar values were also reported by *Moore* et al. [31.341].

Similar to the transient behavior of free excitons, the transient behavior of DBEs also showed a variety of surprising properties. The radiative decay time τ can be estimated from the so-called giant oscillator strength F_{DBE} as $\tau \approx 1/F_{\mathrm{DBE}}$ [31.342]. Experimentally, the radiative lifetime can be measured from the low-temperature decay rate associated with the DBE PL lines. For a thick HVPE-grown ($\approx 100\,\mu$m) GaN layer on sapphire with dislocation density of $10^7\,\mathrm{cm}^{-2}$, a low-intensity decay time of about 200 ps was observed at 2 K as demonstrated in Fig. 31.25, which is believed to be a typical radiative lifetime for DBEs associated with shallow neutral donors [31.343]. For a thinner layer with higher defect density, a shorter lifetime was observed [31.344, 345], probably due to the excitation transfer to defect states. For a homoepitaxial GaN layer, a fast DBE decay was also observed, due to excitation transfer to point defects [31.288]. Recently, PL decay of a DBE no-phonon line for an HVPE-grown thick GaN layer at low temperatures showed a nonexponential behavior with a fast initial slope (a decay time of about 300 ps), and a much slower tail towards longer times (a decay time of about 1000 ps or higher) [31.303]. It is believed that the initial fast decay time may not be the true decay time of DBEs. Instead, it may be due to other processes, such as surface recombination or optical dispersion in the semiconductor medium with exciton resonances accompanied by resonant light scattering at the neutral donors [31.346]. In order to eliminate the contribution from such processes, the DBE decay time was derived from the transient behavior of the LO replicas of DBE, which should reflect the actual density of BE states. Since the two LO phonon replicas of the O- and Si-related donors are not spectrally resolved, average decay times of about 1.3–1.4 ns were observed for the DBE LO replicas in a 1 mm-thick freestanding HVEP sample with a relatively low donor concentration of $\approx 8\times10^{15}\,\mathrm{cm}^{-3}$, which may represent a close approximation to the radiative decay time of the DBEs [31.303]. For a 300 μm-thick freestanding sample with a donor concentration of $\approx 5\times10^{16}\,\mathrm{cm}^{-3}$, a shorter decay time of ≈ 600 ps was recorded. Since two-electron transitions (TETs) also rely on the population of DBE states, the transient behavior of TETs was also used to estimate the decay time of DBEs, and a similar decay time was observed as compared to the decay time extracted from the transient behavior of LO replicas of the DBEs. An example of the transient behavior is presented in Fig. 31.26.

Acceptor-bound excitons in GaN are not as well understood as DBEs, and a major ABE1 line at 3.466 eV

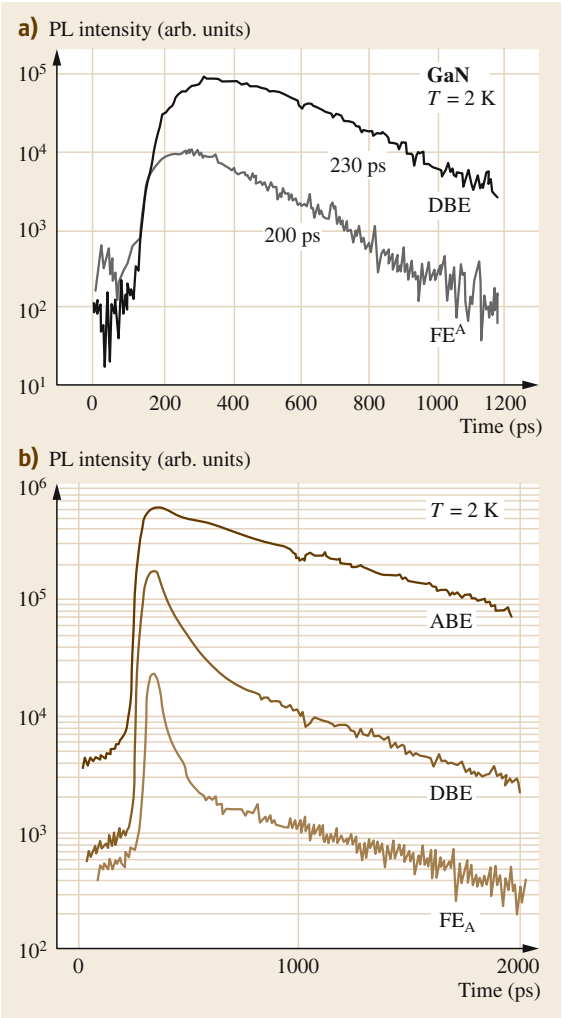


Fig. 31.25 Transient PL **(a)** for a 400 μm-thick GaN layer and **(b)** for a thin strained layer, of which both were grown on a sapphire substrate. (After [31.343])

was observed in strain-free GaN at 2 K [31.347–351], as demonstrated in Fig. 31.27. This peak has been found to be dominant in slightly Mg-doped GaN samples, thus a common assignment of this peak is $\mathrm{Mg_{Ga}}$. This ABE spectrum is stable for n-type or high resistivity materials. However, for p-type (Mg-doped) GaN the emission is clearly unstable and is quenched in a matter of minutes upon photoexcitation, indicating that the acceptor itself may not be stable under these conditions [31.352]. Contradictory evidence comes from the fact that this ABE line disappears in the cathodoluminescence (CL) spectrum under electron irradiation at low temperatures [31.352, 353]. However, one should keep in mind that electron irradiation has been proven to activate the Mg acceptor. Other proposed assignment

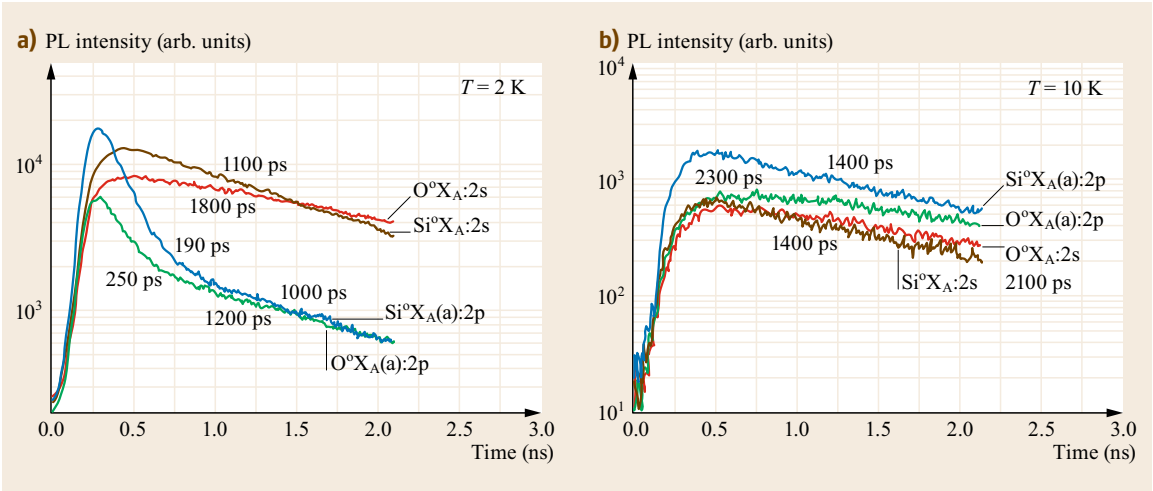


Fig. 31.26a,b Transient PL of TET (two electron transition) related to Si and O donors at **(a)** 2 K and **(b)** 10 K. (After [31.303], courtesy of Dr. B. Monemar)

are: a different acceptor from the substitutional Mg_{Ga}, such as C_N [31.354], and charged donor-bound excitons [31.355–358]. Another broader peak at 3.455 eV (ABE2) has been attributed to Zn based on the fact that this line is close to the position of the dominant ABE observed in Zn-doped samples, if strain-induced shifts are considered [31.359]. However, this still needs further justification. Another tentative assignment for this peak is Mg_{Ga} acceptor due to the fact that the intensity of this broad ABE peak increases with Mg doping [31.360]. This naturally assumes that Mg doping does not facilitate any other process. Recently, both ABE1 and ABE2 lines were proposed to be caused by the Mg doping [31.361], with the ABE2 only observed in highly doped GaN:Mg, which suggest a possible correlation with structural defects, such as basal stacking faults (BSFs), induced by high-Mg doping [31.362]. It was proposed that the A1 acceptor (responsible for the ABE1 emission) will be perturbed by the nearby BSF potential and give rise to the A2 acceptor and its characteristic ABE2 spectrum [31.361]. Decay curves demonstrated the spectral shift of the PL signal from the ABE1 position to ABE2 position, mostly during the first 100 ps, which was interpreted as exciton transfer.

The broader peak at 3.44 eV is probably due to a low-energy acoustic-phonon wing connected to the main peak at 3.455 eV, which is a characteristic for ABEs [31.363, 364]. The decay curves for ABEs in GaN usually show a clean exponential shape as demonstrated in Fig. 31.25b. The radiative lifetime is about 0.7 ns for the shallowest ABE peaked at 3.466 eV, as compared to the much longer decay time of 3.6 ns for the deeper acceptors with an ABE peak in the range of 3.455–3.46 eV [31.348]. The decay time for

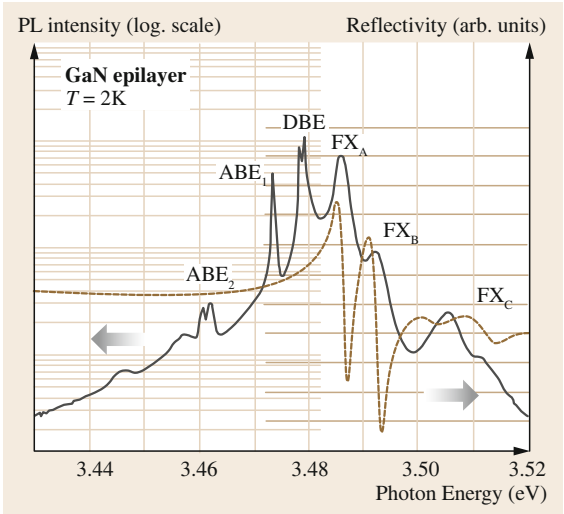


Fig. 31.27 Low-temperature PL spectrum (*solid line*) of a HVPE GaN layer shown on a semilog scale. The reflectivity spectrum (*dashed line*) of the same sample taken at normal incidence geometry is plotted on a linear scale. (After [31.343])

the 3.466 eV ABE peak varies for different samples, presumably due to a nonradiative contribution from excitation transfer effects [31.303].

Donor–Acceptor Transitions in GaN

Due to the various impurities and intrinsic defects, donors and acceptors usually coexist in the same semiconductor material. The electrons bound in the donor states can radiatively recombine with holes bound in the acceptor states, a process termed donor-acceptor

pair (DAP) transition. A commonly observed DAP peak in GaN is located at 3.27 eV, and well-resolved LO-phonon replicas at lower energies were also observed [31.285, 365, 366]. The emission peak is temperature dependent, which is indicative of the change from a DAP signature at low temperatures to a free-to-bound conduction-band-to-acceptor transition at higher temperatures (120 K). This is due to thermal ionization of the shallow donors ($E_D \approx 29$ meV), where an acceptor binding energy of 230 meV was obtained from the temperature-dependent measurements [31.340]. The acceptor responsible for this 3.27 eV DAP is generally believed to be due to the Mg_{Ga} acceptor. There are other interpretations, however, such as an Mg-H complex [31.367], but this interpretation implies that this peak will only be observable in p-type GaN, which is in conflict with the fact that this 3.27 DAP is also prominent in n-type GaN. Another DAP also exists in the blue PL band peaking at about 2.9 eV, which is strong only in MOCVD-grown GaN with high Mg doping. For the same material grown with MBE or HVPE, this blue PL peak does not appear, even though a similar concentration of Mg_{Ga} acceptor would be present. The identity of this blue PL peak is not clear yet, but suggestively it may be related to an H-related deep donor (binding energy 0.37 eV) and the shallow Mg acceptor [31.368, 369].

An interesting feature regarding the ABE as well as 3.27 eV DAP spectra is the long-term instability during prolonged excitation with above bandgap UV light, which was proposed to be due to the instability of the acceptor responsible for the ABE1 and 3.27 eV emission line [31.370]. Additional data suggest that the unstable acceptor might be the Mg-related acceptor as discussed above [31.361]. Short-term instability was also observed, which features a rapid degradation of the PL efficiency in the entire near-bandgap region, indicating a rapid nonradiative process induced by UV excitation. A downshift in energy of the 3.27 eV DAP peak was observed after long-term excitation, which means that the hole occupation of the Mg acceptor decreased [31.371, 372]. This can be explained as that the quasi-Fermi level is raised in the bandgap during the long-term excitation, which can be due to another deeper (nonradiative) defect level. During the recombination process, this deeper defect level is active and can capture photoexcited holes to a metastable state, which decreases the hole occupation in the Mg acceptor [31.361].

Defect-Related Transitions in GaN

Yellow Luminescence (YL). A broad PL band peaking at 2.2 ± 0.1 eV, the so-called yellow emission, has caused enormous debate regarding the nature of this peak for almost 30 years. This PL emission has been systematically observed in undoped or n-GaN [31.373, 374]. The activation energy (thermodynamic transition level) of the YL band was experimentally demonstrated to be 850 meV by fitting the temperature dependence of the YL intensity [31.375]. As studied extensively by *Ogino* and *Aoki* [31.376–379] there seems to be an agreement that this YL can be attributed to transitions from the conduction band or a shallow donor to a deep acceptor. Transition from a deep donor to a shallow acceptor was also proposed to explain the results of magnetic-resonance experiments [31.374]. Early works attributed the YL band to the V_{Ga}-C_{Ga} complex [31.376]. Theoretical calculation from density functional theory suggested that Ga vacancies, both isolated and bound into a complex with oxygen donors, are responsible for the observed YL [31.378, 379], which was supported by positron annihilation experiments, where the YL intensity was found to increase with the concentration of Ga vacancies. Other experimental findings indicated that lattice defects alone do not cause YL, rather its origin should be attributed to some impurities, possibly bound into complexes [31.380]. *Shalish* et al. [31.381] invoked a broad distribution of acceptor-like surface states to account for the YL band. The YL band was also found to shift to lower energies with decreasing energy of the below-gap excitation, which also indicates that the broadening of the YL band may be due to emission from several closely spaced traps [31.382]. More recently, carbon substituting for nitrogen C_N [31.383] and C_N-O_N [31.375] complexes were proposed to be responsible for the YL band by two different groups.

The transient behavior of the YL band has been investigated by several investigators, and a rather long and nonexponential decay in the range of $0.1-100\,\mu s$ at low temperature has been reported by *Hofmann* et al. [31.384] and *Korotkov* et al. [31.385] by using the Thomas–Hopfield model [31.371] of the donor-acceptor pair recombination. Similar work was also demonstrated by *Reshchikov* et al. [31.386] on freestanding high-quality GaN templates grown by HVPE, which concluded that, at temperatures below 40 K, the time decay of the yellow peak is nonexponential and can be explained in the framework of the Thomas–Hopfield model for DAP-type recombination involving shallow donors. For elevated temperatures, the decay becomes exponential with two components, suggesting transitions from the conduction band to two deep acceptors.

31.6.2 Aluminum Nitride

The optical properties of AlN have been investigated in many forms of it, including powders, sintered ce-

ramics, polycrystals, and single-crystal samples. Owing to the availability of large-size single-crystal bulk AlN in recent years, intensive studies on optical properties have been conducted. Because the AlN lattice has a very large affinity to oxygen dissolution, oxygen contamination is difficult to eliminate in AlN. AlN samples contaminated with 1–6% oxygen were found to have two broad emission peaks centered near 3.0 and 4.2 eV [31.387]. *Youngman* and *Harris* [31.388] and *Harris* et al. [31.389] found broad peaks centered at 2.7 and 3.8 eV in large single-crystal AlN with oxygen content of 380 ppm. However, *Strasburg* et al. [31.130] argued that oxygen alone is not able to generate such emission bands by comparing the obtained PL spectra with elemental analysis data obtained from glow discharge mass spectrometry (GDMS), where the emission intensities did not show a direct correlation to the oxygen concentration. GDMS analysis also revealed that in addition to O, Si and C were also present at concentrations that may have a significant influence on the luminescence properties of AlN. An absorption band around 265 nm (≈ 4.7 eV) was found to be more pronounced with the increase of the carbon concentration [31.51], and one possible origin of this absorption band is reported to be due to the substitution of C impurities on the nitrogen site (C_N), which becomes a deep acceptor in AlN [31.390]. Room-temperature PL spectra also showed that a peak at 3.9 eV only becomes prominent as long as the C concentration is much higher than the concentration of other impurities. This suggests a change in the compensation mechanism once the amount of C is not sufficient to act as the main acceptor [31.390], which is consistent with the assignment of C_N to a deep acceptor. However, in another report, the intensity ratios of deep CL bands at 4.6, 3.1 and 3.8 eV to the near-band-edge emission at 11 K were found to increase with V_{Al} concentration, where the V_{Al} concentration was detected from positron annihilation spectroscopy [31.391]. Another proposed origin for the 3.9 eV in the PL spectrum is due to a complex of an aluminum vacancy (V_{Al}) and an oxygen impurity on the N site (V_{Al}-O_N) [31.392]. Another 2.8 eV luminescence peak, whose intensity also increases with increasing carbon doping concentration, was observed even in HVPE-grown samples with relatively low concentrations of O and Si impurities, which reduces the concentration of V_{Al} by increasing the formation energy of V_{Al} [31.379]. One can then conclude that the deep emission around 2.8 eV may be either directly or indirectly related to C impurities [31.51].

The investigation of free-exciton energies has expanded considerably due to the availability of high-quality single-crystalline AlN layers, especially AlN bulk material. The majority of reported free-exciton energies of A are between 6.0–6.05 eV, depending on the strain state of the AlN material, while the free-exciton energies of B and C are difficult to fully resolve, but which would be in the range of 6.2–6.3 eV. A survey of the free-exciton energies is presented in Table 31.8. The free-exciton energy is strongly dependent on the biaxial strain of the AlN lattice, where the free-exciton energy is blueshifted in AlN/sapphire due to compressive strain and redshifted in AlN/SiC and AlN/Si due to tensile strain. A linear relationship between the free-exciton transition energy peak position and in-plane stress was obtained [31.393] with a linearity coefficient of 45 meV/GaP, which is about 88% higher than that in GaN (24 meV/GPa) [31.394]. The dependence of the free-exciton energies on the strain is demonstrated in Fig. 31.28 with four data points, including AlN/sapphire, AlN/AlN bulk, AlN/SiC, and AlN/Si. Based on the free-exciton energy of A and its first excited state, the AlN bandgap can be determined, and the reported values are 6.077 eV [31.395], 6.093(8) eV [31.396], 6.094 eV [31.397, 398], and 6.096 eV for unstrained bulk AlN at low temperatures [31.399]. The binding energy of free excitons can also be determined from the same technique, with reported values of 57 meV [31.400], 48 meV [31.395], 63 meV [31.401], 52.4 meV [31.397], 55 meV [31.402] and 51 meV [31.403] for FX(A). These reported binding energies are much larger than the thermal energy at room temperature, thus enabling one to clearly observe free-exciton luminescence at room temperature [31.404]. Based on a quasicubic model developed for the wurtzite crystal as well as other reported exper-

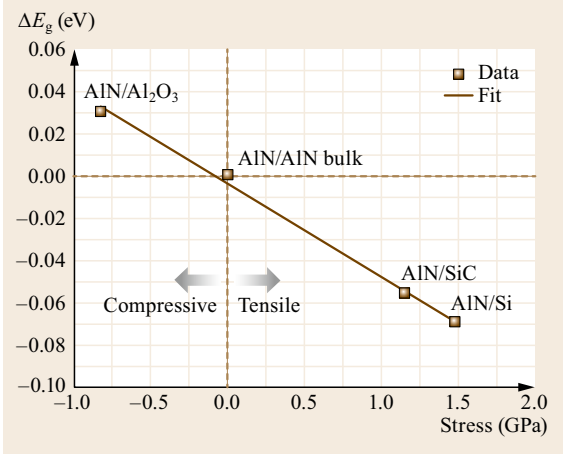


Fig. 31.28 Stress-induced shift of free-exciton energy in AlN epilayers as a function of the in-plane strain. The *solid line* is the least-square linear fit to the experimental data, where a value of 45 meV/GPa is obtained for the linear coefficient of stress-induced bandgap shift. (After [31.393])

imental exciton energy values, the AlN spin-orbit splitting and the crystal-field splitting were estimated to be $\delta = 36\,\text{meV}$ and $\Delta = 225\,\text{meV}$ respectively [31.395]. The obtained value for the crystal-field splitting is close to the value of 230 meV in unstrained AlN as reported earlier [31.399].

Due to the large free-exciton binding energy and strong oscillator strength, the excitons can stably exist at high excitation densities, enabling the study of the interaction between two or more excitons. This unique property can be characterized by the formation of biexcitons (excitonic molecules) and the inelastic scattering of excitons. Exciton-exciton scattering processes occur when, at high enough exciton densities, inelastic collisions of excitons occur. Following scattering, one exciton is scattered into a higher excited state while another decays with equal-energy loss. Excitons can not only be scattered into $n = 2$ state but also scattered to $n = 3 \ldots \infty$, where the number n is usually given as an index to distinguish the different scattering events therefore leading to P_2 to P_∞ bands. Note that under high excitation densities at room temperature, the electron-hole plasma (EHP) has been found to be the dominant recombination process in AlN [31.402]. Biexciton emission (also called the M band) has been observed in ELO-grown *a*-plane AlN epitaxial layers with an energy separation between free-exciton luminescence and biexciton luminescence of 19 meV, which is also the binding energy of biexcitons [31.400]. The PL plots of exciton and biexciton emissions at 7 K at different excitation densities are displayed in Fig. 31.29. By integrating the PL intensities of both the free-exciton line and the biexciton line for different excitation power densities, it was found out that the free-exciton line intensity grows almost linearly with the excitation power density (a power factor of 1) while the biexciton line intensity grows superlinearly with a power factor of around 1.5. This indicates that the biexciton density does not vary as the squares of the free-exciton density, which may be due to a relatively short recombination time of excitons and biexcitons in direct-bandgap semiconductors [31.405]. In the instant case, it may imply that the biexcitons have a shorter radiative lifetime than that of free excitons. Biexciton emission was also observed at a different energy, some 27 meV below the free-exciton emission in AlN layers grown on SiC and sapphire substrates [31.406]. In addition, another band was identified at 36 meV below the free A-exciton transition and was ascribed to the exciton-exciton scattering (P_2 band), which showed a power factor of 1.66 instead of 1.54 for the biexciton emission in the same report [31.406].

A few DBE emission lines have been reported as shown in Fig. 31.30, with D^0X occurring at 9.8, 13.3, 19.0, 22.5, and 28.5 meV below the free exciton X_A [31.396] in AlN layers grown homoepitaxially on AlN substrates, which were in turn grown by physi-

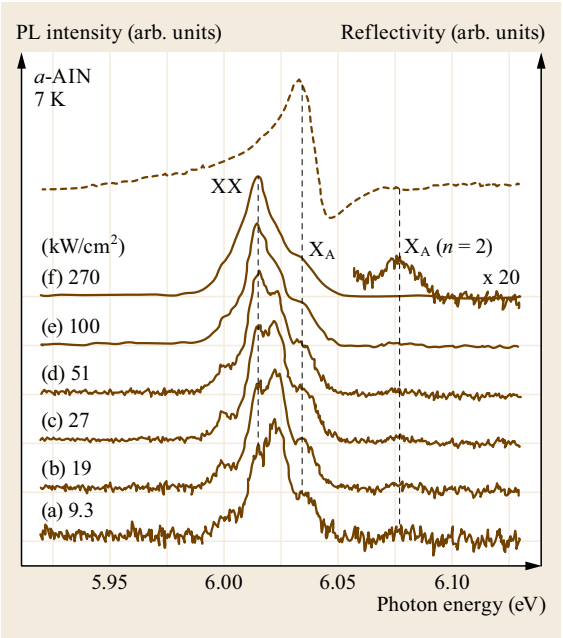


Fig. 31.29 Excitonic photoluminescence spectra (*solid lines*) at 7 K taken from an *a*-plane AlN epitaxial layer under different excitation-power densities ranging from 9.3 to 270 kW/cm^2. Each spectrum is normalized at the peak height. An optical reflectance spectrum is also demonstrated in the upper part by the *dashed line*. The symbol XX represents the biexciton. (After [31.400], courtesy of Yamada))

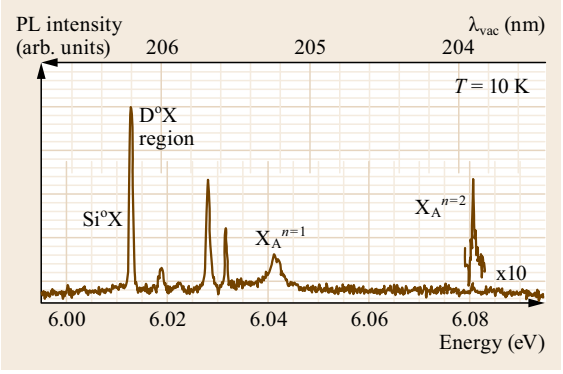


Fig. 31.30 PL emission lines of a nominally undoped AlN homoepitaxial layer with a few donor-bound exciton lines (D^0X) besides the free-exciton emission (X_A) and its first excited state with $n = 2$. The *leftmost line* is ascribed to the excitons bound to silicon (Si^0X) found at 28.5 meV below the X_A emission. (After [31.396])

Table 31.8 Reported free-exciton (FX) energies in AlN

FX(A)	FX(B)	FX(C)	Orientation	Substrate	Strain	Growth	Technique	Reference
6.0430			*c*-plane	AlN	Homoepitaxial	MOVPE	PL	[31.407]
6.041			*c*-plane	AlN	Homoepitaxial	PVT	PL	[31.396]
6.0414			*c*-plane	AlN	Homoepitaxial	MOCVD	PL	[31.397]
5.958	6.216	6.229	*c*-plane	Silicon	Tensile	MOCVD	PL	[31.408]
6.074	6.273	6.283	*c*-plane	Sapphire	Compressive	MOCVD	PL	[31.408]
6.109	6.285	6.295	*c*-plane	SiC	Compressive	PA-MBE	PL	[31.408]
6.0414			*c*-plane	AlN	Homoepitaxial	MOCVD	PL	[31.409]
6.025	6.243	6.257	*c*-plane	AlN	Homoepitaxial	MOCVD	Reflectance	[31.399]
6.1383	6.277	6.304	*c*-plane	Sapphire	Compressive	LP-MOVPE	CL	[31.403]
6.0346			*a*-plane	AlN template	ELO	MOVPE	PL	[31.400]
6.060			*c*-plane	Sapphire	Compressive	MOCVD	PL	[31.393]
5.974			*c*-plane	SiC	Tensile	MOCVD	PL	[31.393]
5.960			*c*-plane	Silicon	Tensile	MOCVD	PL	[31.393]
6.029	6.243	6.268	*c*-plane	Sapphire	Compressive	LP-MOVPE	Reflectance	[31.403]
6.023	6.243	6.268	*a*-plane	Bulk AlN	Bulk	Sublimation	Reflectance	[31.395]
6.023	6.036		*c*-plane	Bulk AlN	Homoepitaxial	OMVPE	CL	[31.401]
6.025	6.243	6.257	*m*-plane	Bulk AlN	Bulk	Sublimation	Reflectance	[31.399]

cal vapor transport. The tentative assignment of these emission lines to DBEs is based on the agreement with earlier reports, the temperature-dependent behavior of the emission bands, and their very narrow line widths (500 μV for the dominant line of 6.0125 eV in Fig. 31.30). The assignment to ABEs was discounted because, in analogy to GaN, the binding energies of common acceptors [31.366] are expected to be larger than those of donors [31.341]. With increasing silicon doping concentration from 1×10^{17} to $1\times10^{19}\,\mathrm{cm}^{-3}$, a clear increase in the intensity was observed for one single D^0X band with 28.5 meV exciton localization energy (located at 6.0125 eV) relative to the intensity of the free-exciton emission [31.396]. This assignment is in agreement with the available data for silicon in AlN, where the localization energy has been reported to be 24 meV [31.410], 22–29 meV [31.411], and 26 meV [31.412]. Based on the PL spectra, the donor binding energy of silicon in AlN was estimated to be 133 meV. The line width of this Si^0X line was broadened due to the increased Si doping, possibly due to the band filling as proposed earlier by *Yoshikawa* et al. [31.413] for silicon-doped GaN, and local potential fluctuations according to $\mathrm{FWHM} = E_0 + a\sqrt{N_{Si}}$ as reported by *Nam* et al. [31.412]. Additionally, Si doping also gave rise to a blue luminescence band at around 400 nm, which may be due to secondary defect creation processes or charge transfer due to higher silicon doping [31.396]. However, another report ascribed the DBE line with localization energy of 22.4 meV to the Si^0X line, most prominent in the sample having the highest silicon concentration, while other DBE lines remained unidentified and may be due to oxygen donors or intrinsic donors [31.397]. An ABE emission line due to an Mg acceptor was claimed to be observed in Mg-doped AlN epilayers at 6.02 eV at 10 K, which is about 40 meV below the free-exciton transition in undoped AlN epilayers [31.414]. The intensity of this ABE line is rather weak, which is about 40 times smaller than that of the free-exciton transition in undoped AlN epilayers.

Time-resolved PL measurement in 1 μm-thick AlN epilayers grown by MOCVD on sapphire substrates revealed that the recombination lifetimes were around 80 ps for DBE and 50 ps for free excitons [31.415]. For Mg-related ABE, the decay kinetics can be described by a single exponential with a decay lifetime constant of about 130 ps at 10 K [31.414]. In another report, time-resolved PL and CL were used to investigate the decay properties in AlN, and the excitonic radiative lifetime (τ_R) at 7 K was measured to be 10 ps, which increased up to 530 ps with increasing amount of impurities and V_{Al} concentrations, irrespective of the dislocation density [31.416]. It was found that radiative lifetime of excitons bound to neutral donors and *e-h* pairs bound in impurity-induced band-tails is much longer than for the intrinsic free excitons.

The temperature dependence of the optical bandgap of AlN has been investigated by a few groups with consistent results. As reported by *Guo* and *Yoshida* [31.417], the bandgap energy increases linearly from room temperature down to about 150 K with a temperature coefficient of 0.55 meV/K. Similar temperature behavior was reported by *Kuokstis* et al. [31.418] for single-crystal *c*-plane AlN with a temperature coefficient of 0.53 meV/K between 150 and 300 K. These two reports are close to each other and also close to the value of 0.51 meV/K reported by *Tang* et al. [31.419]. For temperatures below 150 K, it

was found that for all the measurements a new structure starts to form, where the bandgap does not change linearly with temperature and has a very small temperature coefficient, which is unusual for semiconductors.

31.6.3 Indium Nitride

InN is not as well understood compared to other group III nitrides. Due to the large background electron concentration as well as the relatively low crystalline quality, the excitonic features, neither in the absorption spectra nor in the band-edge photoluminescence spectra, have been observed. Many early efforts determined the bandgap to be in the range of 1.7–2.07 eV [31.421–423]. One widely cited bandgap value for InN was 1.89 eV obtained by *Transley* and *Foley* [31.424]. This value was also often used as an end-point value to interpret experimentally measured composition dependence of the bandgap of InN alloys [31.425]. With improved growth techniques for InN, a strong luminescence transition at energies around 1 eV appeared [31.426–428], which led to a sudden revision of the bandgap down to about 0.7 eV [31.429]. The sudden revision for the bandgap of InN came from the observation of an absorption edge at about 0.7 eV at low temperatures for moderately n-type samples (about $10^{18}\,\mathrm{cm}^{-3}$) [31.430, 431] and the value blueshifts for higher doping [31.427, 432, 433]. *Thakur* et al. [31.434] investigated the optical-absorption spectra of InN thin films with electron densities varying from $\approx 10^{17}$ to $\approx 10^{21}\,\mathrm{cm}^{-3}$. The optical-absorption edge was found to increase from 0.61 to 1.90 eV as the electron density of the film is increased from low to high density. Shifting of the optical-absorption edge is due to a corresponding up-shift of the Fermi level in the conduction band (the so-called Burstein–Moss shift). By subtracting the influence of self-energies and chemical potential from the optical-absorption edge energy, the intrinsic bandgap values were estimated to be ≈ 0.59 eV for the film with lowest carrier density, and ≈ 0.60 to ≈ 0.68 eV for the film with highest carrier density, where the dispersion was attributed to the variation in the electron effective mass and static dielectric constant. However, care must be taken to interpret the carrier density values obtained from the Hall measurements due to the presence of surface electron accumulation for InN [31.435–437] and due to a gradient in defect properties across the film [31.438].

Due to the fact that, in many highly doped samples, the main absorption edge occurs at much higher energies while the PL peak remains at about 0.7–0.8 eV [31.432, 439, 440], an alternative model of this absorption edge as being of extrinsic origin has been suggested in the literature [31.440–442], which ascribed the absorption edge to a deep level to conduction band transition. The value of the absorption coefficient at about 0.5 eV above the absorption edge was measured experimentally to be about $3\times10^{4}\,\mathrm{cm}^{-1}$ (within a factor of two) [31.443], which is a typical value for a direct band-to-band transition. In order for the deep level model to be consistent with the large absorption coefficient measured, the deep donor concentration was estimated to be close to $10^{20}\,\mathrm{cm}^{-3}$ [31.443]. In addition, this concentration has to be the same within

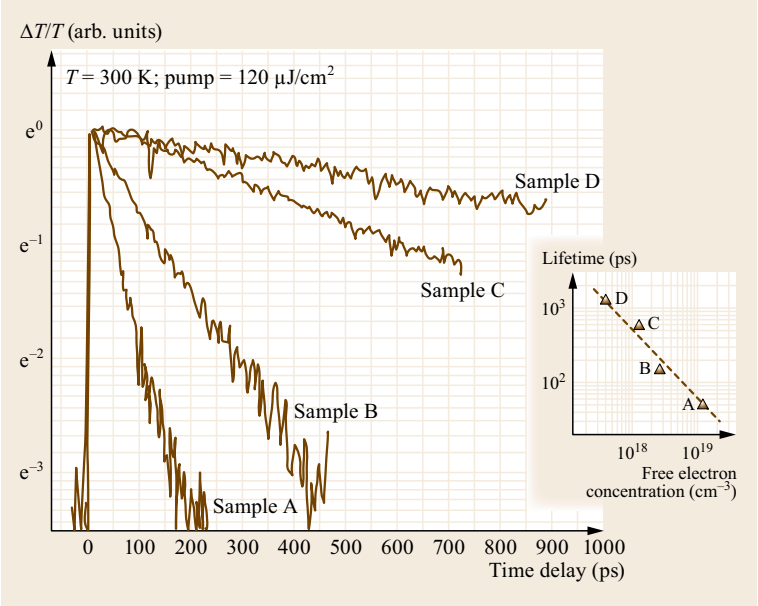


Fig. 31.31 Strength of the differential transmission signal versus time delay under a pump fluence of $120\,\mu\mathrm{J/cm}^2$ at room temperature. The *inset* shows the carrier lifetime as a function of free-electron concentration. (After [31.420], courtesy of Dr. A.N. Cartwright)

a factor of two in all samples studied, which appears implausible and no such deep level could be identified from SIMS data in the highest quality samples studied [31.270, 444]. Besides the optical measurement of absorption edge and direct band nature, the PL signal was found to depend linearly on the excitation laser intensity over a wide intensity range from 10 to 200 mW, indicating that the observed nonsaturable PL peak is due to the emission associated with the direct band-to-band recombination rather than the band-to-defect (or impurity) deep emission [31.445].

Ultrafast differential transmission measurements have been conducted on a series of InN epilayers to determine the carrier lifetimes [31.420], where a fast initial hot carrier cooling followed by a slow recombination process after pulse excitation was observed. The carrier lifetime was measured to be ranging from 48 ps to 1.3 ns, and an inverse proportionality between the carrier lifetime and the free-electron concentration was observed as demonstrated in Fig. 31.31. The ultrafast recombination was assigned to be due to nonradiative recombination related to defects or impurities, which was confirmed by the fact that the defect-related nonradiative lifetime, depending on the defect density, is independent of photogenerated carrier density. The carrier lifetimes were measured as a function of photogenerated carrier density by varying the pump fluence from 60 to 480 $\mu J/cm^2$, the decay curve was observed to remain as a single exponential and there was no observable change in the carrier lifetime.

The temperature dependence of the bandgap of InN has been studied by a few groups [31.421, 446], where a bandgap temperature coefficient of $1.3-1.8\times 10^{-4}$ eV/K was demonstrated near room temperature by using $E_g\ (300\,\mathrm{K}) = 1.89$ eV. For more recent InN samples with a ≈ 0.7 eV bandgap, a small blueshift (nearly linear at 0.1 meV/K) was observed with increasing temperature in the peak energy position [31.430]; although redshift of the peak energy position was also observed with increasing temperature [31.447]. The PL peak broadened with increasing temperature, which may indicate that longitudinal-acoustic phonon scattering dominates the emission broadening process in the InN samples studied [31.447].

31.7 Properties of Nitride Alloys

Many important applications of GaN-based devices require alloying GaN with other group III nitride materials to form, for example, AlGaN, InGaN, AlInN, and even AlInGaN. Ternary alloys of wurtzite polytypes of GaN, AlN, and InN have been obtained continuously whose direct bandgap ranges from ≈ 0.7 eV for InN to 6 eV for AlN. Many material as well as device properties, such as bandgap, effective mass of electrons and holes, and the dielectric constant, depend on the alloy composition. The lattice constants $a(x)$ and $c(x)$ of the wurtzite group III nitride alloys $A_xB_{1-x}N$ ($Al_xGa_{1-x}N$, $In_xGa_{1-x}N$ and $Al_xIn_{1-x}N$) are predicted to follow composition-weighted averages between the end-binary compounds AN and BN (Vegard's law) [31.55], which was confirmed within 2% error over the whole range of possible compositions between theoretical calculations and experimental measurements by a combination of techniques, such as HRXRD, Rutherford and elastic recoil detection analysis [31.63, 448, 449]. However, for pseudomorphic Al-rich $Al_xIn_{1-x}N$ films, the large present strain in the film was found to cause the dispersion between the In contents determined by XRD and RBS, suggesting a strain-driven deviation of the lattice parameters from the Vegard's rule in this case [31.450]. In fact, many materials' properties have been shown to deviate from Vegard's rule, such as the direct bandgap, polarization charge, and so on. The optical bandgap of ternary alloys can be predicted by the following empirical expression

$$E_g(A_xB_{1-x}N) = xE_g(AN) + (1-x)E_g(BN) - x(1-x)b\,, \tag{31.17}$$

where $E_g(AN)$ and $E_g(BN)$ are the optical bandgaps of material AN and BN respectively, and x and b are the AN molar fraction and bowing parameter respectively. For the case of $Al_xGa_{1-x}N$, this empirical equation has been determined to be [31.451]

$$E^g_{AlGaN}(x) = [6.0\,x + 3.42\,(1-x) - 1.0\,x(1-x)]\ \mathrm{eV}\,. \tag{31.18}$$

The above equation can be obtained by simultaneously determining the bandgap and alloy composition precisely. The alloy composition can be obtained from XRD or Rutherford back-scattering (RBS) or both. The bandgap can be determined by optical reflectance and absorption measurements. A plot of the bandgap versus the Al composition is given in Fig. 31.32, where a bowing parameter of 1.0 eV was obtained with a least-square fitting of the data to (31.17). The XRD peak positions as well as the peaks in optical measurement

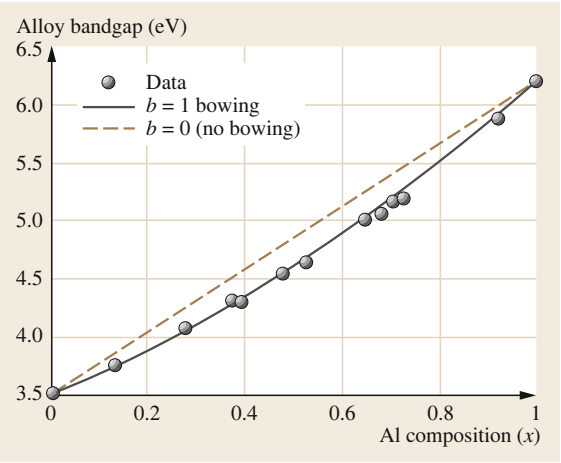


Fig. 31.32 Experimental data for the energy bandgap of AlGaN plotted as a function of Al composition in the whole Al composition range (*solid circle*). The *solid line* is a least-square fit, which gives a bowing parameter of $b = 1.0\,\text{eV}$. The *dashed line* represents the case of zero bowing parameter. (After [31.451])

techniques can be significantly affected by the material crystalline quality and purity, and thus can potentially influence the bowing parameter obtained. Temperature- and composition-dependent bandgaps of $Al_xGa_{1-x}N$ alloys were investigated in the temperature range between 10 and 800 K, where the temperature dependence of the bandgaps at each different composition can be fitted to the Varshini equation yielding the Varshini coefficients [31.452].

The optical properties of Ga-rich $Al_xGa_{1-x}N$ ($x =$ 0.019, 0.038, 0.057, 0.077, and 0.092) alloys in the form of epilayers have been investigated by temperature-dependent PL and time-resolved PL spectroscopy [31.453]. Thermal quenching was observed with two activation energies, which were attributed to the delocalization of excitons and thermal dissociation of excitons. The localization energy increased as the 1.7th power of the PL line width, which reflected a broadening of the density of localized exciton states. Additionally, the luminescence decay of the localized excitons for the five epilayers studied became longer with decreasing emission energy, indicating that the decay of excitons is caused not only by radiative recombination, but also by transfer to lower energy states.

As far as the electrical and doping properties are concerned, there exist several reports of n-type doping of AlGaN with various Al compositions. Hall measurement for n-$Al_{0.09}Ga_{0.91}N$ demonstrated a carrier concentration of $5\times10^{18}\,\text{cm}^{-3}$ and a mobility of $35\,\text{cm}^2/(\text{V s})$ at room temperature [31.454]. A 1.8 μm-thick Si-doped $Al_{0.3}Ga_{0.7}N$ was grown on sapphire by using a superlattice buffer layer that demonstrated a mobility of $80\,\text{cm}^2/(\text{V s})$ and a carrier concentration of $3\times10^{18}\,\text{cm}^{-3}$ at room temperature [31.455]. Crack-free $Al_{0.75}Ga_{0.25}N$ layers with electron concentrations as high as $1\times10^{18}\,\text{cm}^{-3}$ were also demonstrated with Hall mobility of about $50\,\text{cm}^2/(\text{V s})$ at room temperature [31.456]. Highly conductive n-type $Al_xGa_{1-x}N$ ($x = 0.7$) films have been achieved with resistivity as low as 8 mΩ cm (electron concentration $n = 2.6\times10^{19}\,\text{cm}^{-3}$ and mobility $\mu = 30\,\text{cm}^2/(\text{V s})$) for a Si-doping concentration of $3.5\times10^{19}\,\text{cm}^{-3}$ [31.457]. The resistivity was found to increase by one order of magnitude when the Al composition increased by 8%, which is due to the deepening of the Si donor energy level with increasing Al composition, and possibly increasing electron trap density. The Si donor activation energy was estimated to vary from 23 to 180 meV as Al composition was increased from 0.7 to 1.0 [31.457]. A similar trend was also observed that demonstrated that the activation energy of the Si donor increases linearly from 0.02 eV in GaN to 0.32 eV in AlN [31.458]. It should be noted that 23 meV is smaller than the well-established binding energy for dilute GaN, which means that the above-mentioned values represent somewhat screened figures. p-Type doping of the AlGaN alloy was achieved by Mg doping as in the case of GaN. The hole mobility in Mg-doped p-$Al_{0.08}Ga_{0.92}N$ was found to decrease with increasing temperature, reaching a value of about $9\,\text{cm}^2/(\text{V s})$ for a doping density of $1.48\times10^{19}\,\text{cm}^{-3}$, where the low mobility was ascribed to a high carrier concentration and the intergrain scattering present in the samples [31.459]. The Mg doping and resistivity correlation in AlGaN was also investigated [31.460]. In this particular study, the undoped AlGaN layer is highly resistive with resistivity of larger than $10^7\,\Omega\,\text{cm}$, which decreased to $10^3\,\Omega\,\text{cm}$ after doping to a Mg chemical concentration of $2\times10^{19}\,\text{cm}^{-3}$. When the Mg level reached $3.5\times10^{19}\,\text{cm}^{-3}$, the resistivity reduced to $10\,\Omega\,\text{cm}$ with corresponding hole concentration of $2.2\times10^{17}\,\text{cm}^{-3}$ and mobility of $2.7\,\text{cm}^2/(\text{V s})$. Resistivity values increased by several orders of magnitude for Mg concentrations beyond mid-$10^{19}\,\text{cm}^{-3}$ due to generation of structural defects, such as inversion domains and stacking faults, caused by Mg in the lattice. These issues will certainly influence the performance of devices, such as lasers, which depend critically on the overall device series resistance including the p-type AlGaN material. Another issue that needs to be considered for optimum device performance is the thermal conductivity of the $Al_xGa_{1-x}N$ alloy, which demonstrated strong reduction of the thermal conductivity due to even a small portion of alien atoms in host materials [31.461]. The thermal conductivity decreases with increasing Al composition and reaches the minimum

at $x = 0.5$; however, the largest reduction of the thermal conductivity took place when either the Al or Ga composition fell in the band from $x = 1.0$ to 0.9, indicating that profound phonon scattering was induced by even a small portion of alien atoms in host materials. The reduction was observed over a wide temperature range from about 10 to 800 K, but was more sensitive at temperatures higher than 200 K, suggesting that the point-defect and phonon-phonon scattering are the dominant phonon relaxation processes.

The InGaN alloy is used mostly for quantum wells in the active regions of LEDs and lasers, which can emit in the violet or blue wavelength range but can be extended to green. Although high efficiency blue and green LEDs utilizing InGaN active layers are commercially available, this technology might suffer from the phase segregation and other inhomogeneities due to the large disparity between Ga and In, making the determination of the bandgap of InGaN versus composition difficult. Extensive theoretical and experimental studies have been carried out to determine the value of the bowing parameter, which requires the precise information of strain, phase separation and inhomogeneities in InGaN grown by various techniques and on various substrates. Bowing parameter values ranging from 1 to 6 eV have been reported, a summary of which can be found in [31.462]. Considering using either ≈ 1.9 or ≈ 0.8 eV bandgap for InN to fit the experimental results, two bowing factors can be obtained as demonstrated in Fig. 31.33 with bowing parameter of $b = 1.43$ eV for the 0.77 eV bandgap for InN and $b = 2.63$ eV for the 1.9 eV bandgap for InN.

The incorporation of indium in growing InGaN alloys was found to increase with decreasing temperature [31.463–465]. For samples grown at high temperature ($T \geq 750\,^{\circ}$C), prominent near-band-edge emission was observed, while for samples grown at low temperatures ($T \geq 700\,^{\circ}$C), a deep level or impurity transition was found to be dominant [31.465], which can significantly deteriorate the quality of the InGaN material. *Kumakura* et al. [31.466] investigated the electrical properties of Mg-doped InGaN grown by MOCVD with an indium composition smaller than 0.2. The activation energy E_A of the Mg acceptors in InGaN was found to decrease with the increase of the indium content, leading to higher hole concentration at room temperature for higher indium composition within the studied mole fraction range ($x > 0.2$). This phenomenon is consistent with other group III nitride semiconductors, which demonstrated that activation energy increases with energy gap [31.459, 467]. Another report also investigated the p-type doping in $In_xGa_{1-x}N$ alloys grown by MOCVD on semi-insulating *c*-GaN/sapphire templates [31.468]. p-Type layers were achieved for indium composition up to 0.35. The hole concentration was found to continuously increase from 2×10^{17} cm^{-3} for $x = 0$ (GaN) to 5×10^{18} cm^{-3} for $x = 0.22$, while the hole mobility was found to decrease from 15 to 1.8 cm^2/(V s) as x increased from 0 to 0.35. The reduction in hole concentration in p-type $In_xGa_{1-x}N$ with $x > 0.22$ was attributed to the effect of hole compensation by background electrons, which may be due to the defects generated during low-temperature growth, a condition necessary for growing higher indium composition InGaN. The presence of high background electron concentration, most likely due to the low growth temperature, is suggested as the main obstacle for obtaining p-type conductivity in InGaN with indium composition larger than 0.35. The Mg-acceptor level was estimated

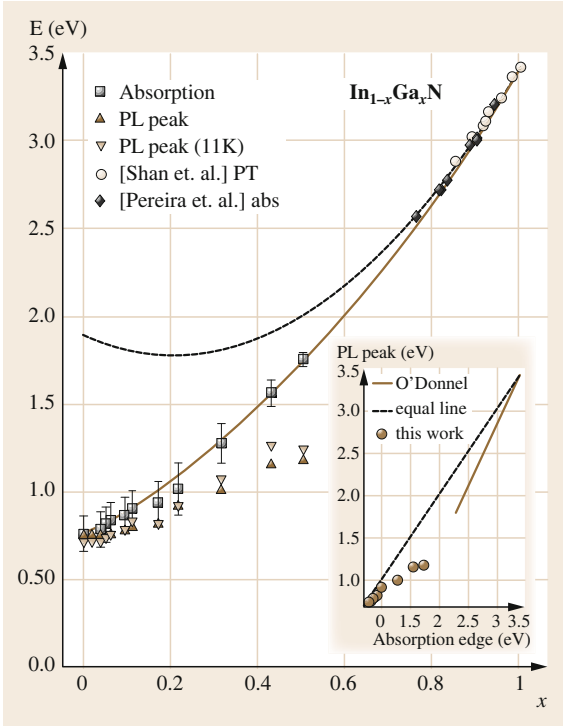


Fig. 31.33 PL peak energy and bandgap determined by optical absorption as a function of composition, including the previously reported data for the Ga-rich end of the ternary, which is not affected by the large InN bandgap previously used. The *solid curve* shows the fit to the bandgap energies using an InN bandgap of 0.77 eV and a bowing parameter of $b = 1.43$ eV. The *dashed line* is the fit to the bandgap energies on the Ga-rich side assuming a bandgap of 1.9 eV for InN. The *inset* shows the PL peak energy plotted against absorption edge energy, where the *solid line* is a least-square fit to experimental data on the Ga-rich side and the *dashed straight line* shows the relationship when the Stokes shift is zero. (After [31.428])

to be about 43 meV for Mg-doped $In_{0.35}Ga_{0.65}N$ alloys, which is roughly four times smaller than that in Mg-doped GaN [31.468]. A room-temperature resistivity as low as 0.4 Ω cm was obtained with a hole concentration of $\approx 5\times10^{18}\,cm^{-3}$ and hole mobility of $\approx 3\,cm^2/(V\,s)$ in Mg-doped $In_{0.22}Ga_{0.78}N$. The PL intensity associated with the Mg-related emission line was found to decrease exponentially with indium composition, which may be related to the incorporation of impurities. As stated above, higher temperature is preferred for growing high-quality InGaN while lower growth temperature is required for growing higher indium composition InGaN; thus it may be concluded that the major challenge for obtaining high-quality InGaN is to find a compromise value for the growth temperature.

The $In_{1-x}Al_xN$ alloy is also an important compound that can be grown lattice-matched to GaN, low-fraction AlGaN and InGaN. Due to their large spontaneous polarization, lattice-matched InAlN/GaN-based HFETs achieve significantly larger two-dimensional electrogas (2DEG) densities and consequently large current densities [31.469]. Owing to the high-index-contrast as well as possible lattice-matching between AlInN and GaN, high-quality AlInN/GaN multipair distributed Bragg reflectors have been achieved [31.470]. The bandgap of AlInN versus the Al composition has been investigated intensively; however, no consensus has been achieved for the bowing parameter due to complexities associated with the precise determination of the composition in AlInN alloys. By adopting a bandgap of ≈ 1.95 eV for InN, which is widely believed to be incorrect, a bowing parameter of 5.4 eV was obtained [31.471]. As mentioned earlier, precise information on the bandgap and alloy composition should be obtained for a fit to determine the bowing parameter. By considering the complications of the band structure of In-rich InAlN, including the Burstein–Moss shift, the nonparabolic conduction band, and the softening of the absorption edge, bandgaps of $In_{1-x}Al_xN$ with $x = 0.017–0.60$ were extracted from optical absorption data [31.472]. As demonstrated in Fig. 31.34, with 0.7 eV for the bandgap of InN, the bowing parameter was calculated to be 4.7 eV, which is reasonably close to the value of 4.96 eV obtained in [31.473]. Also noted is that the value of the bowing parameter is found to be sensitive to the bandgap used for AlN as well as whether the composition is determined from RBS or XRD measurements as a divergence was found between these two measurement techniques [31.450].

Similar to the case of InGaN, the indium incorporation in AlInN strongly depends on the growth temperature [31.474]. In-rich InAlN requires lower growth temperatures; however, lower growth temperatures reduce the AlInN crystalline quality and deteriorate the surface morphology. Transport data for InAlN are extremely scarce. The carrier concentration and the mobility of $In_{1-x}Al_xN$ for $x = 0.04$ were reported as $2\times10^{20}\,cm^{-3}$ and $35\,cm^2/(V\,s)$ respectively, and $8\times10^{19}\,cm^{-3}$ and $2\,cm^2/(V\,s)$ for $x = 0.25$ respectively [31.475]. Room-temperature resistivity and Hall measurements were conducted on In-rich and Al-rich AlInN layers respectively [31.476]. For In-rich AlInN with indium composition of 0.55 and 0.67, the electron concentration n is around 2×10^{19} and $1\times10^{20}\,cm^{-3}$ with associated mobilities around 5 and $7\,cm^2/(V\,s)$ respectively. The Al-rich films showed much higher resistivity, namely $\rho \approx 10^4$ and 10^8 Ω cm for Al composition of 0.66 and 0.72 respectively. These results are consistent with the trend that the mobility decreases substantially with an increasing Al mole fraction and increases with increasing In mole fraction, a close analogy to the parent compounds AlN and InN.

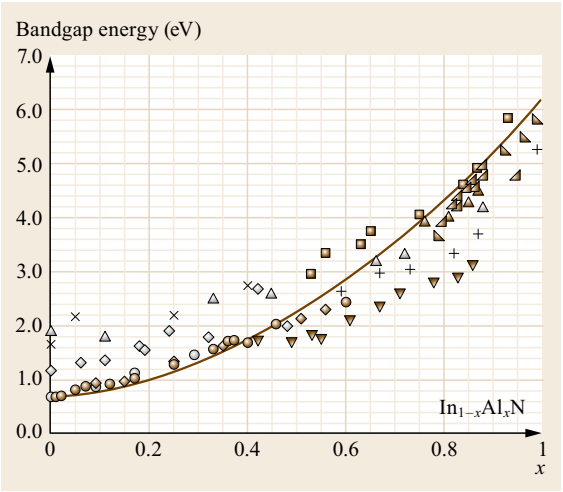


Fig. 31.34 Bandgap energies reported in the literature. A bowing parameter of 4.7 eV was obtained from the fitted *solid curve* with data. (After [31.472], courtesy of Dr. W. Walukiewicz)

By alloying InN with GaN and AlN, the bandgap of the resulting alloy(s) can be expand from the bandgap of InN to that of AlN, which is from 1.9 eV (or ≈ 0.7 eV if we use the recently determined InN bandgap) to a value of 6.0 eV. The benefits of utilizing this quaternary alloy are that: (i) the bandgap of this quaternary can be changed while keeping the lattice constant matched to GaN [31.477, 478]; (ii) an independent control over interface polarization charges and bandgaps is possible, which can result in more uniform carrier distribution among quantum wells while maintaining sufficient electron confinement [31.479, 480]. The InGaN/GaN-based multiquantum wells have been extensively utilized for LED lighting; however, due to the

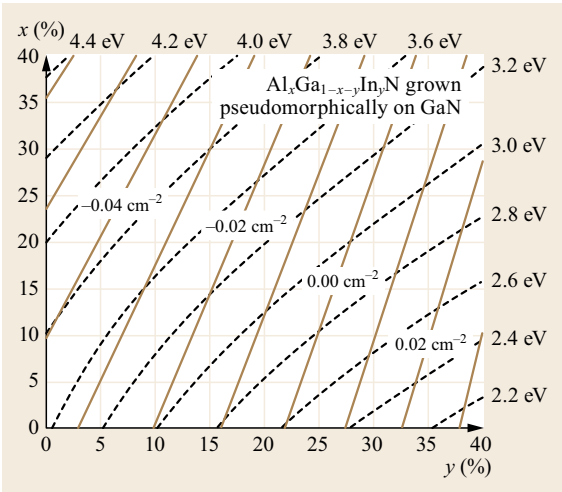


Fig. 31.35 Ga-face polarization charge (*brown*) and bandgap (*black dashed*) contours of quaternary $Al_xGa_{1-x-y}In_yN$ grown pseudomorphically on GaN substrates. (After [31.450], courtesy of Dr. E. Fred Schubert)

polarization charges at the InGaN/GaN interface, the electrons and holes in the quantum wells are separated at two ends of each quantum well due to the polarization charge (mainly of piezoelectric origin) induced electric field, which results in reduced electron-hole overlap and thus radiative recombination. Thus, the radiative recombination would depend on the quantum well width since different well widths result in different electric fields inside the well if all other parameters are fixed, which is the case for InGaN/GaN quantum well structures [31.481]. By utilizing unstrained quantum well layers with quaternary barriers, no emission energy dependence on width was observed due to the elimination of the piezoelectric field. By growing polarization-matched AlGaInN/InGaN LEDs, significant improvements have been achieved in terms of reduced forward voltage, reduced efficiency droop, and improved light-output power at large currents compared to conventional InGaN/GaN LEDs [31.480].

The bandgap and polarization charge versus the In and Al compositions have been calculated for AlGaInN as shown in Fig. 31.35, which clearly demonstrate independent control over the polarization charge and bandgap. However, among other difficulties brought about by the four-component systems, the optimal growth temperature is important to optimize and control, as Al-based compounds generally require higher growth temperatures, while In-based compounds require lower temperatures. Hence, the growth temperature will therefore govern the limits of In and Al incorporation into the AlGaInN quaternary alloy [31.478]. *Ryu* et al. [31.482] reported on optical emission in this quaternary system and AlInGaN/AlInGaN multiple quantum wells grown by pulsed metalorganic chemical-vapor deposition. A strong blueshift with excitation intensity was observed in both the quaternary layers and quantum wells, which was attributed to localization. This would imply that the structures are of inhomogeneous nature and/or presence of band-tail states indicative or early stages of material development and/or serious technological problems.

31.8 Doped GaN

Controlled doping in a broad range of concentrations in wide-bandgap semiconductors is very challenging for p-type varieties, as has been pointed out, for example, by *Walukiewicz* [31.483]. Simple substitution of elements such as C, Si, and Ge on the Ga sites and O, S, and Se on the N sites can potentially form shallow donors in GaN. Simple substitution of elements, such as Be, Mg, Ca, Zn, and Cd on the Ga sites and C, Si, and Ge on the N sites have the potential of forming relatively shallow acceptors in GaN. In practice, however, GaN n-type is achieved by using Si, Ge, and Se, the latter being not that common. On the other hand for p-type GaN only Mg has been successful but the level of hole concentration that can be achieved hit a ceiling of about $10^{18}\,cm^{-3}$.

Electronic and optoelectronic applications require n-type and p-type semiconductors. Even for high-quality GaN, grown by the many different processes (MBE, CVD, and HVPE), is unintentionally n-type doped, a phenomenon that is attributed to native defects (i. e., points defects, dislocation, stacking fault, V_N, antisites) and impurities (i. e., C, O, H) in GaN. Many potential p-type dopants attempted for incorporation into GaN cannot effectively compensate electrons in GaN and lead to p-type GaN. Hence, formation of p-type GaN is not straightforward. On the other hand, n-type doping of GaN is accomplished without major complications. This means that GaN p-type doping is the bottleneck to taking full advantage of the potential of GaN, especially in applications where both n-type GaN and p-type GaN are required, such as in optoelectronics.

The precursors, carrier gas, growth conditions, substrate material, among others, define the incorporation of impurities and imperfections in the crystal that result in the generation of defect-induced free carriers. De-

spite many decades of improvements in GaN technology, the residual electron concentration hovers around $1\times10^{16}\,\mathrm{cm}^{-3}$. Unintentional impurities imply lack of control, which is undesirable. In the following subsections a succinct description of GaN n-type and p-type doping is presented. A detailed treatment of the topic can be found in [31.1].

31.8.1 n-Type Doping

In unintentional GaN samples with a low n-type background, n-type doped GaN with Si is a very well-founded practice for both MBE and MOCVD techniques [31.484]. Controllability of electron concentration $\approx 10^{17}$ to $2\times10^{19}\,\mathrm{cm}^{-3}$ by varying the SiH_4 flow rate in MOCVD growth is possible [31.455]. With GeH_4 an order of magnitude higher electron concentration beyond what is feasible by Si is possible [31.485]. The Si ionization energy level in GaN is about 30 meV (dilution limit) and decreases as the doping level increases due to screening (see, e.g., [31.2, Chap. 5] for details). PL measurements in average-quality GaN yield a binding energy value of 22 meV [31.224]. Measurements on high-quality GaN layers grown on freestanding GaN wafers indicate (as determined by magneto-optical studies [31.486] and detailed analysis of the two-electron satellite transitions [31.487]) a binding energy of 30.18 meV.

Owing to the high solubility of Si ($10^{20}\,\mathrm{cm}^{-3}$) in GaN, Si is suitable for group III nitride doping and it is the most frequently used. Carrier concentrations in Si-doped GaN in the range 10^{17}–$10^{19}\,\mathrm{cm}^{-3}$ were observed, when silane (SiH_4) (used as an Si precursor) molar flow was varied 100 times. Also, Ge doping is well behaved. Ge solubility in GaN is in the range of 10^{17} to $10^{19}\,\mathrm{cm}^{-3}$ when germane (GeH_4) (used as Ge precursor) molar flow was varied 100 times in MOCVD growth. Ge doping has produced materials with electron concentrations of 7×10^{16}–$10^{19}\,\mathrm{cm}^{-3}$. More recently, *Wieneke* et al. [31.488] reported electron concentrations above $10^{20}\,\mathrm{cm}^{-3}$ in nonpolar *a*-plane GaN epilayers grown by MOVPE on *r*-plane sapphire substrates using isobutylgermane (IBGe) as a Ge precursor. A linear electron concentration versus molar flow relationship for both SiH_4 and GeH_4 precursors was observed [31.218, 489]. However, Ge incorporation is less efficient (by one order of magnitude) than Si incorporation.

31.8.2 p-Type Doping

As noted above, p-type doping in wide bandgap semiconductors is challenging and elusive, and nitrides are no exception. p-Type GaN, for the fabrication of p-n junction LEDs, was first achieved by *Akasaki* and *Amano* [31.490] in MOVPE-grown Mg-doped GaN. Doping activation was attained on Mg-doped GaN via low-energy electron-beam irradiation (LEEBI) treatment. The Hall measurement data showed a hole concentration of $\approx 2\times10^{16}\,\mathrm{cm}^{-3}$ with a mobility of $\approx 8\,\mathrm{cm}^2/(\mathrm{V\,s})$ and a resistivity of $\approx 35\,\Omega\,\mathrm{cm}$ at room temperature. *Nakamura* et al., with the same process that Amano used, obtained a hole concentration of $3\times10^{18}\,\mathrm{cm}^3$, a mobility of $9\,\mathrm{cm}^2/(\mathrm{V\,s})$, and a resistivity of $0.2\,\Omega\,\mathrm{cm}$ at room temperature [31.491]. Soon thereafter, *Nakamura* et al. [31.492] developed a doping activation technique based on N_2-ambient thermal annealing at temperatures above 700 °C. After the thermal treatment, the resistivity of Mg-doped GaN films dropped from 1×10^6 to $2\,\Omega\,\mathrm{cm}$ and the hole concentration and mobility were $3\times10^{17}\,\mathrm{cm}^{-3}$ and $10\,\mathrm{cm}^2/(\mathrm{V\,s})$ [31.492]. Nowadays, thermal activation is one of the most widely used techniques to active Mg in GaN. The thermal process is reversible with the GaN reverting to an insulating compensated state when annealed under NH_3, where the decomposed hydrogen from NH_3 was thus considered as the critical compensating agent during the annealing process under the NH_3 atmosphere.

A great deal of effort has been addressed to p-dope GaN and its ternaries by incorporating group II and group IV elements. So far, no other dopant has been as successful as Mg in converting GaN, AlGaN, and InGaN, with low Al and In mole fractions, to p-type materials. Needless to say, p-type doping in GaN and its ternaries remains a topic of interest both in terms of technological aspects and also at the fundamental level. In the case of the chemical vapor phase epitaxy technique [31.493], CCp_2Mg and MCp_2Mg are utilized for the source of Mg. Experimental results indicate CCp_2Mg is less volatile and its incorporation in GaN is more efficient at high temperatures, where a better GaN quality compared with the quality of GaN grown at lower temperatures is attained. Mg concentrations in p-type GaN films may be extremely high: $\approx 1\%$ of the host species. When substitutionally incorporated, only about 1% of the Mg atoms are ionized in GaN.

In Mg-doped GaN grown by MOCVD, the concentration of free holes at room temperature reaches its maximum value of about $10^{18}\,\mathrm{cm}^{-3}$ for a Mg chemical concentration of about $3\times10^{19}\,\mathrm{cm}^{-3}$, and it decreases with further increase of Mg concentration [31.494]. The activation energy of Mg acceptors in GaN and $Al_xGa_{1-x}N$ is in the range of 160–200 meV, which is larger than k_BT at 300 K and increases with Al fraction [31.495–499]. As a result, low conductivity of holes in p-type GaN and p-type $Al_xGa_{1-x}N$ is observed, degrading the performance of light emitters.

Because of the relevance of GaN-based optoelectronic devices, studies conducted to better understand the process and realize highly conductive p-type GaN and AlGaN are in demand, especially for the case of high Al-composition p-type AlGaN. Results on Mg-doped p-type $Al_xGa_{1-x}N$ with low Al content $0 < x < 0.27$ are available in [31.459, 500–502]. An activation energy as high as 310 meV was reported in [31.502] for $x = 0.27$. Demonstration of p-type conductivity in Mg-doped $Al_xGa_{1-x}N$ ($x = 0.35$) epilayers was reported by *Yu* et al. [31.503]. The influence of growth conditions on p-type conductivity was also investigated. The authors found that a proper V/III ratio and a relatively high growth rate were needed to improve the electrical characteristics of the p-type AlGaN epilayers. As a result of optimized growth conditions, a p-type resistivity of 3.5 Ω cm and a hole concentration of $> 5 \times 10^{17}\,\mathrm{cm}^{-3}$ were attained.

Unlike layers grown by MOCVD, MBE-grown p-type GaN layers do not require postgrowth annealing, neither employing activated nitrogen nor employing ammonia as the nitrogen source [31.504]. Mg-doped GaN layers grown by RMBE with ammonia as the nitrogen source exhibit p-type conductivity without any postgrowth treatment [31.496], which is very advantageous. In the case of GaN growth by MBE, Mg incorporation depends on the stoichiometry, the polarity of the growing surface [31.505, 506], the substrate temperature, and the growth rate [31.507–509].

Codoping for Improving p-Type Conductivity

As mentioned above, p-type GaN is challenging and low resistivity is difficult to achieve. One of the reasons is the compensation effect due to the low solubility and the low activation rate owing to the deep energy level of the acceptor. The acceptor energy level is very deep (about a few hundred meV) relative to room temperature (≈ 25 meV), for instance, GaN:Mg (200 meV) and AlN:C (500 meV). A 500 meV-deep acceptor if activated would yield one hole for every 10^6 acceptors at room temperature. To attain low-resistivity in wide bandgap semiconductors, the solubility of dopants should be increased while their compensation avoided. Also, the activation rate of the carriers and their mobility should be increased by reducing the energy level of the acceptors and the scattering rate of the carriers. As a response to mitigate the aforementioned concerns an interesting and promising method called codoping was proposed [31.510]. The codoping method is based on the idea that both n and p-type dopants are provided simultaneously under thermal nonequilibrium crystal growth conditions during MBE or MOCVD. The theoretical foundation of the codoping method was provided by *Reiss* et al. [31.511] [31.512] during the early stages of Si and Ge development using solution theory. Codoping can result in an enhancement of equilibrium solubility of the dopants (problematic for p-type impurities in GaN), low ionization energies, and also might affect the carrier mobility. All the above-mentioned benefits are relevant to GaN. Experimental results on codoping in GaN and in AlGaN are reported in [31.513–517] and [31.513, 518] respectively.

The hole concentration was observed to be enhanced in p-type GaN:Mg codoped with oxygen donors wherein the hole concentration increased linearly from 8×10^{16} to $2 \times 10^{18}\,\mathrm{cm}^{-3}$ with increasing oxygen dopant partial pressure. An increment of 3–5 times in the hole concentration resulted for a fixed oxygen partial pressure during growth of p-type GaN:Mg. Unlike the theoretical predictions, for reduced formation energy/increased solubility, when Si was codoped with GaN:Mg, the hole concentration remained constant [31.510].

Use of Superlattices for Improving p-Type Conductivity

Another promising method to get higher acceptor activation and lower resistivity is by using $Al_xGa_{1-x}N$/GaN doped superlattices [31.519–521]. The idea of using AlGaN/GaN heterojunctions for enhancement of p-type doping has been proposed as a technique to increase the average hole concentration [31.521–523]. Increased hole concentrations at room temperature through the use of AlGaN/GaN superlattices have been reported in the literature [31.524–526]. The mechanism for hole enhancement is the periodic oscillation of the valence band edge. Acceptors are ionized where the band edge is far below the Fermi energy and the resulting holes accumulate where the band edge is close to the Fermi level, forming a confined sheet of carriers. Although the free carriers are separated into parallel sheets, their spatially averaged density will be much higher than in a simple bulk film. High electric fields due to both spontaneous and piezoelectric polarization within the strained AlGaN layers strongly impact the band bending within the superlattice. The fields created by the superlattice/carriers exhibit sawtooth variation in the band diagram. The hole concentration expected from an AlGaN/GaN superlattice was calculated in [31.523].

The spatially averaged measured hole concentration and mobility for the Mg-doped $Al_{0.2}Ga_{0.8}N$/GaN superlattices with equal thickness, L, of the GaN and AlGaN layers (superlattice period $2L$) exhibit approximately constant mobility of $10\,\mathrm{cm}^2/(\mathrm{V\,s})$ for $2\,\mathrm{nm} < L < 14\,\mathrm{nm}$ and an average hole concentration which peaks at $2 \times 10^{18}\,\mathrm{cm}^{-3}$ for $L \approx 8\,\mathrm{nm}$. Further improvements of the p-type conductivity of the lattice were obtained by minimizing the ionized and neutral impurity scattering mechanisms via doping that maximize the separation of

the dopants (ionized scattering centers) from the multiple two-dimensional hole gas (2DHG) sheets.

Samples incorporating modulation-doped (MD) and shifted-modulation-doped (SMD) superlattices have been shown to have superior electrical properties compared to uniformly doped (UD) samples, especially at low temperatures [31.525].

After all the efforts invested to achieve high-p-type doping levels in GaN and AlGaN, the maximum attainable doping level is still a limitation for optoelectronic devices, particularly for shorter wavelength devices such as solar-blind ($\approx 290\,\mathrm{nm}$) detectors, UV LEDs, LDs and even visible LEDs, which need p-type AlGaN with high-Al composition and large hole concentration for efficiency retention. For example, a UV LED using an AlGaN p-n junction with an emission wavelength shorter than 290 nm can only give submilliwatt or milliwatt power [31.527].

31.9 Defects in GaN

Regardless of the degree of perfection of the material, defects are always present in semiconductors, particularly during the early stages of development. In fact, defects in complex semiconductors may never be fully solved [31.528]. By having to be grown on nonnative substrates due to the prohibitively expensive bulk material, GaN layers exhibit a wide variety of defects. Extended defects, such as threading dislocations and staking faults, and intrinsic defects, such as point defects and vacancies, among many others (for an in-depth discussion on defects the reader is encouraged to read [31.2]), may coexist in GaN eptaxial layers. In order to improve the quality of the layers a good deal of effort still is devoted to attaining material perfection via heteroepitaxy. However, after decades of research and investigations, both extended and intrinsic defects are still present to varying degrees. Nevertheless, defect densities have been reduced to a level where the realization of commercial devices based on GaN is feasible.

In what follows we center our discussion on two different kinds of defects. Because of their importance in the determination of electrical properties, we address the intrinsic point defects followed by staking fault defects, which is particularly important in nonpolar GaN for optoelectronic applications.

31.9.1 Points Defects

The importance of intrinsic point defects in nitrides is due to their role in determining the electrical and optical properties of the material, as well as their influence on the material behavior in various processing steps such as ion implantation and thermal annealing. Point defects are created in semiconductor materials during growth, where their formation is governed by thermodynamics and growth kinetics. Because GaN is unintentionally n-type, it is assumed that an acceptor-like defect is ionized and negatively charged as the Fermi level is close to the conduction band. Then, the important question is what is the source of these negative charges? To answer this question we need to take into account two factors in our analysis, namely the effect of incorporated impurities and charged defects. Both entities may behave like n or p donors. In GaN, point defects are identified as nitrogen vacancy (V_N), gallium vacancy (V_{Ga}) and Ga interstitial. A large variety of analysis tools have been used to study point defects and their impact on GaN physical properties. Some of the utilized techniques are photoluminescence (introduced already in Sect. 31.6), deep-level transient spectroscopy, minority carrier lifetime measurements, positron annihilation, electron paramagnetic resonance (EPR), and optically detected magnetic resonance (ODMR).

Deep-level transient spectroscopy (DLTS) [31.529, 530] is a popular method to get insight into electrical characteristics of point defects; both donor- or acceptor-like traps. The drawback of DLTS is its incapability to give local information or the chemical nature of the defect centers. These shortcomings can be alleviated to some extent if DLTS studies are realized in conjunction with systematic studies, such as irradiation damage and measurements in a pressure cell. Even though DTLS is a mature method, the characteristics of defects as analyzed by DLTS for GaN are widely scattered. A large dispersion in the basic reported parameters of the defects in GaN hide their genesis and nature. As many as eight point defects active in DLTS, some of which may be the same due to the aforementioned dispersion, have been reported in n-type GaN, and only one in p-type GaN.

GaN samples prepared by OMVPE on sapphire substrates have been investigated by DLTS by several groups [31.531–533]. Typical defect states are termed as E1, E2, E3, E4, and E5 with energy level in the range of 0.14−0.27, 0.49−0.598, 0.62−0.67, 0.81−0.85, and 1.07−1.44 eV respectively. The concentration of the same defects are in the ranges of 1.6×10^{13}−7.7×10^{14}, 2×10^{13}−1.2×10^{15}, 1.6×10^{14}−3.8×10^{14}, 1.5×10^{14}−2×10^{15}, and 2.4×10^{14}−$3.34\times10^{15}\,\mathrm{cm}^{-3}$ for E1, E2, E3, E4, and E5 traps respectively. An illustration of various defect states detected in DLTS and photoemission tran-

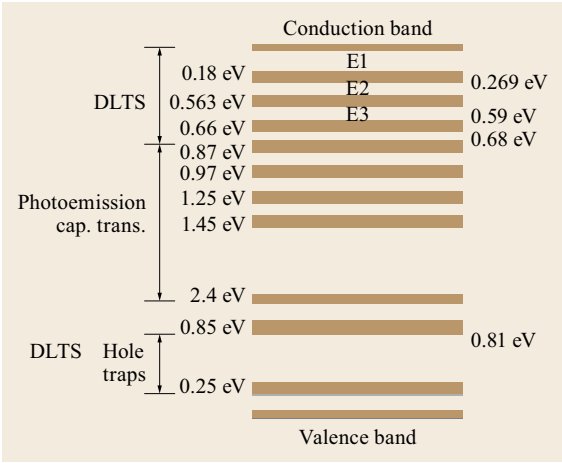


Fig. 31.36 Various deep levels within the gap energy of GaN grown by OMVPE, as observed by DLTS and photoemission transient spectroscopy. The two hole trap energies are with respect to the valence band maximum. (After [31.1])

sient capacitance methods are schematically shown in Fig. 31.36.

Minority carriers are widely trapped by defects, easily reducing their lifetime and diffusion length in the material. Minority carrier lifetime measurements are of importance in the evaluation of device performance and assessment of the material quality. Minority carrier lifetimes, τ, can be estimated using diffusion length L and measured mobility μ, through the expression $L = \sqrt{D\tau}$, where the diffusivity, D, and τ are related via the Einstein relation. The minority hole diffusion length can be derived from the line-scan electron beam-induced current (EBIC) measurements [31.534, 535]. The measurements are carried out at a distance from the material/substrate interface i. e., the GaN/sapphire interface [31.536]. In this kind of measurements the electron beam (positioned perpendicular to the sample edge) is moved from the vertical wall of the Au/n-GaN Schottky barrier toward another Au contact. The diffusion length L, can be obtained from the EBIC current decay versus distance from the edge of the Schottky contact (for distances $> 2L$). The methodology for diffusion length extraction can be found elsewhere [31.534, 535]. TEM and DLTS results clearly indicate that the defect density is reduced away from the GaN/sapphire interface. Then it is reasonable to evaluate the minority carrier lifetime as a function of distance away from the interface. Comparing the depth dependence of the dislocation density with that of the diffusion length, it is observed that the decrease in minority carrier diffusion length toward the GaN/sapphire interface is correlated with an increase in the dislocation density. The minority carrier diffusion length at a certain distance from the interface agrees well with the spacing between two adjacent dislocations, provided that the sample thickness is four times the minority carrier diffusion length. In which case, no sample thickness correction is needed because the sample can be construed as bulk [31.514].

Certain intrinsic defects (point defects/vacancies) can be studied with nondestructive positron annihilation spectroscopy (PAS) [31.537–558], which together with theoretical calculations provides a means to deduce both the nature and the concentration of the vacancies in the material. Parametric measurements with varying temperature provide information on the charge states of the detected vacancies. Positrons impingent onto a sample can get trapped in and localize neutral and negative vacancies due to the missing positive ion core. This positron lifetime and the momentum distribution of the annihilating positron-electron pair (Doppler broadening) represent the observable changes that the method relies on. Positron studies have been performed in nitride semiconductors for decades after the pioneering work of *Saarinen* et al. [31.559], which resulted in the identification of Ga vacancies in bulk GaN crystals. Many others studies were carried out [31.545, 547–549, 551–553, 555, 558]. The following is the most relevant data, regarding vacancies in GaN, InN, and AlN [31.540].

Vacancies in GaN

Most of the average positron lifetimes (τ_{ave}) measured in a variety of thick bulk or quasibulk GaN samples are above the bulk lifetime of 160 ps [31.560] for temperatures below 240 K indicating that positrons are annihilated as they get trapped at vacancy defects. The defects are identified as Ga vacancies by the second lifetime component $\tau_2 = 235 \pm 10$ ps deconvolved (separated) from the lifetime spectra [31.559, 561] as demonstrated in Fig. 31.37 [31.562].

The temperature dependence of τ_{ave} in O-doped and undoped GaN samples grown by HVPE at low temperatures points to an enhanced positron trapping at Ga vacancies, indicating that Ga vacancies are negatively charged. The lifetime, $\tau_{ave,}$ in GaN grown by the high nitrogen pressure (HNP) method indicates that nonopen volume defects, identified as MgGa-acceptors [31.559, 560], act as shallow traps competing with Ga vacancies in positron trapping at low temperatures.

No evidence of, or very low levels of, shallow traps in undoped and O-doped GaN grown by HVPE [31.561, 563, 564] set Ga vacancies as the dominant intrinsic acceptor defects in n-type GaN [31.565] (Fig. 31.37). The above-mentioned assessment is supported by a very good agreement between the vacancy and total acceptor densities, acquired by combining PAS and temperature-

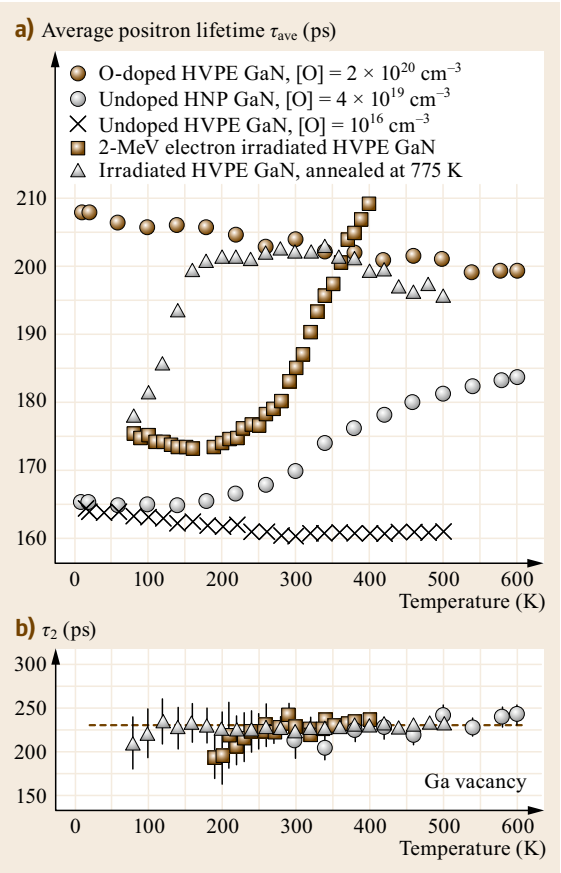


Fig. 31.37 **(a)** Average positron lifetime as a function of temperature. Vacancies are present when the average positron lifetime is above the bulk lifetime of $\tau_B \approx 160$ ps. The temperature dependence of the average positron lifetime evidences the different effects of negative charge of both the vacancies and acceptor-type impurities. **(b)** The second lifetime component extracted from the lifetime spectra, $\tau_2 \approx 235$ ps. This component is attributed to Ga vacancies or Ga-O complexes. (After [31.562], courtesy of Dr. F. Tuomisto)

dependent Hall measurements [31.564]. Experimental data show the Ga vacancy as being the most important compensating acceptor (over four orders of magnitude of the intentional oxygen doping) [31.563]. Ga vacancies in as-grown and highly O-doped GaN were identified as V_{Ga}–O_N pairs [31.563, 566], while isolated Ga vacancies are generated in electron irradiation experiments [31.561, 567]. Second lifetime components (Fig. 31.37b) are similar in grown O-complexed (V_{Ga}–O_N) and in isolated Ga vacancies, in good agreement with theory [31.563].

The τ_{ave} in as-irradiated (2 meV at room temperature) samples exhibit the coexistence of a high-Ga vacancy concentration and irradiation-induced negative ions. Annealed at 775 K, irradiated samples display a change in threshold temperature at which positrons can escape from negative ions (Fig. 31.37a). This can be interpreted as a change of positron-defects binding energy, i. e., change in the charge state of negatives ions, from 2^- to 1^-. Results from electron irradiation experiments [31.561, 567] showed a migration barrier $E_M = 1.8 \pm 0.2$ eV for isolated Ga vacancies. On the other hand, in HVPE GaN samples annealed at high temperatures and high nitrogen pressure, a redistribution of the V_{Ga}-O_N pairs [31.566–569] was observed, giving a way to calculate the binding energy of the pair as $E_B = 1.6 \pm 0.2$ eV, which is in good agreement with theoretical results [31.378, 379]. Studies on the effect of polarity (N-polar, Ga-polar, and nonpolar in both homo- and heteroepitaxy by HVPE) on Ga vacancies provided further understanding of the V_{Ga}–O_N pair formation [31.560, 570–572]. Experimental results indicate that Ga vacancies are created thermally as isolated defects in GaN growth at high temperatures in both HVPE and HNP methods and that the formation of V_{Ga}–O_N complexes is determined by the fraction of Ga vacancies that remain after the cool-down process.

O incorporation and subsequent vacancy depends on GaN growth polarity, owing to the different surface sticking coefficients of oxygen on different polarities. This effect is not observed in Si doping as the binding energy between Si donors and Ga vacancies is weaker, due to a larger distance between a Ga vacancy (acceptor) and substitutional Si (donor) [31.573].The above-mentioned results show that Ga vacancies act as dominant compensating centers in n-type GaN. In p-type GaN, because of its acceptor nature, Ga vacancy formation is energetically unfavorable. Studies of N vacancies (V_N) by PAS are useless owing to the small open volume generated by N vacancies. However, evidence of V_N-Mg_{Ga} complexes was obtained by PAS in Mg-doped (p-type) GaN grown by MOCVD [31.574, 575]. In [31.561] it is reported that N vacancies in p-type GaN samples that are electron irradiated (0.5 meV) behave like isolated neutral vacancies. Vacancies in both n-type and p-type GaN are a few percent of the doping densities. In n-type GaN, vacancies are dominantly compensating centers. On the other hand, in p-type GaN, other defects and impurities, such as hydrogen, play a more relevant role from the point of view of compensating centers.

The effect of growth conditions in thin-film (MBE, MOCVD), and bulk or quasibulk (HVPE)-grown samples were also studied by slow positron beam. Showing that stoichiometry has a very strong impact on Ga vacancy formation, the richer the growth in N and the higher the growth temperature, the higher the Ga va-

cancy concentration. This indicates that Ga vacancy is ruled by the thermodynamics of the reaction, but not necessarily by the growth method. The latter statement is supported by the results obtained from samples grown with various polarities [31.560, 570, 572, 576], where it is observed that the sticking coefficient and density of O on different growing surfaces determine the formation of Ga vacancies.

Vacancies in InN

InN layers of different thicknesses grown on sapphire by MBE [31.577, 578] have also been studied by PSA. The low electron-momentum parameter S and the high electron-momentum parameter W were calculated at room temperature to evaluate the positron annihilation with the valence electrons and the annihilation contribution from the core electrons, respectively [31.579]. In the context of the S parameter, the larger the value the higher the vacancy density in the layer. Regardless of the sample thickness, a value of $S = 0.48$ is registered for all samples for positron implantation energies in the range of (E) 0–1 keV (0–5 nm) (Fig. 31.38). For higher values of E and up to 30 keV, S remains constant. The S value and the extent of E values for which S remains constant is a characteristic of each sample. Note that for thinner InN layers, S decreases when the layer thickness increases. This is an indication of improvement in the quality of InN layer growth. Close to the InN-sapphire interface, the parameter S starts to decrease, approaching a value of 0.41, due to the increasing fraction of positrons annihilating.

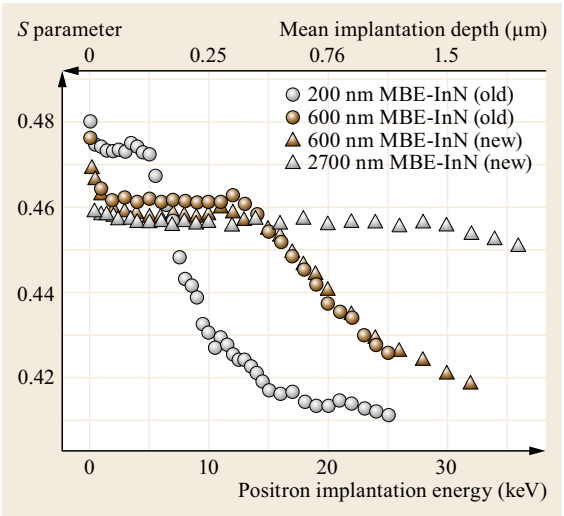


Fig. 31.38 The S parameter measured as a function of positron implantation energy in selected MBE-grown InN layers. *Old* from [31.577], *new* from [31.578]. (After [31.562], courtesy of Dr. F. Tuomisto)

Indium vacancy concentration is typically not sufficient to compensate electron conductivity in n-type InN. Indium vacancies are, as evaluated by electron mobility, effective scattering centers in InN. Improved electron mobility for thicker layers (up to 1 μm) suggest that In vacancy concentration is dramatically reduced [31.577]. An interesting point is that In vacancy formation does not seem to be promoted by the growth stoichiometry in MOVCD grown InN [31.580] and the same has been suggested to hold for MBE growth as well [31.581].

Vacancies in AlN

Average positron lifetime as a function of the temperature was measured in bulk AlN [31.582] grown by physical vapor transport (PVT) [31.172] and in thin layers grown heteropitaxily on various substrates by MBE [31.391] and MOCVD [31.583]. Qualitatively, the temperature dependence of τ_{ave} in AlN is similar to that on GaN. Any deviation in τ_{ave} as function of temperature in GaN and AlN can be interpreted as a competition between the negative ion type defects (shallow traps with no open volume) and vacancy defects for trapping positrons.

Taking an Al vacancy trapping coefficient of $3 \times 10^{15}\,\mathrm{cm^3\,s^{-1}}$ at 300 K and its temperature dependency, a lower limit for Al vacancy concentration is estimated as $1 \times 10^{17}\,\mathrm{cm^{-3}}$ at 600 K. On the other hand, negative ion concentration is $4 \times 10^{18}\,\mathrm{cm^{-3}}$ from data taken below 300 K, which indicates that negative ions are the dominant acceptor centers in AlN. A comparison between impurity and negative ion concentrations suggests that negative ions are related to oxygen incorporation and not to carbon impurity as is usually the case. In contrast with the case in GaN, O in AlN is predicted to behave as a negative charge.

In thin AlN films PAS results are similar to those acquired in GaN and InN thin films. Larger AlN-substrate mismatch and higher growth temperatures lead to higher Al vacancy concentrations. Also the growth stoichiometry effect is more notorious at very high V/III ratios. Measurements as a function of temperature provide valuable information on the charge states of the detected defects. Briefly, group III sublattice vacancies are common defects in all the group III nitrides. They compensate donors either by forming vacancy-impurity complexes or by providing deep states for electrons. Additionally, the presence of N vacancies was observed.

31.9.2 Extended Defects

As mentioned in Sect. 31.7 GaN-based LEDs and lasers rely for their functionality on QWs, which are formed,

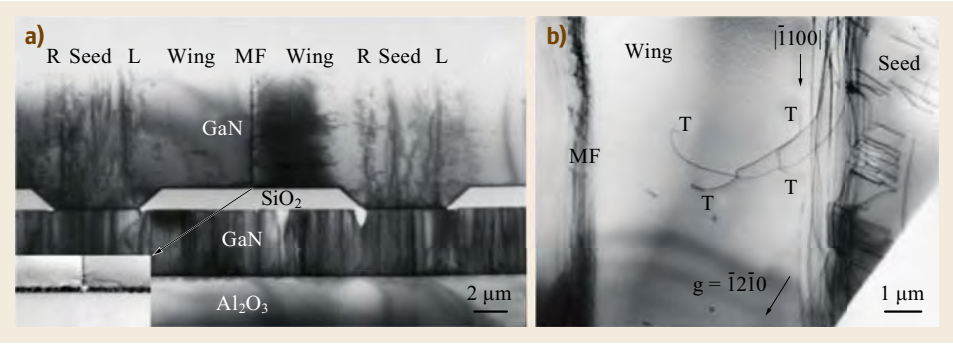


Fig. 31.39 **(a)** Cross-sectional TEM image showing defect distribution in a LEO layer. The seeds, left (L) and right (R) wings, meeting fronts (MF) of two wings, dielectric (SiO_2) layer and original GaN deposited on the Al_2O_3 (seed layer) are indicated. The *inset* shows a formation of a small void where two wings originally meet each other; **(b)** Plan view of the laterally overgrown area showing a meeting front of the two wings (MF) and one wing with a small part of the seed on the *right-hand side* of the image. The indicated threading parts of the dislocation, which originate either at the interface with the substrate or at some obstacles in the seed, are indicated as T. (After [31.584], courtesy of Dr. Zuzanna Liliental-Weber)

basically, by sandwiching a semiconductor material layer (GaN-based active region) with layers of semiconductor material with larger bandgap. On the other hand, because of economical, practical and functional reasons, the material for GaN-based optoelectronic devices is usually grown on SiC (for high-intensity LEDs) and sapphire substrates (note that the improved quality of GaN grown on Si substrates paved the way for high-brightness LEDs, which already made it into the marketplace); the latter is of low cost and light-transparent (from UV to IR). The lattice mismatch between the epitaxial layer and the substrate result in high-strain fields, which in turn lead to a high density of extended defects, i. e., threading dislocations (TD) and staking faults (SFs). Residual strain is also present in epitaxial layers upon cool-down due to the difference in temperature expansion coefficients. In fact the tensile strain when grown on SiC and Si leads to cracking, which must be addressed by strain engineering during growth.

Extended defects do affect device performance. TDs can act as both scattering centers and nonradiative recombination centers lowering either or both carrier mobility [31.585] and quantum efficiency [31.586]. The SFs are linked to high leakage current [31.587]. In short, extended defect density needs to be decreased in order to attain better device performance. A well-known method to decrease the defect density, for GaN growth on *c*-plane, is the epitaxial lateral overgrowth method (ELO) [31.588–590] (Fig. 31.39), which has been developed [31.591–593]. This approach yields a much lower density of defects (SF $\approx 1\times10^4\,cm^{-1}$), especially those that propagate along the growth direction. The ELO technique is also used to grow GaN along the semipolar and nonpolar directions.

When group III nitride samples grown on *c*-plane sapphire are heavily doped to produce n-type GaN, structural defect formation occurs. High-resolution transmission electron microscopy (HRTEM) investigation has revealed that these defects are small dislocation loops (in the range of few nanometers) associated with extrinsic SFs. This problem can be mitigated with the use of Ge doping instead of Si, as discussed below. The demonstrated superiority of Ge doping compared to Si, on GaN layers on *c*-plane sapphire, in terms of achievable free-carrier concentrations (up to $2.9\times10^{20}\,cm^{-3}$), surface smoothness, and reduced layer stress, has been reported in [31.485]. The authors observed a reduction in tensile stress for any Ge doping level, which could lead to a reduction of crystal defect density at high doping levels. On the other hand and in contrast with the structural defects formation in GaN heavily doped with Si, TEM investigations on Si-doped and undoped GaN samples grown by MOVCD on the (11−20) plane of sapphire revealed an improvement of the GaN layer quality with Si doping. For a Si concentration of $3\times10^{18}\,cm^{-3}$ dislocation density decreased from $5\times10^9\,cm^{-2}$ (undoped sample) to $7\times10^8\,cm^{-2}$ [31.594]. Also, a correlation between the Mg doping level and SF concentration in highly Mg-doped *c*-plane homoepitaxial GaN layers was found. As the Mg concentration increased from 2×10^{18} to $5\times10^{19}\,cm^{-3}$, the density of small, 3−10 nm-sized, SFs increased [31.595].

A scheme, which overcomes ELO limitations to a great extent, is the nanoheteroepitaxy (NHE) as shown schematically in Fig. 31.40. NHE has the potential

to eliminate defects more efficiently than conventional ELO. In the NHE technique coalescence would occur at a nanodimension scale reducing subsequent overall strain for minimal defect propagations. The NHE growth process begins with the definition of a two-dimensional array of e.g., 10–300 nm-sized nucleation islands on a substrate. It is then followed by a selectively grown epitaxial material vertically on the islands (in the case of AlN/SiC template, AlN array is generated with lithography followed by etching). What follows is the lateral overgrowth to achieve coalescence as indicated in Fig. 31.40. In contrast to the millimeter scale of patterns in ELO, patterning is in the nanometer scale in NHE, which allows for stress in the NHE sample to decay exponentially away from the heterointerface. The characteristic decay length is proportional to the diameter of the island. According to the NHE theory, a mismatch dislocation formation can be eliminated from materials systems with a lattice mismatch in the range of 0–4%. For a higher mismatch, defects are probably unavoidable, such as the case for GaN on Si (20% lattice mismatch); however, NHE may still result in a significant reduction in the local defect density.

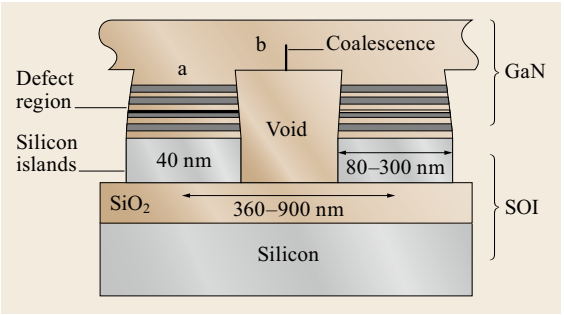


Fig. 31.40 Patterned silicon-on-insulator (SOI) substrate with islands created in the upper Si layer. The selective growth of GaN is also illustrated. A defected region is located at the heterointerface. *Point* a corresponds to selective growth of islands on a patterned substrate. *Point* b corresponds to the case where lateral epitaxy has been performed already. (After [31.596])

31.10 GaN-Based Nanostructures

At the nanometer scale, semiconductor structures exhibit novel electronic and optical properties, primarily owing to the dimensions that are comparable to the length of intrinsic processes (Fig. 31.41). The aforementioned novel properties are observed in GaN devices based on quantum wells [31.597–601], nanowires [31.601–605], or nanodots [31.606, 607]. In addition to the common issues related to conventional semiconductors such as band discontinuity, and dopant and carrier redistribution, polarization effects have to be taken into account in GaN devices. Polarization field effects may be in some cases desirable as in HFETs. But intense polarization fields foster an undesirable effect, the stark effect, which reduces the internal quantum efficiency in quantum confined structures. To mitigate the polarization effect, a quaternary alloy, InGaAlN, is used to neutralize the strain-induced component of the polarization field. Spontaneous polarization can be avoided by growing the heterostructure on $(1\bar{1}00)$ (*m*-plane) or $(11\bar{2}0)$ (*a*-plane), which represent the nonpolar planes.

31.10.1 Quantum Wells

As already discussed, the polarization-induced fields can drastically reduce the oscillator strength (Sect. 31.6.1, *Free Exciton in GaN* and Sect. 31.6.2, *Bound Exciton in GaN* for details on oscillator strength concept) for the ground-state optical transition [31.609–611]. The observed Stokes shift in two-dimensional (2-D) systems such as QWs [31.612, 613] is partly due to polarization effects. In this subsection we focus on the impacts of the piezoelectric effect and how to minimize them by manipulating well size and the crystal orientation. In multiple quantum wells

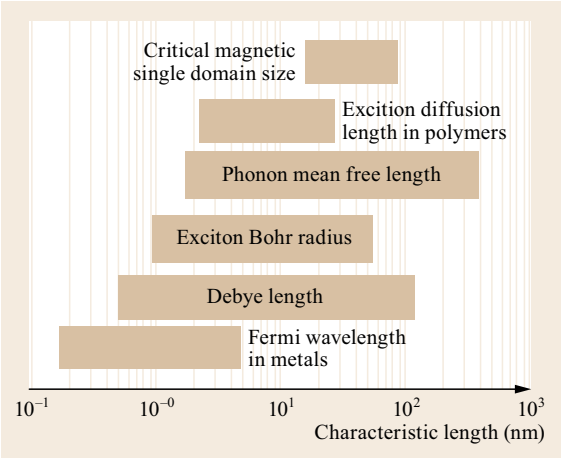


Fig. 31.41 Some of the well-known processes taking place in a semiconductor. Indicated characteristic lengths at 300 K go from tens of nanometers to hundreds of nanometers. The exciton Bohr radius is the most critical parameter in optical processes. (After [31.1])

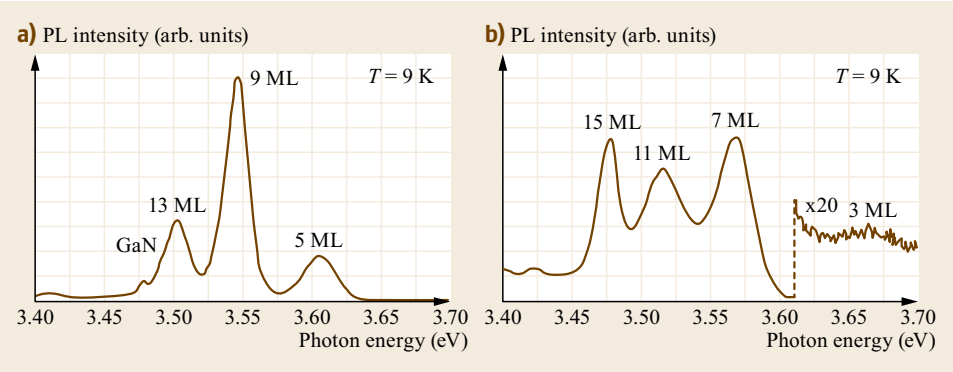


Fig. 31.42a,b Photoluminescence spectra at 9 K of $GaN/Al_{0.11}Ga_{0.89}N$ single QW. **(a)** corresponds to series A, **(b)** corresponds to series B. Note that in **(a)** the GaN buffer layer emission at 3.478 eV. (After [31.608], courtesy of Dr. N. Grandjean)

(MQWs), polarization effects manifest themselves as a redshift excitonic transition peak position [31.614, 615]. The effect can be compensated (blueshifting) by making the well size small enough to induce quantum confinement. The balance between redshift and blueshift is attained at a well thickness of 2 nm or less ($\approx$ 10 monolayers – MLs). The effect of QW thickness on the low-temperature photoluminescence spectra was studied by *Grandjean* et al. [31.608]. A single QW with 50 Å-thick $Al_{0.11}Ga_{0.89}N$ barrier layers grown by ammonia MBE were used. The data were from two samples: sample A consisting of 5, 9, and 13 ML well width, and sample B consisting of 3, 7, 11, and 15 ML width. PL spectra at 9 K are shown in Fig. 31.42.

It is evident from Fig. 31.42b, for the samples therein in particular, that the balance between the blueshift due to the confinement and the redshift, a consequence of the polarization effect, takes place at a QW thickness of 15 ML ($\approx$ 4 nm). The latter is wider than the values reported by other studies where a barrier layer with higher Al content was used. It is not necessarily a discrepancy and it may be because at a higher Al content the polarization effect is larger, which means the larger redshift needs to be compensated by a larger blueshift afforded by thinner QWs. Similar studies on AlGaN/GaN quantum wells grown by MOCVD but with high Al content (50%) in the barrier layer have been reported [31.616]. PL transition energies (not shown here) display a clear dominant transition from a redshift of 100 meV for a 4 nm QW width to a blueshift of 300 and 600 meV for QW widths of 2 and 1 nm respectively. The difference among the QW widths reported for the redshift-blueshift crossover, is mainly due to a larger than expected Al mole fraction in the PL reference samples and the growth technique, which may limit to some extent the heterostructure quality, particularly in terms of different degrees of localizing potentials. In the GaN system, the barrier thickness redistributes the polarization difference between wells and barriers, and thinner barrier layers were found to lead to larger blueshifts [31.600]. A point worth noting is that the trend of the electric field dependence on the barrier layer thickness can be found by using simple electrostatic calculations. However, not only piezoelectric-induced polarization, but also spontaneous polarization should be taken into account.

Nitrides are usually grown heteropitaxially on the basal plane of sapphire and SiC as well as Si substrates. Under normal growth conditions, nitrides assume wurtzite crystal structures, which have a hexagonal symmetry and a polar axis, *c*-axis [0001], along which both piezoelectric and spontaneous polarization are manifested. For optoelectronics, the ideal situation could be to grow nitride layers on nonpolar planes, i. e., $[1\bar{1}00]$ (*m*) and $[11\bar{2}0]$ (*a*) planes, in which case the polarization-induced shift would be eliminated, or at least the effect of the polarization-induced field associated with Stark shift could be suppressed. However, the inconvenience with nonpolar nitrides has to do with their film quality, which is lower than for those grown on *c*-planes. The most intensively investigated nonpolar directions are the $[1\bar{1}00]$ *m*-plane [31.617, 618] and the $[11\bar{2}0]$ *a*-plane [31.619, 620], which are free of polarization-induced internal electric fields along the growth direction.

At the present time, *a*-plane GaN is routinely grown on $(1\bar{1}02)$ or $(10\bar{1}2)$ *r*-plane sapphire by both MBE [31.619, 621] and MOCVD [31.618, 622–624]. However, heterostructures grown on *a*-plane, as well as the other nonpolar orientations, such as the *m*-plane, suffer from large stacking fault densities due to the low formation energy, making the radiative recombination at the desired wavelength relatively weak (this correction is a matter of style). Effects of systematic variation of well width on $GaN/Al_{0.16}Ga_{0.84}N$ quantum wells with a fixed 10 nm-thick barrier layer and 10-period *a*- or *c*-plane MQWs regrown by MBE on MOCVD-grown GaN/*r*-plane sapphire templates was investigated. Because of the absence of a polarization-

induced field, *a*-plane MQWs showed enhanced recombination efficiency as compared with *c*-plane wells, even though the *a*-plane variety presented a higher dislocation density [31.625]. It should be pointed out that this limited comparison does not necessarily hold in general.

31.10.2 Quantum Dots

In semiconductor quantum dots (QDs) electronic states are spatially localized and the energy is fully quantizied for all three directions. The atomic-like density of states near the bandgap is higher than in three-dimensional (3-D) and 2-D systems, leading to relatively high efficiency optical transitions. The latter is of great interest in terms of its potential in light emitting applications and optical sensors. The caveat of course is the carrier transport, which is impeded as well. Another interesting characteristic of the QDs is they are less sensitive to thermal perturbations, due to confinement, which is important for thermal stability of devices. Also, because the carriers are strongly localized their migration toward nonradiative centers such as dislocations outside of the dot region is somewhat inhibited [31.626].

Assuming good uniformity and control over the shape, quantum dots present some advantages as compared with QWs. Among them are a sharp energy level transition leading to reduced lasing thresholds, tunable bandgaps, and additionally reduced defects in nitrides. As in quantum wells, the optical properties of group III nitride QDs are assessed via photoluminescence. There is not an established consensus on the spectra emitted by GaN-based QDs. Published data indicate PL peak energies from 2.15 to 3.9 eV, depending on sample and temperature. QDs exhibit strong, albeit broad, PL intensity peaks with energies ranging from 2.6 to 3.9 eV (Fig. 31.43). The spread in the spectrum is most likely due to the divergence in dot size and shape.

As mentioned already, quantum confinement shifts the effective bandgap to higher energies. The amount of energy shift depends on the size and shape of the QDs, as well as the material optical properties of both the QD and its surrounding matrix. For small dots (disk, cubic, or sphere-like shape) of few nanometers in size, confinement energy is very sensitive to the dot size, becoming lower and less sensitive to the dot size for large-size quantum dots. Depending on the dot shape, confinement energy may change from 20 meV to 1 V for dot size from 2−10 nm [31.627].

Because of the relevance of strain-induced effects on group III nitrides, it should also be taken into account in QDs. In GaN QDs grown on $Al_xGa_{1-x}N$ and $In_xGa_{1-x}N$, which in turn is grown on GaN, strain is compressive, leading to a blueshift in the bandgap. For example, for a fully strained GaN on $Al_xGa_{1-x}N$, assuming the validity of Vergad's law, the bandgap energy shift is of 0.04 x eV, where x is the Al composition. On the other hand, as in the case of QWs, the polarization effect induces a redshift in the bandgap. Both spontaneous and piezoelectic (strain-induced) polarization induce very strong electric fields in nitride-based QDs, affecting significantly their optical properties. It should also be noted that the small dimensions associated with dots reduce the extent of the energy shift induced by polarization effects.

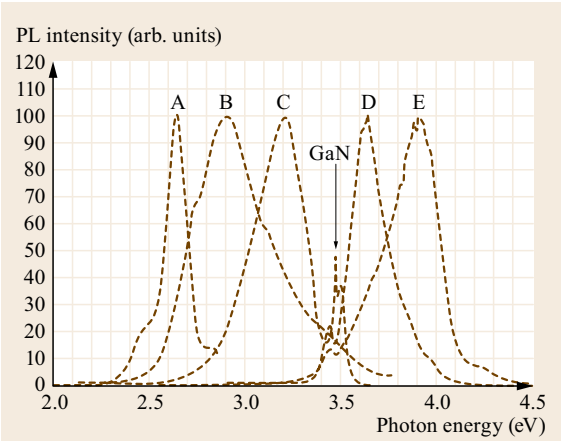


Fig. 31.43 PL spectra of GaN/AlN QD samples grown by MBE on different substrates and under different conditions. Samples A and B were grown on Si substrates; C, D, and E on sapphire substrates. Spectra were taken at 15 K under Ti-sapphire laser excitation (photon energy 5.06 eV). Intensity is normalized with respect to QD maximum signal. Also, for comparison, GaN bulk PL spectrum is displayed at 3.45 eV. (After [31.2])

31.10.3 Vertical Cavities

In terrestrial long-range and high-data-density communication systems based on fiber optics, where data transfer speed is as high as 100 G/s on 10 Tb/s [31.628], zero dispersion at a wavelength 1.3 μm [31.598, 599, 629–644], and lowest attenuation at 1.5 μm [31.597, 645–648] are possible. Then, it is desirable to count with powerful lasers emitting at the mentioned fiber-optic transmission windows. Vertical cavity surface emitting lasers (VCSELs) based on more recently developed dilute nitrides (GaAsN), are touted to satisfy the mentioned requirements (8 W at 1300 nm [31.649] and 12 W at 1120 nm [31.650]), but the quality requirements have not yet been met. A schematic cross-section description of a particular VCSEL design is presented in Fig. 31.44.

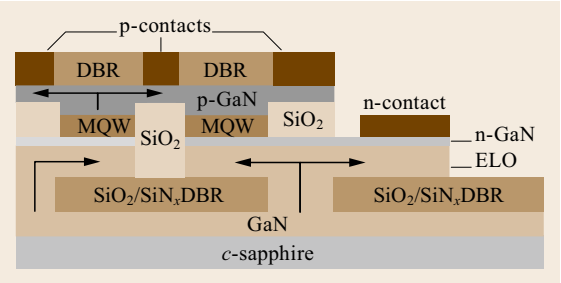


Fig. 31.44 Cross-sectional schematic of VCU's (Virginia Commonwealth University) VCSEL design. The design is meant for current injection. Lateral epitaxial overgrowth allows the implementation of distributed Bragg reflectors (DBR). The ELO growth technique allows all-dielectric DBR implementation. The design exhibits a SiO_2-defined current aperture. Current is injected and distributed via n and p GaN layers (after [31.651])

Aspects fostering the development of VCSELs are its small size ($\approx 10\,\mu m$ diameter), easier packaging and testing, possible fabrication and integration of 2-D arrays, and manufacturability. The smallest lasers available commercially today are vertical-cavity surface-emitting lasers [31.652]. The core part of VCSELs is the microcavity (MC). The constituting parts of a MC are top- and bottom-distributed Bragg reflectors (DBRs), which consist of $\lambda/4$-thick (where λ is the designed wavelength) alternating semiconductor or dielectric layers with low and high refractive index. Located between the DBRs is the active region of the VCSEL formed by multiple quantum wells (MQWs). The higher the refractive index dissimilarity, the contrast, between the alternating layers forming the dielectric stack, the fewer the number of stack pairs needed to attain DBRs with high reflectivity, which should be as close as possible to 100% for low-threshold current density. The cavity thickness should be an integer number of $\lambda/2$ times the length of the active region.

Difficulties arise when we shift from the 850 nm wavelength, either to shorter wavelengths (toward UV) or to longer wavelengths (toward infrared). In the case of short wavelengths, the number of quarter-wavelength DBR mirror pairs required increases considerably, which lowers the thermal conductivity and also increases the electrical contact's resistance. As a result, the heat dissipation problem increases substantially, which limits the VCSEL's thermal operating range.

Reflectors are commonly composed of either heterostructures or dielectric stacks and thus cleaved and/or etched facets are not applicable for vertical cavities. In the case of reflectors based on heterostructures, layers are made of materials with similar crystal structures but with different reflection indices; the larger the difference between their reflection indices the fewer the number of layers and the wider the blocking wavelength band. Due to the epitaxial growth of the active region of the VCSEL, bottom reflector stacks are commonly made of compounds compatible with the active region. Thanks to the development of the epitaxial lateral overgrowth technique (Sect. 31.10.2) for the GaN material system, it is possible to use dielectric stacks (SiN/SiO_2) as reflectors. Dielectrics constituting the stacks offer large reflection index contrast, which is beneficial to obtain a wider blocking band as well as a high reflection coefficient. On the other hand, top reflectors can be made of either semiconductors or dielectrics. Due to the short length of the active region in a vertical cavity, the gain is low. A minimum reflective coefficient for top and bottom DBR of 90% or higher is required, in order to have optical efficiency for lasing. The ideal case could be a reflection of 100% for the bottom DBR while the value for the top one should be slightly lower to allow light extraction, the extent of which depends on the desired output power.

Also in VCSELs based on GaN, AlN/GaN DBRs have been investigated. The main interest in such stacks stems from the fact that they possess the highest refractive index contrast among all the III–V compounds, so that a fewer number of stacks is required to obtain high reflectivity and relatively wide stop bands. But owing to the lattice mismatch between the AlN and GaN, which leads to crack formation (promoted by tensile strained GaN on AlN, and causing generation and propagation of defects), stack thickness is somehow limited. A way to mitigate the mentioned issues is by using AlGaN/GaN DBRs with Al concentration lower than 40% [31.653–657], and also introducing a digital equivalent of lower index materials as opposed to constant composition.

Figure 31.45 shows a stop-band-wide comparison between a 10-pair SiO_2/Si_3N_4 quarter wavelength DBR and a 30-pair $Al_{0.5}Ga_{0.5}N/GaN$ quarter wavelength DBR, both designed for a central frequency of 380 nm. The relevant point here is that even though few stacks were used, dielectric DBR achieves a higher reflectivity and a wider stop band compared with those obtained with semiconductor DBRs. GaN-based VCSELs with semiconductor-dielectric and all-dielectric reflectors were compared by *Okur* et al. [31.651]. The active region of VCSELs was formed by 6× $In_{0.01}Ga_{0.99}N$ (12 nm)/$In_{0.15}Ga_{0.85}N$ (2 nm) QWs. Sample 1 had semiconductor-dielectric DBRs, the top one being composed of 13.5 pairs of SiO_2/SiN_x while the bottom DBR was formed by 29 AlN/GaN pairs on top of a freestanding GaN (*c*-plane) substrate. Sample 2

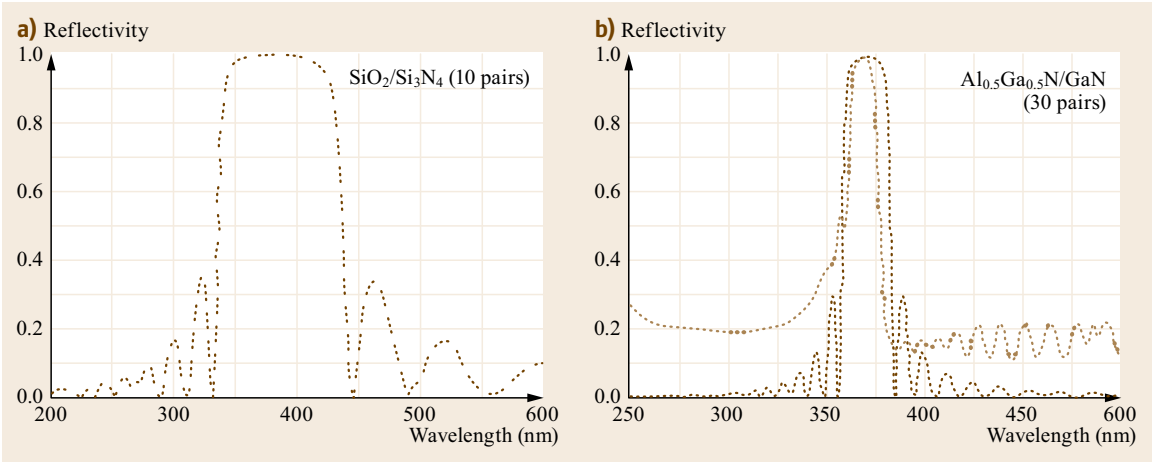


Fig. 31.45a,b Calculated reflection spectra. **(a)** for a 10-pair SiO_2/Si_3N_4 quarter wavelength DBR **(b)** for a 30-pair $Al_{0.5}Ga_{0.5}N/GaN$ quarter wavelength DBR. The *dotted line* is the calculated reflection spectra. Note that even though fewer pair stacks were used, dielectric DRBs exhibit a better reflectivity and wider stop band. (After [31.3])

consisted of all dielectric reflectors, where the top and bottom DBRs consisted of 13.5 pairs of SiO_2/SiN_x. In the latter case the entire structure was grown on a 3 μm GaN layer prepared by the ELO technique on top of a *c*-plane sapphire. Sample 2 exhibited a quality factor of 1200 at high excitation against a quality factor of 800 for sample 1, both at a wavelength of 403 nm. Also, sample 1 demonstrated one order of magnitude lower stimulated emission density ($\approx 5\,\mu J/cm^2$). On the other hand, *Butté* et al. [31.658] reported on VCSELs formed by a $5\lambda/2$ GaN cavity containing three $In_{0.15}Ga_{0.85}N/GaN$ QWs composed of a bottom 50-pair $Al_{0.83}In_{0.17}N/GaN$ DBR and a top 20-pair SiO_2/Si_3N_4 DBR with a quality factor as high as 6400 as determined from PL spectra at a wavelength of 421.1 nm.

Even though VCSELs are advantageous in many respects, some technological aspects delayed the GaN-based VCSEL development. One of those aspects is related to the fabrication of high-Q optical resonators. Another issue is the electrical current injection, which should incorporate lateral current spreading, because of the low conductivity of the p-GaN and its alloys used in the construction of some DBRs. Also, smooth dielectric and semiconductor interfaces are required of VCSELs to a greater extent due to the relatively short wavelengths involved. A roughness of smaller than 2–3 nm for areas of several hundred square micrometers is required. VCSELs based on nonpolar GaN have also been demonstrated by *Holder* et al. [31.659]. Vertical cavities were implemented on *m*-plane freestanding GaN substrates grown by HVPE. The active region, grown by MOCVD, consisted of five $In_{0.1}Ga_{0.9}N$ QWs (7 nm-thick) separated by 5 nm GaN barriers. DBRs were formed by 10-pair SiO_2/Ta_2O_5 (bottom) and 13-pair SiO_2/Ta_2O_5 (top).

Kasahara et al. [31.660] demonstrated current injection stimulated lasing at room temperature in GaN-based VCSELs on GaN substrates. CW operation for blue (8 μm aperture) and under pulsed (pulse width: 1.0 μs; pulse period: 500 μs) current operation (10 μm aperture) was reported. Devices were formed by InGaN/GaN quantum wells in conjunction with dielectric DBRs and a vertical current injection configuration. Blue CW lasing at 451 nm had a threshold current and voltage of 1.5 mA and 3.3 V respectively. Pulsed green lasing (503 nm) was attained at a threshold current of 22 mA and a threshold voltage of 6.3 V. Maximum output power for blue emission was 0.7 mW, while for green light it was 0.8 mW. The authors suggest that the difference in the threshold current, from blue to green emission, is caused by the quantum confinement Stark effect caused by the piezoelectric polarization-induced field, the genesis of which is the strain due to the mismatch between GaN and InGaN and which increases for higher In content. Electrically injected stimulated emission at a wavelength of 411.9 nm and a maximum power output of 19.5 μW in these devices was observed. However, high threshold current and very low forward voltage suggests high leakage current. Laser emission was polarized-locked with an optical electric field oriented parallel to the *a*-plane, in contrast to that which occurs in *c*-plane VCSELs, where the emitted light is randomly polarized. This can be a key characteristic of VCSELs built on m-plane orientations for applications where high-intensity polarized light is desired.

31.10.4 Nitride Nanorods

Nanorods, also known as nanowires (NWs) and whiskers, can be synthesized with a 5−500 nm diameters, and be dozens of micrometers long given the opportunity, as mentioned in the introduction of this section, to explore the realm of the mesoscopic phenomena. Some of the potential and interesting applications are nanowire light-emitting diodes (NW LEDs) in all visible spectra [31.606, 661–664], white light generation [31.11], nanowire lasers (NW lasers) [31.665–668] with emission wavelengths from the deep ultraviolet (UV) [31.669] to the near-infrared [31.670–672] spectral range, and photodetectors [31.673–676]. However, for successful insertion of nitride nanowires into the actual market, electrical injection, ultralow-threshold, and continuous wave (CW) operation at room temperature are crucial. In this vein some obstacles need to be overcome first, such as the carrier loss to defect-related nonradiative and surface recombination processes [31.602], electron overflow [31.677, 678], and poor hole transport [31.679]. In order to enhance group III nitride NW performance, thermal management in NW lasers, which plays a principal role, also needs to be addressed.

Because of the reduced diameter of the NWs, thermal conductivity is also lower compared with that of the bulk. Thermal conductivity of single-GaN NWs for different diameters has been reported [31.680–682]. Thermal conductivities as low as 0.13−0.19 $\mathrm{Wcm^{-1}K^{-1}}$ at 300 K (while for bulk Si, GaN and 4HSiC is 1.5, 2.3, 4.5 $\mathrm{W\,cm^{-1}\,K^{-1}}$ respectively) were reported. NW doping, for both n and p-type materials, still remains a challenging problem, especially the p-type variety, which makes electrical injection difficult to attain. n-ype NWs doped with Si have been demonstrated [31.683, 684], while p-type doping, as in the case of planar structures, is a main concern, especially for InGaN and InAlN with high Al and In compositions. Generally, lasing in NW lasers is demonstrated via optical pumping, while electrical injection in group III nitride NW lasers remains elusive. At this time there is no report on electrically pumped group III nitride NW lasers.

One-dimensional nanostructures are usually grown by a variety of techniques. Among them are vapor phase, laser ablation, chemical vapor deposition [31.685, 686], chemical vapor transport methods, molecular beam epitaxy [31.687], sputtering [31.688–692], vapor-liquid-solid (VLS) [31.693–697], inclusive of self-catalytic and vapor phase varieties, vapor-solid (VS) growth [31.698, 699], confined chemical reaction [31.700–702] such as that utilized for carbon nanotube preparation, and anodized Al membrane confined reactions [31.703, 704]. No matter what growth techniques is used, it must have attributes such as reproducibility, uniformity, scalability, cost-effectiveness, and a well understood growth process. Nanorods are typically grown by patterning techniques as an integral part of the growth process. For instance, high-quality InGaN NWs have been grown by many methods including CVD [31.603, 705], MBE [31.603, 705–708], and hydride vapor phase epitaxy [31.709, 710] to produce a diverse class of heterostructures: core-shell [31.604, 705], core-multishell [31.604], well/disk-in-a-wire [31.711, 712], and dot-in-a-wire [31.606, 706]. Also, InN nanostructures have been grown in many different shapes: nanocrystals [31.713], nanocolumns [31.714], nanorods [31.715], NWs [31.684], nanotubes [31.716], and nanobelts [31.602, 717]. For more details on group III nitride nanorods, the reader may wish to see reviews [31.667, 718–723].

As in the case of layers grown by heteroepitaxy, the material perfection is highly coveted, and in order to master the growth process, an in-depth understanding of the kinetics and thermodynamics of reactions applied to a wide variety of material systems is necessary [31.724]. Note that for group III nitride nanowires there are no affordable lattice-matched substrates available. The small base area, however, together with the large surface/volume ratio leads to relaxed nanowires without extended defects on nonnative substrates. As a result, nearly structural defect-free nanorods are possible to attain. A relaxed material is beneficial in terms of minimization of the strain-induced polarization effects in the polar direction, which is advantageous for heterojunction systems, such as green LDs based on InGaN in a GaN matrix.

For optoelectronic applications, sapphire and Si are of most interest. The optical window of sapphire is consistent with UV, visible, and near-infrared applications. Even though the Si transmission window begins at the far-deep-infrared spectrum, the main interest in Si wafers lies in their relatively low cost and possibility to integrate light emitters with electronics in cases where the performance of the emitter does not rely on substrate transparency i. e., VCSELs, NWs Fabry–Pérot cavities and longitudinal heterostructures. The cost issue alone has been driving the LED industry to migrate to Si substrates along with processes that are already in place for substrate removal.

As noted already, nanowire growth in a variety of cross-sectional shapes (hexagonal, triangular, circular, and rectangular) has been achieved. The triangular cross-section is to some extent related to the growth method and growth conditions employed. Hexagonal or triangular cross-section NWs are attained commonly by MOCVD [31.605, 725]. Nearly extended defect-free wurtzite single-crystal GaN nanocolumns grown

by plasma-assisted molecular beam epitaxy with a diameter in the range of 20–40 nm were grown along the [0001] *c*-direction on Si substrates with (001) and (111) surfaces [31.726]. This represents another piece of evidence that, when not bounded by the substrate, the wurzite phase is the preferential polytype with dominant growth parallel to the polar direction. Vertically aligned [0001] GaN nanorods on sapphire were grown by MBE using an RF source for reactive nitrogen [31.687]. Nanorod growth that took place under GaN growth conditions deviated from that used for standard growth (planar growth). In other words, the growth takes place at relatively high temperatures (> 800 °C), some 200 °C higher than that used to obtain uniform layers under Ga-rich conditions, and under very high nitrogen fluxes. This approach exploits the very high sticking coefficient of nitrogen and the high evaporation rate of Ga at high temperatures.

Because high-quality devices based on heterostructures are commonly produced by MOCVD and to a lesser extent by MBE, these methods could facilitate a smooth integration of nanowire growth with compositional and doping modulation in the growth direction. In practical applications, more precisely those based on electrical injection, a p-type semiconductor is crucial. The p-type doping in GaN systems for functional device structures was reported in [31.727]. The authors used a chemical vapor transport vapor-solid process to introduce Cp_2Mg to accomplish p-type doping. In another report [31.601], the authors used the vapor-liquid-solid (VLS) method to grow nanowires. The tuned incorporation of Mg during synthesis resulted in 35 nm-diameter nanowires with rectifying p-n junction properties. GaN nanowires synthesized by a catalytic chemical vapor deposition, with hexagonal single-crystalline structure and diameters of 10–50 nm and lengths of tens of micrometers, exhibited UV bands at 3.481 and 3.285 eV in low-temperature PL measurements associated with donor-bound excitons and donor-acceptor pairs respectively [31.728]. The observed blueshift of UV bands was attributed to quantum confinement effects in thin GaN nanowires with diameters of about 11 nm. The turn-on field emission of nanowires was 8.5 V mm^{-1} and the current density was about 0.2 mA cm^{-2} at 17.5 V mm^{-1}, which is sufficient for field emission display applications and vacuum microelectronic devices. Also, GaN nanowires formed via the VLS mechanism with gold, iron, or nickel as catalysts, exhibited orientations predominantly along directions [001] and [101]. The NWs with growth orientation [210] featured a triangular cross-section with an n-type character and a mobility of 65 cm^2 V^{-1}s^{-1} [31.729].

In an attempt to accomplish highly efficiency light emission, *Qian* et al. [31.604] prepared core multishell nanowire (CMS) radial heterostructures by MOVPE having triangular cross-sections, which allowed synthetically tunable multicolor light emitting diodes. Structures consisted of n-GaN cores surrounded by $In_xGa_{(1-x)}N$/GaN/p-AlGaN/p-GaN shells. Electrical carrier injection was achieved by forming metal contacts on the p-type outer shell and n-type cores at the ends of individual nanowires. I–V curves exhibited p-n diode characteristics with a knee voltage of around 3.5 V. CMS nanowires display electroluminescence spectra with emission peaks at 367, 412, 459, 510, and 577 nm, consistent with $In_xGa_{1-x}N$ band edge emissions for In compositions of 1, 10, 20, 25, and 35%, respectively. Estimation of quantum efficiencies (QEs) for these CMS nanowire LEDs gave a value of 5.8% at 440 nm and 3.9% at 540 nm, both of which are several times better than the QEs reported for nanoscale LEDs and approach InGaN-based SQW thin-film LEDs at similar emission wavelengths. InGaN/GaN multiple quantum disk (MQD) nanocolumn LEDs are fabricated on n-type Si(111) substrates [31.730] as well. The columnar growth was preceded by the deposition of GaN dots. The latter was obtained by exposing the surface of the wafer to a Ga beam and RF plasma excited nitrogen at 530 °C. In order to attain p-type GaN, growth conditions were changed to those suitable for GaN-Mg doping. Resulting nanocolumn LEDs exhibited diode I–V characteristics with a typical turn-on voltage of 2.5–3 V at room temperature [31.711]. With this process, functional LEDs emitting in the entire visible spectrum were demonstrated.

Regardless of the growth direction of the NWs, the junction area between InGaN and GaN is reduced. Therefore, strain along the direction perpendicular to the growth is relaxed. Then a lower density of defects and weaker piezoelectric polarization-induced fields are expected, which is an advantage in comparison with InGaN-GaN quantum wells in planar structures where strain degrades the internal quantum efficiency (IQE) of the LED. In contrast, and in addition to a higher extraction efficiency (because of the NW aspect ratio), NWs exhibit an improved IQE. By way of background, white LEDs are commonly achieved by blue LED pumping a yellow phosphor-dye and leaking through it [31.731], a technique where a blue LED chip is packaged with a precise phosphor arrangement. As a result, blue light is converted to white light. *Nguyen* et al. [31.606, 732] obtained phosphor-free nanowire white light LEDs by the incorporation of different In content levels in stacked InGaN/GaN QWs and the use of an p-doped AlGaN electron blocking layer between the active region and the uppermost NW layer, the p-GaN layer. The authors attained stable white light emission, which was attributed to the large variation of QW In composition,

uniform carrier distribution in the LED active region, and reduced quantum confinement stark effect [31.733]. Also, the authors claim a negligible degradation up to $\approx 2200\,\mathrm{A/cm^2}$.

Lasers have also been attained in NWs, albeit with optical pumping at this stage. The possibility to grow NWs on Si wafers paves the way for their integration with very-large-scale integration circuits (VLSIs) if suitable applications can be visualized. Along these lines, recent investigations have focused on making NW lasers compatible with VLSI circuits. The integration of optics and digital electronics in the same chip is of great interest in high-speed communication systems, fiber optic communication, and in all the relevant potential applications where optoelectronic at the nanometer scale is required. The reduction of laser dimensions is associated with some beneficial features, such as low-threshold, high efficiency, and low power consumption, all of which are difficult to achieve with VCSELs. It should be noted that group III nitride NW lasers under electrical injection at room temperature and continuous wave face considerable challenges. Carrier loss mechanisms (nonradiative centers), electron overflow, poor hole transport, thermal management, and poor ohmic contact to the p-GaN layer all need to be overcome in order to fully exploit the potential of group III nitride NW lasers.

Among the studied micro- and nanocavities are Fabry–Pérot [31.665, 669, 717, 734–738], nanoplasmonics [31.739, 740], photonic crystals [31.741], micro stadium lasers [31.666], and ring cavities [31.742]. The Fabry–Pérot (FP) cavity, the simplest one, has been very widely investigated. The NW lasers based on FP cavities consist of two crystalline facets at the ends of the NWs. The optical field propagates along the longitudinal direction of the NW, being amplified and absorbed inside the active region and reflected back by the crystalline facets. Lasing threshold conditions in NWs result from the balance of the round-trip losses and the round-trip gains in the cavity. For details on NW Fabry–Pérot lasers see [31.668].

Optoelectronic devices based on nitrides, such as LEDs and lasers, although commercially available, are still far from being fully exploited from the viewpoint of what GaN has to offer. This is mainly because the material quality, in spite of continual improvement, is still far from being nearly defect-free. In addition, limitations imposed by p-type doping make it difficult to obtain good contacts and provide a sufficient number of holes to match that of electrons in optical devices. As a result, electrical carrier injection is limited. Optically pumped nanolasers have been demonstrated by many groups around the world. However, electrical carrier injection in group III nitride NW lasers is still lacking. To succeed in practice, nitride-based NW lasers need to be electrically driven, and therefore they should feature good electric contacts. Also their manipulation should be well controlled.

31.11 Summary and Conclusions

Group III nitride semiconductors continue to attract a good deal of interest due to their numerous industrial applications. To begin with, AlN exhibits many useful mechanical, electronic, and thermal properties, for example, hardness, resistance to high temperature and caustic chemicals, and high thermal conductivity in crystalline shape, and so on. The wide bandgap is also the reason AlN is touted as an insulating material in semiconductor device applications, for example, AlN has been attempted as surface passivation layer in AlGaN/GaN-based heterojunction field effect transistors (HFETs) with the benefit of better heat dissipation compared to other common passivation materials. Piezoelectric properties also make AlN suitable for surface-acoustic-wave devices and sensor applications. However, the majority of interest in this semiconductor, in the context of electronic and optoelectronic devices, stems from its ability to form alloys with GaN and InN to tune the bandgap, polarization charge, and lattice constant. A few usages of AlGaN alloy are as a barrier layer in AlGaN/GaN HFETs, as an electron blocking layer in InGaN-based LEDs, as buffer layers in AlGaN/GaN HFETs to reduce the buffer leakages, and in deep UV LEDs, and so on. AlN also forms a crucial component of the nitride-based AlInGaN quaternary, which allows tuning of the bandgap independently of the composition over a reasonably wide range of bandgaps. Currently, AlN crystals up to 2 in in diameter are available; however, many challenges still remain that need to be addressed to further exploit its usage in electronic and optoelectronic applications.

Although GaN has been studied far more extensively than the other group III nitrides, there is still a great need for further investigations to approach the level of understanding of technologically important materials such as Si and GaAs. GaN growth still suffers from the insufficient control of impurity levels, defects and dislocations. Although native substrates are commercially available, which can alleviate the above-mentioned challenges to a certain extent, these issues

may remain for a fairly long time when foreign substrates such as sapphire and silicon are used for cost reduction as well as other considerations. These, together with the difficulties in obtaining high-quality p-type doping layers, will certainly slow down the process of further improvements in the performance of GaN-based electronic and optoelectronic devices.

Oddly enough, InN received much of the early attention compared to the other nitride binaries. However it is still difficult to grow high-quality crystalline InN layers. Although InN can be alloyed with GaN, which enables the tuning of the bandgap from $\approx$ 0.7 to ≈ 3.4 eV, the high-In composition InGaN has to be grown at low temperatures, rendering rather low-quality layers. Thus, small-bandgap InN does not escape competition by the existing alternatives: well-characterized/developed semiconductors such as AlGaAs and In (Ga, Al)AsP. Consequently, practical applications of InN may be restricted to its alloys with GaN and AlN, especially low-indium composition alloys, in addition to tandem solar cells, which are very unlikely to be successful. The growth of high-quality InN/high-indium-composition alloys and the enumeration of their fundamental physical properties remain, for the present, a scientific enterprise.

References

31.1 H. Morkoç: *Handbook of Nitride Semiconductors and Devices*, Vol. 1 (Wiley, Weinheim 2008)

31.2 H. Morkoç: *Handbook of Nitride Semiconductors and Devices*, Vol. 2 (Wiley, Weinheim 2008)

31.3 H. Morkoç: *Handbook of Nitride Semiconductors and Devices*, Vol. 3 (Wiley, Weinheim 2008)

31.4 L. Sang, M. Liao, M. Sumiya: A comprehensive review of semiconductor ultraviolet photodetectors: From thin film to one-dimensional nanostructures, Sens. Switz. **13**(8), 10482 (2013)

31.5 M. de la Mata, X. Zhou, F. Furtmayr, J. Teubert, S. Gradecak, M. Eickhoff, A. Fontcuberta i Morral, J. Arbiol: A review of MBE grown 0-D, 1-D and 2-D quantum structures in a nanowire, J. Mater. Chem. C **1**(28), 4300 (2013)

31.6 M.P. Ulmer: A review of UV detectors for astrophysics: Past, present, and future, Proc. SPIE–Int. Soc. Opt. Eng. **7222**, 722210 (2009)

31.7 M. Meneghini, A. Tazzoli, G. Mura, G. Meneghesso, E. Zanoni: A review on the physical mechanisms that limit the reliability of GaN-based LEDs, IEEE Trans. Electron Devices **57**(1), 108 (2010)

31.8 N. Trivellin, M. Meneghini, E. Zanoni, K. Orita, M. Yuri, T. Tanaka, D. Ueda, G. Meneghesso: A review on the reliability of GaN-based laser diodes, Proc. 2010 IEEE Int. Reliab. Phys. Symp. (IRPS 2010), Piscataway (2010) p. 6

31.9 M. Meneghini, L.-R. Trevisanello, G. Meneghesso, E. Zanoni: A review on the reliability of GaN-based LEDs, IEEE Trans. Device Mater. Reliab. **8**(2), 323 (2008)

31.10 T. Hashimoto, E. Letts, S. Hoff: Current status and future prospects ammonothermal bulk GaN growth, Sens. Mater. Jpn. **25**(3), 155 (2013)

31.11 E. Okahisa, S. Masui, T. Yanamoto, S. Nagahama: Current status and issues of blue laser diodes, Rev. Laser Eng. Jpn. **41**(4), 230 (2013)

31.12 D.F. Feezell, M.C. Schmidt, S.P. DenBaars, S. Nakamura: Development of nonpolar and semipolar InGaN/GaN visible light-emitting diodes, MRS Bull. **34**(5), 318 (2009)

31.13 E. Zanoni, M. Meneghini, N. Trivellin, M. Dal Lago, G. Meneghesso: GaN-based LEDs: State of the art and reliability-limiting mechanisms, Proc. 2014 15th Int. Conf. Thermal, Mech. Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), Piscataway (2014) p. 5

31.14 Z.J. Liu, W.C. Chong, K.M. Wong, K.M. Lau: GaN based light-emitting diode on silicon (LEDoS) micro-displays for BLU-free full-color projector application, Proc. 2013 IEEE Photonics Conf. (IPC), Piscataway (2013) p. 171

31.15 T. Paskova, D.A. Hanser, K.R. Evans: GaN substrates for III-nitride devices, Proc. IEEE **98**(7), 1324 (2010)

31.16 V. Avrutin, D.J. Silversmith, Y. Mori, F. Kawamura, Y. Kitaoka, H. Morkoc: Growth of bulk GaN and AlN: Progress and challenges, Proc. IEEE **98**(7), 1302 (2010)

31.17 J. Piprek: III-nitride LED efficiency droop models: A critical status review, Proc. 2013 13th Int. Conf. on Numerical Simulation of Optoelectronic Devices (NUSOD), Piscataway (2013) p. 107

31.18 M. Beeler, E. Trichas, E. Monroy: III-nitride semiconductors for intersubband optoelectronics: A review, Semicond. Sci. Technol. **28**(7), 074022 (2013)

31.19 M. Razeghi: III-Nitride optoelectronic devices: From ultraviolet toward terahertz, IEEE Photonics. J. **3**(2), 263 (2011)

31.20 R. Paiella, K. Driscoll, Y. Li, Y. Liao, A. Bhattacharyya, C. Thomidis, L. Zhou, D.J. Smith, T.D. Moustakas: Intersubband transitions in GaN-based quantum wells: A new materials platform for infrared device applications, Proc. SPIE Int. Soc. Opt. Eng. **7808**, 780807 (2010)

31.21 S.P. Najda, P. Perlin, T. Suski, L. Marona, M. Bockowski, M. Leszczynski, P. Wisniewski, R. Czernecki, R. Kucharski, G. Targowski: Latest developments in AlGaInN laser diode technology, Proc. SPIE Int. Soc. Opt. Eng. **8625**, 862517 (2013)

31.22 R. Kucharski, M. Zajac, R. Doradzinski, M. Rudzinski, R. Kudrawiec, R. Dwilinski: Non-polar

and semi-polar ammonothermal GaN substrates, Semicond. Sci. Technol. **27**(2), 024007 (2012)
31.23 H.T. Grahn: Nonpolar-oriented GaN films for polarization-sensitive and narrow-band photodetectors, MRS Bulletin **34**(5), 341 (2009)
31.24 H. Ohta, K. Okamoto: Nonpolar/semipolar GaN technology for violet, blue, and green laser diodes, MRS Bulletin **34**(5), 324 (2009)
31.25 S. Albert, A. Bengoechea-Encabo, M.A. Sanchez-Garcia, F. Barbagini, E. Calleja, E. Luna, A. Trampert, U. Jahn, P. Lefebvre, L.L. Lopez, S. Estrade, J.M. Rebled, F. Peiro, G. Nataf, P. De Mierry, J. Zuniga-Perez: Ordered GaN/InGaN nanorods arrays grown by molecular beam epitaxy for phosphor-free white light emission, Frontiers in Electronics. Selected Papers from the Workshop on Frontiers in Electronics 2011 (WOFE-11), Singapore (2013) p. 109
31.26 D. Zhu, D.J. Wallis, C.J. Humphreys: Prospects of III-nitride optoelectronics grown on Si, Rep. Prog. Phys. **76**(10), 106501 (2013)
31.27 M. Tchernycheva, F.H. Julien, E. Monroy: Review of nitride infrared intersubband devices, Proc. SPIE-Int. Soc. Opt. Eng. **7602**, 76021A (2010)
31.28 S.J. Pearton, R. Deist, F. Ren, L. Liu, A.Y. Polyakov, J. Kim: Review of radiation damage in GaN-based materials and devices, J. Vac. Sci. Technol. A: Vac. Surf. Films **31**(5), 050801 (2013)
31.29 V. Consonni: Self-induced growth of GaN nanowires by molecular beam epitaxy: A critical review of the formation mechanisms, Phys. Status Solidi RRL Rapid Res. Lett. Ger. **7**(10), 699 (2013)
31.30 F. Scholz: Semipolar GaN grown on foreign substrates: A review, Semicond. Sci. Technol. **27**(2), 024002 (2012)
31.31 H. Amano: Progress and prospect of the growth of wide-band-gap group III nitrides: Development of the growth method for single-crystal bulk GaN, Jpn. J. Appl. Phys. **52**(5), 050001 (2013)
31.32 W.A. Harrison: *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond* (Dover, New York 1989)
31.33 C.-Y. Yeh, Z. Lu, S. Froyen, A. Zunger: Zinc-blende–wurtzite polytypism in semiconductors, Phys. Rev. B **46**(16), 10086 (1992)
31.34 G.A. Jeffrey, G.S. Parry, R.L. Mozzi: Study of the Wurtzite-Type Binary Compounds. I. Structures of Aluminum Nitride and Beryllium Oxide, J. Chem. Phys. **25**(5), 1024 (1956)
31.35 H. Schulz, K.H. Thiemann: Crystal structure refinement of AlN and GaN, Solid State Commun. **23**(11), 815 (1977)
31.36 A. Zubrilov: *Properties of Advanced Semiconductor Materials: GaN, AlN, InN, BN, SiC, SiGe* (Wiley, New York 2001)
31.37 E. Dimakis, E. Iliopoulos, K. Tsagaraki, A. Adikimenakis, A. Georgakilas: Biaxial strain and lattice constants of InN (0001) films grown by plasma-assisted molecular beam epitaxy, Appl. Phys. Lett. **88**(19), 191918 (2006)
31.38 M.F. Wu, S.Q. Zhou, A. Vantomme, Y. Huang, H. Wang, H. Yang: High-precision determination of lattice constants and structural characterization of InN thin films, J. Vac. Sci. Technol. A: Vac. Surf. Films **24**(2), 275 (2006)
31.39 Q. Xia, H. Xia, A.L. Ruoff: Pressure-induced rock-salt phase of aluminum nitride: A metastable structure at ambient condition, J. Appl. Phys. **73**(12), 8198 (1993)
31.40 P. Perlin, C. Jauberthie-Carillon, J. Itie, A. San Miguel, I. Grzegory, A. Polian: Raman scattering and x-ray-absorption spectroscopy in gallium nitride under high pressure, Phys. Rev. B **45**(1), 83 (1992)
31.41 M. Ueno, M. Yoshida, A. Onodera, O. Shimomura, K. Takemura: Stability of the wurtzite-type structure under high pressure: GaN and InN, Phys. Rev. B **49**(1), 14 (1994)
31.42 M.J. Paisley: Growth of cubic phase gallium nitride by modified molecular-beam epitaxy, J. Vac. Sci. Technol. Vac. Surf. Films **7**(3), 701 (1989)
31.43 T. Lei, M. Fanciulli, R.J. Molnar, T.D. Moustakas, R.J. Graham, J. Scanlon: Epitaxial growth of zinc blende and wurtzitic gallium nitride thin films on (001) silicon, Appl. Phys. Lett. **59**(8), 944 (1991)
31.44 R.C. Powell, N.-E. Lee, Y.-W. Kim, J.E. Greene: Heteroepitaxial wurtzite and zinc-blende structure GaN grown by reactive-ion molecular-beam epitaxy: Growth kinetics, microstructure, and properties, J. Appl. Phys. **73**(1), 189 (1993)
31.45 M. Mizuta, S. Fujieda, Y. Matsumoto, T. Kawamura: Low Temperature Growth of GaN and AlN on GaAs Utilizing Metalorganics and Hydrazine, Jpn. J. Appl. Phys. **25**(12), L945 (1986)
31.46 M. Leszczynski, J. Bak-Misiuk, J. Domagala, J. Muszalski, M. Kaniewska, J. Marczewski: Lattice dilation by free electrons in heavily doped GaAs:Si, Appl. Phys. Lett. **67**(4), 539 (1995)
31.47 G. Cargill, A. Segmüller, T. Kuech, T. Theis: Lattice strain from DX centers and persistent photocarriers in Sn-doped and Si-doped $Ga_{1-x}Al_xAs$, Phys. Rev. B **46**(16), 10078 (1992)
31.48 M. Leszczynski, H. Teisseyre, T. Suski, I. Grzegory, M. Bockowski, J. Jun, S. Porowski, K. Pakula, J.M. Baranowski, C.T. Foxon, T.S. Cheng: Lattice parameters of gallium nitride, Appl. Phys. Lett. **69**(1), 73 (1996)
31.49 J. Härtwig, S. Grosswig: Measurement of x-ray diffraction angles of perfect monocrystals with high accuracy using a single crystal diffractometer, Phys. Status Solidi A **115**(2), 369 (1989)
31.50 W.L. Bond: Precision lattice constant determination, Acta Crystallogr. **13**(10), 814 (1960)
31.51 T. Nagashima, Y. Kubota, T. Kinoshita, Y. Kumagai, J. Xie, R. Collazo, H. Murakami, H. Okamoto, A. Koukitu, Z. Sitar: Structural and optical properties of carbon-doped AlN substrates grown by hydride vapor phase epitaxy using AlN substrates prepared by physical vapor transport, Appl. Phys. Express **5**(12), 125501 (2012)
31.52 P. Lu, R. Collazo, R.F. Dalmau, G. Durkaya, N. Dietz, B. Raghothamachar, M. Dudley, Z. Sitar: Seeded growth of AlN bulk crystals in *m*- and *c*-orientation, J. Cryst. Growth **312**(1), 58 (2009)

31.53 Y. Kumagai, Y. Kubota, T. Nagashima, T. Kinoshita, R. Dalmau, R. Schlesser, B. Moody, J. Xie, H. Murakami, A. Koukitu, Z. Sitar: Preparation of a freestanding AlN substrate from a thick AlN layer grown by hydride vapor phase epitaxy on a bulk AlN substrate prepared by physical vapor transport, Appl. Phys. Express **5**(5), 055504 (2012)

31.54 J. Domagala, M. Leszczynski, P. Prystawko, T. Suski, R. Langer, A. Barski, M. Bremser: Strain relaxation of $Al_xGa_{1-x}N$ epitaxial layers on GaN and SiC substrates, J. Alloys Compd. **286**(1/2), 284 (1999)

31.55 A. Zoroddu, F. Bernardini, P. Ruggerone, V. Fiorentini: First-principles prediction of structure, energetics, formation enthalpy, elastic constants, polarization, and piezoelectric constants of AlN, GaN, and InN: Comparison of local and gradient-corrected density-functional theory, Phys. Rev. B **64**(4), 045208 (2001)

31.56 T. Paskova, E.A. Preble, A.D. Hanser, K.R. Evans, R. Kröger, P.P. Paskov, A.J. Cheng, M. Park, J.A. Grenko, M.A.L. Johnson: Polar and nonpolar HVPE GaN substrates: Impact of doping on the structural, electrical and optical characteristics, Phys. Status Solidi C **6**(S2), S344 (2009)

31.57 D. Gogova, D. Siche, A. Kwasniewski, M. Schmidbauer, R. Fornari, C. Hemmingsson, R. Yakimova, B. Monemar: HVPE GaN substrates: Growth and characterization, Phys. Status Solidi C **7**(7/8), 1756 (2010)

31.58 D. Gogova, P.P. Petrov, M. Buegler, M.R. Wagner, C. Nenstiel, G. Callsen, M. Schmidbauer, R. Kucharski, M. Zajac, R. Dwilinski, M.R. Phillips, A. Hoffmann, R. Fornari: Structural and optical investigation of non-polar (1–100) GaN grown by the ammonothermal method, J. Appl. Phys. **113**(20), 203513 (2013)

31.59 R. Kucharski, M. Zajac, R. Doradziński, M. Rudziński, R. Kudrawiec, R. Dwiliński: Non-polar and semi-polar ammonothermal GaN substrates, Semicond. Sci. Technol. **27**(2), 024007 (2012)

31.60 T. Deguchi, D. Ichiryu, K. Toshikawa, K. Sekiguchi, T. Sota, R. Matsuo, T. Azuhata, M. Yamaguchi, T. Yagi, S. Chichibu, S. Nakamura: Structural and vibrational properties of GaN, J. Appl. Phys. **86**(4), 1860 (1999)

31.61 M. Leszczynski, P. Prystawko, T. Suski, B. Lucznik, J. Domagala, J. Bak-Misiuk, A. Stonert, A. Turos, R. Langer, A. Barski: Lattice parameters of GaN single crystals, homoepitaxial layers and heteroepitaxial layers on sapphire, J. Alloys Compd. **286**(1/2), 271 (1999)

31.62 T. Detchprohm, K. Hiramatsu, K. Itoh, I. Akasaki: Relaxation process of the thermal strain in the GaN/α-Al_2O_3 heterostructure and determination of the intrinsic lattice constants of GaN free from the strain, Jpn. J. Appl. Phys. **31**(10B), L1454 (1992)

31.63 H. Angerer, D. Brunner, F. Freudenberg, O. Ambacher, M. Stutzmann, R. Höpler, T. Metzger, E. Born, G. Dollinger, A. Bergmaier, S. Karsch, H.-J. Körner: Determination of the Al mole fraction and the band gap bowing of epitaxial $Al_xGa_{1-x}N$ films, Appl. Phys. Lett. **71**(11), 1504 (1997)

31.64 C. Stampfl, C. Van de Walle: Density-functional calculations for III–V nitrides using the local-density approximation and the generalized gradient approximation, Phys. Rev. B **59**(8), 5521 (1999)

31.65 W. Paszkowicz, J. Adamczyk, S. Krukowski, M. Leszczyński, S. Porowski, J.A. Sokolowski, M. Michalec, W. Łasocha: Lattice parameters, density and thermal expansion of InN microcrystals grown by the reaction of nitrogen plasma with liquid indium, Philos. Mag. A **79**(5), 1145 (1999)

31.66 I. Petrov, E. Mojab, R.C. Powell, J.E. Greene, L. Hultman, J.-E. Sundgren: Synthesis of metastable epitaxial zinc-blende-structure AlN by solid-state reaction, Appl. Phys. Lett. **60**(20), 2491 (1992)

31.67 M.E. Sherwin, T.J. Drummond: Predicted elastic constants and critical layer thicknesses for cubic phase AlN, GaN, and InN on β-SiC, J. Appl. Phys. **69**(12), 8423 (1991)

31.68 H. Xia, Q. Xia, A. Ruoff: High-pressure structure of gallium nitride: Wurtzite-to-rocksalt phase transition, Phys. Rev. B **47**(19), 12925 (1993)

31.69 H. Vollstädt, E. Ito, M. Akaishi, S. Akimoto, O. Fukunaga: High pressure synthesis of rocksalt type of AlN, Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. **66**(1), 7 (1990)

31.70 R. Dwiliński, R. Doradziński, J. Garczyński, L. Sierzputowski, J. Baranowski, M. Kamińska: AMMONO method of GaN and AlN production, Diam. Relat. Mater. **7**(9), 1348 (1998)

31.71 T. Hashimoto, K. Fujito, F. Wu, B.A. Haskell, P.T. Fini, J.S. Speck, S. Nakamura: Structural characterization of thick GaN films grown on freestanding GaN seeds by the ammonothermal method using basic ammonia, Jpn. J. Appl. Phys. **44**(25), L797 (2005)

31.72 B. Wang, M.J. Callahan: Transport growth of GaN crystals by the ammonothermal technique using various nutrients, J. Cryst. Growth **291**(2), 455 (2006)

31.73 R. Dwiliński, R. Doradziński, J. Garczyński, L.P. Sierzputowski, A. Puchalski, Y. Kanbara, K. Yagi, H. Minakuchi, H. Hayashi: Excellent crystallinity of truly bulk ammonothermal GaN, J. Cryst. Growth **310**(17), 3911 (2008)

31.74 T. Paskova, K.R. Evans: GaN substrates – Progress, status, and prospects, IEEE J. Sel. Top. Quantum Electron. **15**(4), 1041 (2009)

31.75 C. van de Walle: Effects of impurities on the lattice parameters of GaN, Phys. Rev. B **68**(16), 165209 (2003)

31.76 X.H. Zheng, Y.T. Wang, Z.H. Feng, H. Yang, H. Chen, J.M. Zhou, J.W. Liang: Method for measurement of lattice parameter of cubic GaN layers on GaAs (001), J. Cryst. Growth **250**(3/4), 345 (2003)

31.77 S.V. Novikov, N. Zainal, A.V. Akimov, C.R. Staddon, A.J. Kent, C.T. Foxon: Molecular beam epitaxy as a method for the growth of freestanding zinc-blende (cubic) GaN layers and substrates, J. Vac.

Sci. Technol. B **28**(3), C3B1 (2010)
31.78 A. Muoz, K. Kunc: High-pressure phase of gallium nitride, Phys. Rev. B **44**(18), 10372 (1991)
31.79 J.G. Lozano, F.M. Morales, R. García, D. González, V. Lebedev, C.Y. Wang, V. Cimalla, O. Ambacher: Cubic InN growth on sapphire (0001) using cubic indium oxide as buffer layer, Appl. Phys. Lett. **90**(9), 091901 (2007)
31.80 A. Lima, A. Tabata, J. Leite, S. Kaiser, D. Schikora, B. Schöttker, T. Frey, D. As, K. Lischka: Growth of cubic InN on InAs (001) by plasma-assisted molecular beam epitaxy, J. Cryst. Growth **201/202**, 396 (1999)
31.81 D. Chandrasekhar, D.J. Smith, S. Strite, M.E. Lin, H. Morkoç: Characterization of Group III-nitride semiconductors by high-resolution electron microscopy, J. Cryst. Growth **152**(3), 135 (1995)
31.82 V. Cimalla, J. Pezoldt, G. Ecke, R. Kosiba, O. Ambacher, L. Spieß, G. Teichert, H. Lu, W.J. Schaff: Growth of cubic InN on *r*-plane sapphire, Appl. Phys. Lett. **83**(17), 3468 (2003)
31.83 F.D. Murnaghan: The compressibility of media under extreme pressures, Proc. Natl. Acad. Sci. **30**(9), 244 (1944)
31.84 A. Polian, M. Grimsditch, I. Grzegory: Elastic constants of gallium nitride, J. Appl. Phys. **79**(6), 3343 (1996)
31.85 M. Kazan, E. Moussaed, R. Nader, P. Masri: Elastic constants of aluminum nitride, Phys. Status Solidi C **4**(1), 204 (2007)
31.86 Y. Ohashi, M. Arakawa, J. Kushibiki, B.M. Epelbaum, A. Winnacker: Ultrasonic microspectroscopy characterization of AlN single crystals, Appl. Phys. Express **1**, 077004 (2008)
31.87 K. Tsubouchi, K. Sugai, N. Mikoshiba: AlN material constants evaluation and SAW properties on AlN/Al2O3 and AlN/Si, IEEE 1981 Ultrasonics Symp. (1981) p. 375
31.88 A.F. Wright: Elastic properties of zinc-blende and wurtzite AlN, GaN, and InN, J. Appl. Phys. **82**(6), 2833 (1997)
31.89 K. Kim, W. Lambrecht, B. Segall: Elastic constants and related properties of tetrahedrally bonded BN, AlN, GaN, and InN, Phys. Rev. B **53**(24), 16310 (1996)
31.90 Y. Taniyasu, M. Kasu, T. Makimoto: Threading dislocations in heteroepitaxial AlN layer grown by MOVPE on SiC (0001) substrate, J. Cryst. Growth **298**, 310 (2007)
31.91 R. Thokala, J. Chaudhuri: Calculated elastic constants of wide band gap semiconductor thin films with a hexagonal crystal structure for stress problems, Thin Solid Films **266**(2), 189 (1995)
31.92 L.E. McNeil, M. Grimsditch, R.H. French: Vibrational spectroscopy of Aluminum Nitride, J. Am. Ceram. Soc. **76**(5), 1132 (1993)
31.93 M. Ueno, A. Onodera, O. Shimomura, K. Takemura: X-ray observation of the structural phase transition of aluminum nitride under high pressure, Phys. Rev. B **45**(17), 10123 (1992)
31.94 D. Gerlich, S.L. Dole, G.A. Slack: Elastic properties of aluminum nitride, J. Phys. Chem. Solids **47**(5), 437 (1986)
31.95 P. Van Camp, V. Van Doren, J. Devreese: High-pressure properties of wurtzite- and rocksalt-type aluminum nitride, Phys. Rev. B **44**(16), 9056 (1991)
31.96 E. Ruiz, S. Alvarez, P. Alemany: Electronic structure and properties of AlN, Phys. Rev. B **49**(11), 7115 (1994)
31.97 I. Yonenaga, T. Shima, M.H.F. Sluiter: Nano-indentation hardness and elastic moduli of bulk single-crystal AlN, Jpn. J. Appl. Phys. **41**(7A), 4620 (2002)
31.98 S.-R. Jian, J.-Y. Juang: Indentation-induced mechanical deformation behaviors of AlN thin films deposited on *c*-plane sapphire, J. Nanomater. **2012**, 1 (2012)
31.99 M. Fujikane, M. Leszczyński, S. Nagao, T. Nakayama, S. Yamanaka, K. Niihara, R. Nowak: Elastic–plastic transition during nanoindentation in bulk GaN crystal, J. Alloys Compd. **450**(1/2), 405 (2008)
31.100 M.A. Moram, Z.H. Barber, C.J. Humphreys: Accurate experimental determination of the Poisson's ratio of GaN using high-resolution x-ray diffraction, J. Appl. Phys. **102**, 023505 (2007)
31.101 Z. Yang, R.N. Wang, S. Jia, D. Wang, B.S. Zhang, K.M. Lau, K.J. Chen: Mechanical characterization of suspended GaN microstructures fabricated by GaN-on-patterned-silicon technique, Appl. Phys. Lett. **88**(4), 041913 (2006)
31.102 V.A. Savastenko, A.U. Sheleg: Study of the elastic properties of gallium nitride, Phys. Status Solidi A **48**(2), K135 (1978)
31.103 R.B. Schwarz, K. Khachaturyan, E.R. Weber: Elastic moduli of gallium nitride, Appl. Phys. Lett. **70**(9), 1122 (1997)
31.104 C. Deger, E. Born, H. Angerer, O. Ambacher, M. Stutzmann, J. Hornsteiner, E. Riha, G. Fischerauer: Sound velocity of $Al_xGa_{1-x}N$ thin films obtained by surface acoustic-wave measurements, Appl. Phys. Lett. **72**(19), 2400 (1998)
31.105 C.-Y. Nam, P. Jaroenapibal, D. Tham, D.E. Luzzi, S. Evoy, J.E. Fischer: Diameter-dependent electromechanical properties of GaN nanowires, Nano Lett. **6**(2), 153 (2006)
31.106 J.-E. Jørgensen, J.M. Jakobsen, J.Z. Jiang, L. Gerward, J.S. Olsen: High-pressure x-ray diffraction study of bulk- and nanocrystalline GaN, J. Appl. Crystallogr. **36**(3), 920 (2003)
31.107 R. Nowak, M. Pessa, M. Suganuma, M. Leszczynski, I. Grzegory, S. Porowski, F. Yoshida: Elastic and plastic properties of GaN determined by nano-indentation of bulk crystal, Appl. Phys. Lett. **75**(14), 2070 (1999)
31.108 M. Fujikane, T. Yokogawa, S. Nagao, R. Nowak: Yield shear stress dependence on nanoindentation strain rate in bulk GaN crystal, Phys. Status Solidi C **8**(2), 429 (2011)
31.109 M. Martyniuk, G. Parish, H. Marchand, P.T. Fini, S.P. DenBaars, L. Faraone: Nanoindentation of laterally overgrown epitaxial gallium nitride, Electron. Mater. Lett. **8**(2), 111 (2012)

31.110 X. Wang, S.-B. Che, Y. Ishitani, A. Yoshikawa: Effect of epitaxial temperature on N-polar InN films grown by molecular beam epitaxy, J. Appl. Phys. **99**(7), 073512 (2006)

31.111 Y.-S. Lu, C.-H. Hsieh, S. Gwo, M.T. Hou, J.-S. Yao, J.A. Yeh: An investigation of the Young's modulus of single-crystalline wurtzite indium nitride using an atomic force microscopy based micromechanical bending test, Appl. Phys. Lett. **101**(22), 221906 (2012)

31.112 F. Peng, Y. Han, H. Fu, X. Yang: First-principles calculations on structure and elasticity of wurtzite-type indium nitride under pressure, J. Alloys Compd. **475**(1/2), 885 (2009)

31.113 J. Serrano, A. Bosak, M. Krisch, F.J. Manjón, A.H. Romero, N. Garro, X. Wang, A. Yoshikawa, M. Kuball: In N thin film lattice dynamics by grazing incidence inelastic x-ray scattering, Phys. Rev. Lett. **106**(20), 205501 (2011)

31.114 S.-R. Jian, C.-Y. Huang, W.-C. Ke: Nanoindentation responses of InN thin films, J. Alloys Compd. **609**, 125 (2014)

31.115 S. Kataria, T.-W. Liu, C.-L. Hsiao, S. Dhara, L.-C. Chen, K.-H. Chen, S. Dash, A.K. Tyagi: Growth orientation dependent hardness for epitaxial Wurtzite in N films, J. Nanosci. Nanotechnol. **10**(8), 5170 (2010)

31.116 C.-H. Tsai: Nanoindention study of indium nitride thin films grown using RF plasma-assisted molecular beam epitaxy, Vacuum **86**(9), 1328 (2012)

31.117 J.H. Edgar, C.H. Wei, D.T. Smith, T.J. Kistenmacher, W.A. Bryden: Hardness, elastic modulus and structure of indium nitride thin films on AlN-nucleated sapphire substrates, J. Mater. Sci. Mater. Electron. **8**(5), 307 (1997)

31.118 P.E. Van Camp, V.E. Van Doren, J.T. Devreese: High pressure structural phase transformation in gallium nitride, Solid State Commun. **81**(1), 23 (1992)

31.119 I. Yonenaga, K. Motoki: Yield strength and dislocation mobility in plastically deformed bulk single-crystal GaN, J. Appl. Phys. **90**(12), 6539 (2001)

31.120 M. Yamaguchi, T. Yagi, T. Azuhata, T. Sota, K. Suzuki, S. Chichibu, S. Nakamura: Brillouin scattering study of gallium nitride: Elastic stiffness constants, J. Phys. Condens. Matter **9**(1), 241 (1997)

31.121 I. Yonenaga: High-temperature strength of III–V nitride crystals, J. Phys. Condens. Matter **14**(48), 12947 (2002)

31.122 I. Yonenaga: High-temperature strength of bulk single crystals of III–V nitrides, J. Mater. Sci. Mater. Electron. **14**(5–7), 279 (2003)

31.123 G. Kaczmarczyk, A. Kaschner, S. Reich, A. Hoffmann, C. Thomsen, D.J. As, A.P. Lima, D. Schikora, K. Lischka, R. Averbeck, H. Riechert: Lattice dynamics of hexagonal and cubic InN: Raman-scattering experiments and calculations, Appl. Phys. Lett. **76**(15), 2122 (2000)

31.124 T. Azuhata, T. Sota, K. Suzuki, S. Nakamura: Polarized Raman spectra in GaN, J. Phys. Condens. Matter **7**(10), L129 (1995)

31.125 M. Kazan, C. Zgheib, E. Moussaed, P. Masri: Temperature dependence of Raman-active modes in AlN, Diam. Relat. Mater. **15**(4–8), 1169 (2006)

31.126 A. Link, K. Bitzer, W. Limmer, R. Sauer, C. Kirchner, V. Schwegler, M. Kamp, D.G. Ebling, K.W. Benz: Temperature dependence of the E_2 and A_1(LO) phonons in GaN and AlN, J. Appl. Phys. **86**(11), 6256 (1999)

31.127 D.Y. Song, M. Holtz, A. Chandolu, S.A. Nikishin, E.N. Mokhov, Y. Makarov, H. Helava: Optical phonon decay in bulk aluminum nitride, Appl. Phys. Lett. **89**(2), 021901 (2006)

31.128 J. Senawiratne, M. Strassburg, N. Dietz, U. Haboeck, A. Hoffmann, V. Noveski, R. Dalmau, R. Schlesser, Z. Sitar: Raman, photoluminescence and absorption studies on high quality AlN single crystals, Phys. Status Solidi C **2**(7), 2774 (2005)

31.129 M. Bickermann, B.M. Epelbaum, P. Heimann, Z.G. Herro, A. Winnacker: Orientation-dependent phonon observation in single-crystalline aluminum nitride, Appl. Phys. Lett. **86**(13), 131904 (2005)

31.130 M. Strassburg, J. Senawiratne, N. Dietz, U. Haboeck, A. Hoffmann, V. Noveski, R. Dalmau, R. Schlesser, Z. Sitar: The growth and optical properties of large, high-quality AlN single crystals, J. Appl. Phys. **96**(10), 5870 (2004)

31.131 K. Shimada, T. Sota, K. Suzuki: First-principles study on electronic and elastic properties of BN, AlN, and GaN, J. Appl. Phys. **84**(9), 4951 (1998)

31.132 K. Miwa, A. Fukumoto: First-principles calculation of the structural, electronic, and vibrational properties of gallium nitride and aluminum nitride, Phys. Rev. B **48**(11), 7897 (1993)

31.133 K. Fujito, K. Kiyomi, T. Mochizuki, H. Oota, H. Namita, S. Nagao, I. Fujimura: High-quality nonpolar m-plane GaN substrates grown by HVPE, Phys. Status Solidi A **205**(5), 1056 (2008)

31.134 H. Gao, F. Yan, H. Zhang, J. Li, J. Wang, J. Yan: First and second order Raman scattering spectroscopy of nonpolar *a*-plane GaN, J. Appl. Phys. **101**(10), 103533 (2007)

31.135 V. Darakchieva, T. Paskova, M. Schubert, H. Arwin, P.P. Paskov, B. Monemar, D. Hommel, M. Heuken, J. Off, F. Scholz, B.A. Haskell, P.T. Fini, J.S. Speck, S. Nakamura: Anisotropic strain and phonon deformation potentials in GaN, Phys. Rev. B **75**(19), 195217 (2007)

31.136 A. Goñi, H. Siegle, K. Syassen, C. Thomsen, J.-M. Wagner: Effect of pressure on optical phonon modes and transverse effective charges in GaN and AlN, Phys. Rev. B **64**(3), 035205 (2001)

31.137 J.-M. Wagner, F. Bechstedt: Properties of strained wurtzite GaN and AlN: Ab initio studies, Phys. Rev. B **66**(11), 115202 (2002)

31.138 D. Gogova, A. Kasic, H. Larsson, B. Pécz, R. Yakimova, B. Magnusson, B. Monemar, F. Tuomisto, K. Saarinen, C. Miskys, M. Stutzmann, C. Bundesmann, M. Schubert: Optical and structural characteristics of virtually unstrained bulk-like GaN, Jpn. J. Appl. Phys. **43**(4A), 1264 (2004)

31.139 D. Huang, F. Yun, M.A. Reshchikov, D. Wang, H. Morkoç, D.L. Rode, L.A. Farina, Ç. Kurdak, K.T. Tsen, S.S. Park, K.Y. Lee: Hall mobility and carrier concentration in free-standing high quality GaN templates grown by hydride vapor phase epitaxy, Solid-State Electron. **45**(5), 711 (2001)

31.140 X. Wang, S.-B. Che, Y. Ishitani, A. Yoshikawa: Experimental determination of strain-free Raman frequencies and deformation potentials for the E2 high and A1(LO) modes in hexagonal InN, Appl. Phys. Lett. **89**(17), 171907 (2006)

31.141 C. Pinquier, F. Demangeot, J. Frandon, J.-C. Chervin, A. Polian, B. Couzinet, P. Munsch, O. Briot, S. Ruffenach, B. Gil, B. Maleyre: Raman scattering study of wurtzite and rocksalt InN under high pressure, Phys. Rev. B **73**(11), 115211 (2006)

31.142 X.D. Pu, J. Chen, W.Z. Shen, H. Ogawa, Q.X. Guo: Temperature dependence of Raman scattering in hexagonal indium nitride films, J. Appl. Phys. **98**(3), 033527 (2005)

31.143 V. Darakchieva, P.P. Paskov, E. Valcheva, T. Paskova, B. Monemar, M. Schubert, H. Lu, W.J. Schaff: Deformation potentials of the E_1(TO) and E_2 modes of InN, Appl. Phys. Lett. **84**(18), 3636 (2004)

31.144 V.Y. Davydov, V.V. Emtsev, I.N. Goncharuk, A.N. Smirnov, V.D. Petrikov, V.V. Mamutin, V.A. Vekshin, S.V. Ivanov, M.B. Smirnov, T. Inushima: Experimental and theoretical studies of phonons in hexagonal InN, Appl. Phys. Lett. **75**(21), 3297 (1999)

31.145 J.S. Dyck, K. Kim, S. Limpijumnong, W.R.L. Lambrecht, K. Kash, J.C. Angus: Identification of Raman-active phonon modes in oriented platelets of InN and polycrystalline InN, Solid State Commun. **114**(7), 355 (2000)

31.146 T. Prokofyeva, M. Seon, J. Vanbuskirk, M. Holtz, S. Nikishin, N. Faleev, H. Temkin, S. Zollner: Vibrational properties of AlN grown on (111)-oriented silicon, Phys. Rev. B **63**(12), 125313 (2001)

31.147 D.G. Zhao, S.J. Xu, M.H. Xie, S.Y. Tong, H. Yang: Stress and its effect on optical properties of GaN epilayers grown on Si(111), 6H-SiC(0001), and *c*-plane sapphire, Appl. Phys. Lett. **83**(4), 677 (2003)

31.148 W. Götz, N.M. Johnson, D.P. Bour, M.D. McCluskey, E.E. Haller: Local vibrational modes of the Mg–H acceptor complex in GaN, Appl. Phys. Lett. **69**(24), 3725 (1996)

31.149 C. Van de Walle: Interactions of hydrogen with native defects in GaN, Phys. Rev. B **56**(16), R10020 (1997)

31.150 D. Gogova, A. Kasic, H. Larsson, C. Hemmingsson, B. Monemar, F. Tuomisto, K. Saarinen, L. Dobos, B. Pécz, P. Gibart, B. Beaumont: Strain-free bulk-like GaN grown by hydride-vapor-phase-epitaxy on two-step epitaxial lateral overgrown GaN template, J. Appl. Phys. **96**(1), 799 (2004)

31.151 C. Pinquier, F. Demangeot, J. Frandon, J.W. Pomeroy, M. Kuball, H. Hubel, N.W.A. van Uden, D.J. Dunstan, O. Briot, B. Maleyre, S. Ruffenach, B. Gil: Raman scattering in hexagonal InN under high pressure, Phys. Rev. B **70**(11), 113202 (2004)

31.152 J.W. Pomeroy, M. Kuball, H. Lu, W.J. Schaff, X. Wang, A. Yoshikawa: Phonon lifetimes and phonon decay in InN, Appl. Phys. Lett. **86**(22), 223501 (2005)

31.153 S. Figge, H. Kröncke, D. Hommel, B.M. Epelbaum: Temperature dependence of the thermal expansion of AlN, Appl. Phys. Lett. **94**(10), 101915 (2009)

31.154 K. Wang, R.R. Reeber: Thermal expansion Of GaN and AlN, MRS Proc. **482**, 863 (1997)

31.155 K.M. Taylor, C. Lenie: Some properties of Aluminum Nitride, J. Electrochem. Soc. **107**(4), 308 (1960)

31.156 S.N. Ivanov, P.A. Popov, G.V. Egorov, A.A. Sidorov, B.I. Kornev, L.M. Zhukova, V.P. Ryabov: Thermophysical properties of aluminum nitride ceramic, Phys. Solid State **39**(1), 81 (1997)

31.157 W. Paszkowicz, M. Knapp, S. Podsiadlo, G. Kamler, J.B. Pelka: Lattice parameters of aluminium nitride in the range 10–291 K, Acta Phys. Pol. A **101**(5), 781 (2002)

31.158 G.A. Slack, S.F. Bartram: Thermal expansion of some diamondlike crystals, J. Appl. Phys. **46**(1), 89 (1975)

31.159 W.M. Yim, R.J. Paff: Thermal expansion of AlN, sapphire, and silicon, J. Appl. Phys. **45**(3), 1456 (1974)

31.160 H.P. Maruska: The preparation and properties of vapor-deposited single-crystal-line GaN, Appl. Phys. Lett. **15**(10), 327 (1969)

31.161 M. Leszczynski, T. Suski, H. Teisseyre, P. Perlin, I. Grzegory, J. Jun, S. Porowski, T.D. Moustakas: Thermal expansion of gallium nitride, J. Appl. Phys. **76**(8), 4909 (1994)

31.162 V. Kirchner, H. Heinke, D. Hommel, J.Z. Domagala, M. Leszczynski: Thermal expansion of bulk and homoepitaxial GaN, Appl. Phys. Lett. **77**(10), 1434 (2000)

31.163 C. Roder, S. Einfeldt, S. Figge, D. Hommel: Temperature dependence of the thermal expansion of GaN, Phys. Rev. B **72**(8), 085218 (2005)

31.164 R.R. Reeber, K. Wang: Lattice parameters and thermal expansion of GaN, J. Mater. Res. **15**(1), 40 (2000)

31.165 A.U. Sheleg, V.A. Sevastenko: Investigation of thermal expansion of indium and gallium nitrides, Vestsi Akad Navuk BSSR **3**, 126 (1976)

31.166 K. Wang, R.R. Reeber: Thermal expansion and elastic properties of InN, Appl. Phys. Lett. **79**(11), 1602 (2001)

31.167 W. Paszkowicz, R. Černý, S. Krukowski: Rietveld refinement for indium nitride in the 105–295 K range, Powder Diffr. **18**(2), 114 (2003)

31.168 G.A. Slack: Nonmetallic crystals with high thermal conductivity, J. Phys. Chem. Solids **34**(2), 321 (1973)

31.169 G.A. Slack, T.F. McNelly: Ain single crystals, J. Cryst. Growth **42**, 560 (1977)

31.170 A. Witek: Some aspects of thermal conductivity of isotopically pure diamond – A comparison with nitrides, Diam. Relat. Mater. **7**(7), 962 (1998)

31.171 G.A. Slack, L.J. Schowalter, D. Morelli, J.A. Freitas: Some effects of oxygen impurities on AlN and GaN,

J. Cryst. Growth **246**(3/4), 287 (2002)
31.172 M. Bickermann, B.M. Epelbaum, A. Winnacker: Characterization of bulk AlN with low oxygen content, J. Cryst. Growth **269**(2–4), 432 (2004)
31.173 A. AlShaikhi, G.P. Srivastava: Thermal conductivity of single crystal and ceramic AlN, J. Appl. Phys. **103**(8), 083554 (2008)
31.174 A. Nikolaev, I. Nikitina, A. Zubrilov, M. Mynbaeva, Y. Melnik, V. Dmitriev: AlN wafers fabricated by hydride vapor phase epitaxy, MRS Proc. **595**, 6.5 (1999)
31.175 D.I. Florescu, V.M. Asnin, F.H. Pollak: Thermal conductivity measurements of GaN and AlN, Compd. Semicond. **7**(2), 62 (2001)
31.176 L. Qiao, H. Zhou, C. Li: Microstructure and thermal conductivity of spark plasma sintering AlN ceramics, Mater. Sci. Eng. B **99**(1–3), 102 (2003)
31.177 K. Watari, K. Ishizaki, F. Tsuchiya: Phonon scattering and thermal conduction mechanisms of sintered aluminium nitride ceramics, J. Mater. Sci. **28**(14), 3709 (1993)
31.178 E. Richter, M. Gründer, B. Schineller, F. Brunner, U. Zeimer, C. Netzel, M. Weyers, G. Tränkle: GaN boules grown by high rate HVPE, Phys. Status Solidi C **8**(5), 1450 (2011)
31.179 A. Jeżowski, P. Stachowiak, T. Plackowski, T. Suski, S. Krukowski, M. Boćkowski, I. Grzegory, B. Danilchenko, T. Paszkiewicz: Thermal conductivity of GaN crystals grown by high pressure method, Phys. Status Solidi B **240**(2), 447 (2003)
31.180 H. Shibata, Y. Waseda, H. Ohta, K. Kiyomi, K. Shimoyama, K. Fujito, H. Nagaoka, Y. Kagamitani, R. Simura, T. Fukuda: High thermal conductivity of gallium nitride (GaN) crystals grown by HVPE process, Mater. Trans. **48**(10), 2782 (2007)
31.181 W. Liu, A.A. Balandin, C. Lee, H.-Y. Lee: Increased thermal conductivity of free-standing low-dislocation-density GaN films, Phys. Status Solidi A **202**(12), R135 (2005)
31.182 E.K. Sichel, J.I. Pankove: Thermal conductivity of GaN, 25–360 K, J. Phys. Chem. Solids **38**(3), 330 (1977)
31.183 A. AlShaikhi, S. Barman, G.P. Srivastava: Theory of the lattice thermal conductivity in bulk and films of GaN, Phys. Rev. B **81**(19), 195320 (2010)
31.184 C. Mion, J.F. Muth, E.A. Preble, D. Hanser: Accurate dependence of gallium nitride thermal conductivity on dislocation density, Appl. Phys. Lett. **89**(9), 092123 (2006)
31.185 D.I. Florescu, V.A. Asnin, L.G. Mourokh, F.H. Pollak, R.J. Molnar: Doping dependence of the thermal conductivity of hydride vapor phase epitaxy grown n-GaN/Sapphire(0001) using a scanning thermal microscope, MRS Proc. **595**, F99W3.89 (1999)
31.186 D.I. Florescu, V.M. Asnin, F.H. Pollak, R.J. Molnar, C.E.C. Wood: High spatial resolution thermal conductivity and Raman spectroscopy investigation of hydride vapor phase epitaxy grown n-GaN/Sapphire(0001): Doping dependence, J. Appl. Phys. **88**(6), 3295 (2000)
31.187 S. Krukowski, A. Witek, J. Adamczyk, J. Jun, M. Bockowski, I. Grzegory, B. Lucznik, G. Nowak, M. Wróblewski, A. Presz, S. Gierlotka, S. Stelmach, B. Palosz, S. Porowski, P. Zinn: Thermal properties of indium nitride, J. Phys. Chem. Solids **59**(3), 289 (1998)
31.188 A.X. Levander, T. Tong, K.M. Yu, J. Suh, D. Fu, R. Zhang, H. Lu, W.J. Schaff, O. Dubon, W. Walukiewicz, D.G. Cahill, J. Wu: Effects of point defects on thermal and thermoelectric properties of InN, Appl. Phys. Lett. **98**(1), 012108 (2011)
31.189 A.D. Mah, W.W. Weller, A.U. Christensen: Thermodynamic Properties of Aluminum Nitride, Report of Investigation (US Bureau of Mines), Vol. RI-5716 (1961) p. 78
31.190 V.P. Glushko, L.V. Gurevich, G.A. Bergman, I.V. Weitz, V.A. Medvedev, G.A. Chachkurov, V.S. Yungman: *Thermodinamicheskiie Swoistwa Indiwidualnych Weshchestw (the old USSR)*, Vol. 1 (Nauka, Moscow 1979) p. 164
31.191 G.A. Slack, R.A. Tanzilli, R.O. Pohl, J.W. Vandersande: The intrinsic thermal conductivity of AlN, J. Phys. Chem. Solids **48**(7), 641 (1987)
31.192 J. Leitner, A. Strejc, D. Sedmidubský, K. Růžička: High temperature enthalpy and heat capacity of GaN, Thermochim. Acta **401**(2), 169 (2003)
31.193 K. Itagaki, K. Yamaguchi: High temperature heat contents of III–V semiconductor systems, Thermochim. Acta **163**, 1 (1990)
31.194 B.A. Danilchenko, T. Paszkiewicz, S. Wolski, A. Jeżowski, T. Plackowski: Heat capacity and phonon mean free path of wurtzite GaN, Appl. Phys. Lett. **89**(6), 061901 (2006)
31.195 R. Kremer, M. Cardona, E. Schmitt, J. Blumm, S. Estreicher, M. Sanati, M. Bockowski, I. Grzegory, T. Suski, A. Jezowski: Heat capacity of α-GaN: Isotope effects, Phys. Rev. B **72**(7), 075209 (2005)
31.196 J.C. Nipko, C.-K. Loong, C.M. Balkas, R.F. Davis: Phonon density of states of bulk gallium nitride, Appl. Phys. Lett. **73**(1), 34 (1998)
31.197 J. Leitner, P. Maršík, D. Sedmidubský, K. Růžička: High temperature enthalpy, heat capacity and other thermodynamic functions of solid InN, J. Phys. Chem. Solids **65**(6), 1127 (2004)
31.198 I. Zięborak-Tomaszkiewicz, R. Świerzewski, P. Gierycz: Heat capacity of indium nitride, J. Therm. Anal. Calorim. **91**(2), 649 (2008)
31.199 W. Götz, L.T. Romano, J. Walker, N.M. Johnson, R.J. Molnar: Hall-effect analysis of GaN films grown by hydride vapor phase epitaxy, Appl. Phys. Lett. **72**(10), 1214 (1998)
31.200 D.L. Rode, R.K. Willardson, A.C. Beer: *Semiconductors and Semimetals*, Vol. 10 (Academic, New York 1975)
31.201 K. Seeger: *Semiconductor Physics*, 2nd edn. (Springer, Berlin, Heidelberg 1982)
31.202 V.W.L. Chin, T.L. Tansley, T. Osotchan: Electron mobilities in gallium, indium, and aluminum nitrides, J. Appl. Phys. **75**(11), 7365 (1994)
31.203 D.L. Rode, D.K. Gaskill: Electron Hall mobility of n-GaN, Appl. Phys. Lett. **66**(15), 1972 (1995)

31.204 M.E. Lin, B. Sverdlov, G.L. Zhou, H. Morkoc: A comparative study of GaN epilayers grown on sapphire and SiC substrates by plasma-assisted molecular-beam epitaxy, Appl. Phys. Lett. **62**(26), 3479 (1993)

31.205 H.M. Ng, D. Doppalapudi, T.D. Moustakas, N.G. Weimann, L.F. Eastman: The role of dislocation scattering in n-type GaN films, Appl. Phys. Lett. **73**(6), 821 (1998)

31.206 D.C. Look, J.R. Sizelove: Dislocation scattering in GaN, Phys. Rev. Lett. **82**(6), 1237 (1999)

31.207 H.W. Choi, J. Zhang, S.J. Chua: Dislocation scattering in n–GaN, Proc. ICMAT 2001 Symp. Mater. for Optoelectron. High Freq. Electron. Appl (2001) p. 567

31.208 J.Y. Shi, L.P. Yu, Y.Z. Wang, G.Y. Zhang, H. Zhang: Influence of different types of threading dislocations on the carrier mobility and photoluminescence in epitaxial GaN, Appl. Phys. Lett. **80**(13), 2293 (2002)

31.209 N.G. Weimann, L.F. Eastman, D. Doppalapudi, H.M. Ng, T.D. Moustakas: Scattering of electrons at threading dislocations in GaN, J. Appl. Phys. **83**(7), 3656 (1998)

31.210 Z.-Q. Fang, D.C. Look, W. Kim, Z. Fan, A. Botchkarev, H. Morkoc: Deep centers in n–GaN grown by reactive molecular beam epitaxy, Appl. Phys. Lett. **72**(18), 2277 (1998)

31.211 W. Kim, A.E. Botohkarev, H. Morkoc, Z.-Q. Fang, D.C. Look, D.J. Smith: Effect of ammonia flow rate on impurity incorporation and material properties of Si-doped GaN epitaxial films grown by reactive molecular beam epitaxy, J. Appl. Phys. **84**(12), 6680 (1998)

31.212 Q.-S. Zhu, N. Sawaki: Nitrogen vacancy scattering in n-GaN grown by metal-organic vapor phase epitaxy, Appl. Phys. Lett. **76**(12), 1594 (2000)

31.213 Z. Chen, H. Yuan, D.-C. Lu, X. Sun, S. Wan, X. Liu, P. Han, X. Wang, Q. Zhu, Z. Wang: Nitrogen vacancy scattering in GaN grown by metal-organic vapor phase epitaxy, Solid-State Electron. **46**(12), 2069 (2002)

31.214 D.C. Look, R.J. Molnar: Degenerate layer at GaN/Sapphire interface: Influence on Hall-effect measurements, Appl. Phys. Lett. **70**(25), 3377 (1997)

31.215 S. Nakamura: GaN growth using GaN buffer layer, Jpn. J. Appl. Phys. Part 2 **30**(10A), 1705 (1991)

31.216 S. Nakamura, T. Mukai: High-quality InGaN films grown on GaN films, Jpn. J. Appl. Phys. Part 2 **31**(10B), 1457 (1992)

31.217 D. Huang, F. Yun, M.A. Reshchikov, D. Wang, H. Morkoc, D.L. Rode, L.A. Farina, C. Kurdak, K.T. Tsen, S.S. Park, K.Y. Lee: Hall mobility and carrier concentration in free-standing high quality GaN templates grown by hydride vapor phase epitaxy, Solid-State Electron. **45**(5), 711 (2001)

31.218 S. Nakamura, T. Mukai, M. Senoh: Si- and Ge-doped GaN films grown with GaN buffer layers, Jpn. J. Appl. Phys. **31**, 2883 (1992)

31.219 S. Nakamura, T. Mukai, M. Senoh: In situ monitoring and hall measurements of GaN grown with GaN buffer layers, J. Appl. Phys. **71**(11), 5543 (1992)

31.220 S.C. Binari, H.C. Dietrich: III-V Nitride electronic devices. In: *GaN and Related Materials*, ed. by S.J. Pearson (Gordon Breach, New York 1997) pp. 509–534

31.221 F. Yun, M.A. Reshchikov, K. Jones, P. Visconti, H. Morkoc, S.S. Park, K.Y. Lee: Electrical, structural, and optical characterization of free-standing GaN template grown by hydride vapor phase epitaxy, Solid-State Electron. **44**(12), 2225 (2000)

31.222 B. Heying, I. Smorchkova, C. Poblenz, C. Elsass, P. Fini, S. Den Baars, U. Mishra, J.S. Speck: Optimization of the surface morphologies and electron mobilities in GaN grown by plasma-assisted molecular beam epitaxy, Appl. Phys. Lett. **77**(18), 2885 (2000)

31.223 D.D. Koleske, R.L. Henry, M.E. Twigg, J.C. Culbertson, S.C. Binari, A.E. Wickenden, M. Fatemi: Influence of AlN nucleation layer temperature on GaN electronic properties grown on SiC, Appl. Phys. Lett. **80**(23), 4372 (2002)

31.224 W. Gotz, N.M. Johnson, C. Chen, H. Liu, C. Kuo, W. Imler: Activation energies of Si donors in GaN, Appl. Phys. Lett. **68**(22), 3144 (1996)

31.225 D.C. Look, D.C. Reynolds, J.W. Hemsky, J.R. Sizelove, R.L. Jones, R.J. Molnar: Defect donor and acceptor in GaN, Phys. Rev. Lett. **79**(12), 2273 (1997)

31.226 M. Ilegems, H.C. Montgomery: Electrical properties of n-type vapor-grown gallium nitride, J. Phys. Chem. Solids **35**(5), 885 (1973)

31.227 M.A. Littlejohn, J.R. Hauser, T.H. Glisson: Monte Carlo calculation of the velocity-field relationship for gallium nitride, Appl. Phys. Lett. **26**(11), 625 (1975)

31.228 U.V. Bhapkar, M.S. Shur: Monte Carlo calculation of velocity-field characteristics of wurtzite GaN, J. Appl. Phys. **82**(4), 1649 (1997)

31.229 J.D. Albrecht, R.P. Wang, P.P. Ruden, M. Farahmand, K.F. Brennan: Monte Carlo calculation of electron transport properties of bulk AlN, J. Appl. Phys. **83**(3), 1446 (1998)

31.230 S. Dhar, S. Ghosh: Low field electron mobility in GaN, J. Appl. Phys. **86**(5), 2668 (1999)

31.231 J.G. Kim, A.C. Frenkel, T. Liu, R.M. Park: Growth by molecular beam epitaxy and electrical characterization of Si-doped zinc blende GaN films deposited on –SiC coated(001) Si substrates, Appl. Phys. Lett. **65**(1), 91 (1994)

31.232 R.J. Molnar, W. Götz, L.T. Romano, N.M. Johnson: Growth of gallium nitride by hydride vapor-phase epitaxy, J. Cryst. Growth **178**(1/2), 147 (1997)

31.233 S. Yamaguchi, M. Kariya, S. Nitta, T. Takeuchi, C. Wetzel, H. Amano, I. Akasaki: Structural properties of InN on GaN grown by metalorganic vapor-phase epitaxy, J. Appl. Phys. **85**(11), 7682 (1999)

31.234 H. Morkoc: Wurtzite GaN-based heterostructures by molecular beam epitaxy, IEEE J. Sel. Top. Quantum Electron. **4**(3), 537 (1998)

31.235 S. Keller, B.P. Keller, Y.-F. Wu, B. Heying, D. Kapolnek, J.S. Speck, U.K. Mishra, S.P. DenBaars: Influence of sapphire nitridation on properties of gallium nitride grown by metalorganic chemical

vapor deposition, Appl. Phys. Lett. **68**(11), 1525 (1996)
31.236 B. Wang, P. Yu, B. Kucukgok, A.G. Melton, N. Lu, I.T. Ferguson: Characterization of undoped and Si-doped bulk GaN fabricated by hydride vapor phase epitaxy, Phys. Status Solidi C Curr. Top, Solid State Phys. **11**(3/4), 573 (2014)
31.237 T. Sochacki, Z. Bryan, M. Amilusik, M. Bobea, M. Fijalkowski, I. Bryan, B. Lucznik, R. Collazo, J.L. Weyher, R. Kucharski, I. Grzegory, M. Bockowski, Z. Sitar: HVPE-GaN grown on MOCVD-GaN/sapphire template and ammonothermal GaN seeds: Comparison of structural, optical, and electrical properties, J. Cryst. Growth **394**(0), 55 (2014)
31.238 T.T. Mnatsakanov, M.E. Levinshtein, L.I. Pomortseva, S.N. Yurkov, G.S. Simin, M.A. Khan: Carrier mobility model for GaN, Solid-State Electron. **47**(1), 111 (2003)
31.239 D.M. Caughey, R.E. Thomas: Carrier mobilities in silicon empirically related to doping and field, Proceeding IEEE **55**(12), 2192 (1967)
31.240 J. Edwards, K. Kawabe, G. Stevens, R.H. Tredgold: Space charge conduction and electrical behaviour of aluminium nitride single crystals, Solid State Commun. **3**(5), 99 (1965)
31.241 K. Kawabe, R.H. Tredgold, Y. Inuishi: Electrical and optical properties of AIN – A thermostable semiconductor, Electr. Eng. Jpn. **87**(1), 62 (1967)
31.242 G.A. Cox, D.O. Cummins, K. Kawabe, R.H. Tredgold: On the preparation, optical properties and electrical behaviour of aluminium nitride, J. Phys. Chem. Solids **28**(4), 543 (1967)
31.243 W.M. Yim, E.J. Stofko, P.J. Zanzucchi, J.I. Pankove, M. Ettenberg, S.L. Gilbert: Epitaxially grown AIN and its optical band gap, J. Appl. Phys. **44**(1), 292 (1973)
31.244 S. Yoshida, S. Misawa, Y. Fujii, S. Takada, H. Hayakawa, S. Gonda, A. Itoh: Reactive molecular beam epitaxy of aluminium nitride, J. Vac. Sci. Technol. **16**(4), 990 (1979)
31.245 A.G. Gorbatov, V.M. Kamyshoc: Sov. Powder Metall. Met. Ceram. **9**, 917 (1970)
31.246 R.F. Rutz: Ultraviolet electroluminescence in AIN, Appl. Phys. Lett. **28**(7), 379 (1976)
31.247 R.F. Rutz, E.P. Harris, J.J. Cuomo: IBM J. Res. Dev. **17**, 61 (1973)
31.248 A. Fara, F. Bernardini, V. Fiorentini: Theoretical evidence for the semi-insulating character of AIN, J. Appl. Phys. **85**(3), 2001 (1999)
31.249 K. Wongchotigul, N. Chen, D.P. Zhang, X. Tang, M.G. Spencer: Low resistivity aluminum nitride: Carbon (AIN:C) films grown by metal organic chemical vapor deposition, Mater. Lett. **26**(4/5), 223 (1996)
31.250 H.J. Hovel, J.J. Cuomo: Electrical and optical properties or rf-sputtered GaN and InN, Appl. Phys. Lett. **20**(2), 71 (1972)
31.251 T.L. Tansley, C.P. Foley: Electron mobility in indium nitride, Electron. Lett. **20**(25/26), 1066 (1984)
31.252 K.L. Westra, M.J. Brett: Near IR optical properties of sputtered InN films, Thin Solid Films **192**(2), 227 (1990)
31.253 J.S. Morgan, T.J. Kistenmacher, W.A. Bryden, S.A. Ecelberger: Sputter deposition of indium nitride on the (111) face of elemental and compound semiconductors, Proc. Evolution of Thin-Film and Surface Microstructure Symp (1991) p. 383
31.254 T.J. Kistenmacher, W.A. Bryden: Overgrowth of indium nitride thin films on aluminum nitride nucleated (00.1) sapphire by reactive magnetron sputtering, Appl. Phys. Lett. **59**(15), 1844 (1991)
31.255 S. Strite, D. Chandrasekhar, D.J. Smith, J. Sariel, H. Chen, N. Teraguchi, H. Morkoç: Structural properties of InN films grown on GaAs substrates: Observation of the zincblende polytpe, J. Cryst. Growth **127**(1–4), 204 (1993)
31.256 C.R. Abernathy, S.J. Pearton, F. Ren, P.W. Wisk: Growth of InN for ohmic contact formation by electron cyclotron resonance metalorganic molecular-beam epitaxy, J. Vac. Sci. Technol. B **11**(2), 179 (1993)
31.257 W.A. Bryden, S.A. Ecelberger, M.E. Hawley, T.J. Kistenmacher: ECR-assisted reactive magnetron sputtering of InN, Proc. 1994 MRS Spring Meet (1994) p. 497
31.258 T. Maruyama, T. Morishita: Indium nitride thin films prepared by radio-frequency reactive sputtering, J. Appl. Phys. **76**(10), 5809 (1994)
31.259 H. Lu, W.J. Schaff, J. Hwang, H. Wu, W. Yeo, A. Pharkya, L.F. Eastman: Improvement on epitaxial grown of InN by migration enhanced epitaxy, Appl. Phys. Lett. **77**(16), 2548 (2000)
31.260 Y. Saito, N. Teraguchi, A. Suzuki, T. Araki, Y. Nanishi: Growth of high-electron-mobility InN by RF molecular beam epitaxy, Jpn. J. Appl. Phys. Part 2 Lett. **40**(2A), 91 (2001)
31.261 J. Aderhold, V.Y. Davydov, F. Fedler, H. Klausing, D. Mistele, T. Rotter, O. Semchinova, J. Stemmer, J. Graul: InN thin films grown by metalorganic molecular beam epitaxy on sapphire substrates, J. Cryst. Growth **222**(4), 701 (2001)
31.262 H. Lu, W.J. Schaff, J. Hwang, H. Wu, G. Koley, L.F. Eastman: Effect of an AIN buffer layer on the epitaxial growth of InN by molecular-beam epitaxy, Appl. Phys. Lett. **79**(10), 1489 (2001)
31.263 E. Motlan, M. Goldys, T.L. Tansley: Optical and electrical properties of InN grown by radio-frequency reactive sputtering, J. Cryst. Growth **241**(1/2), 165 (2002)
31.264 Y. Saito, T. Yamaguchi, H. Kanazawa, K. Kano, T. Araki, Y. Nanishi, N. Teraguchi, A. Suzuki: Growth of high-quality InN using low-temperature intermediate layers by RF-MBE, Proc. Thirteen. Int. Conf. Cryst. Growth Conj. Unction Elev. Int. Conf. Vap. Growth Epitaxy (2002) p. 1017
31.265 M. Higashiwaki, T. Matsui: High-quality InN film grown on a low-temperature-grown GaN intermediate layer by plasma-assisted molecular-beam epitaxy, Jpn. J. Appl. Phys. Part 2 Lett. **41**(5B), 540 (2002)
31.266 H.W. Lu, J. Schaff, L.F. Eastman, J. Wu, W. Walukiewicz, K.M. Yu, J.W. Auger III, E.E. Haller, O. Ambacher: Conf. Digest of the 44th Electron.

Mat. Conf., Santa Barbara (2002) p. 2
31.267 M. Higashiwaki, T. Matsui: Epitaxial growth of high-quality InN films on sapphire substrates by plasma-assisted molecular-beam epitaxy, J. Cryst. Growth **252**(1–3), 128 (2003)
31.268 T.L. Tansley, C.P. Foley: Proc. 3rd Inter. Conf. Semi-Insulating III-V-Materials, Warm Springs, 1984, ed. by J. Blakemore (Shiva, London 1985) pp. 497–500
31.269 C. Stampfl, C.G. Van de Walle, D. Vogel, P. Kruger, J. Pollmann, D. Vogel: Native defects and impurities in InN: First-principles studies using the local-density approximation and self-interaction and relaxation-corrected pseudopotentials, Phys. Rev. B Condens. Matter **61**(12), 7846 (2000)
31.270 D.C. Look, H. Lu, W.J. Schaff, J. Jasinski, Z. Liliental-Weber: Donor and acceptor concentrations in degenerate InN, Appl. Phys. Lett. **80**(2), 258 (2002)
31.271 J.D. Albrecht, R.P. Wang, P.P. Ruden, M. Farahmand, K.F. Brennan: Electron transport characteristics of GaN for high temperature device modeling, J. Appl. Phys. **83**(9), 4777 (1998)
31.272 S.K. O'Leary, B.E. Foutz, M.S. Shur, U.V. Bhapkar, L.F. Eastman: Monte Carlo simulation of electron transport in wurtzite aluminum nitride, Solid State Commun. **105**(10), 621 (1998)
31.273 B. Gelmont, K. Kim, M. Shur: Monte Carlo simulation of electron transport in gallium nitride, J. Appl. Phys. **74**(3), 1818 (1993)
31.274 N.S. Mansour, K.W. Kim, M.A. Littlejohn: Theoretical study of electron transport in gallium nitride, J. Appl. Phys. **77**(6), 2834 (1995)
31.275 J. Kolník, İ.H. Oğuzman, K.F. Brennan, R. Wang, P.P. Ruden, Y. Wang: Electronic transport studies of bulk zincblende and wurtzite phases of GaN based on an ensemble Monte Carlo calculation including a full zone band structure, J. Appl. Phys. **78**(2), 1033 (1995)
31.276 M. Shur, B. Gelmont, M.A. Khan: Electron mobility in two-dimensional electron gas in AlGaN/GaN heterostructures and in bulk GaN, Proc. 37th Electronic Materials Conf. III–V Nitrides and Silicon Carbide, 1995 (1996) p. 777
31.277 S.K. O'Leary, B.E. Foutz, M.S. Shur, U.V. Bhapkar, L.F. Eastman: Electron transport in wurtzite indium nitride, J. Appl. Phys. **83**(2), 826 (1998)
31.278 E. Bellotti, B.K. Doshi, K.F. Brennan, J.D. Albrecht, P.P. Ruden: Ensemble Monte Carlo study of electron transport in wurtzite InN, J. Appl. Phys. **85**(2), 916 (1999)
31.279 B.E. Foutz, S.K. O'Leary, M.S. Shur, L.F. Eastman: Transient electron transport in wurtzite GaN, InN, and AlN, J. Appl. Phys. **85**(11), 7727 (1999)
31.280 C.G. Rodrigues, V.N. Freire, A.R. Vasconcellos, R. Luzzi: Velocity overshoot onset in nitride semiconductors, Appl. Phys. Lett. **76**(14), 1893 (2000)
31.281 M. Wraback, H. Shen, S. Rudin, E. Bellotti: Experimental and theoretical studies of transient electron velocity overshoot in GaN, Phys. Status Solidi B **234**(3), 810 (2002)
31.282 R. Collazo, R. Schlesser, Z. Sitar: Experimental observation of electron velocity overshoot in AlN, Appl. Phys. Lett. **81**(27), 5189 (2002)
31.283 B. Monemar, U. Lindefelt, W.M. Chen: Electronic structure of bound excitons in semiconductors, Phys. B Condens. Matter **146**(1), 2–256 (1987)
31.284 A. Rodina, M. Dietrich, A. Göldner, L. Eckey, A. Hoffmann, A. Efros, M. Rosen, B. Meyer: Free excitons in wurtzite GaN, Phys. Rev. B **64**(11), 115204 (2001)
31.285 R. Dingle, M. Ilegems: Donor-acceptor pair recombination in GaN, Solid State Commun. **9**(3), 175 (1971)
31.286 E.L. Ivchenko: *Excitons* (North-Holland, Amsterdam 1982)
31.287 J.J. Song, W. Shan: *Group III Nitride Semiconductor Compounds: Physics and Applications* (Oxford Univ. Press, Clarendon, New York 1998)
31.288 B. Monemar, J.P. Bergman, I.A. Buyanova: Optical characterization of GaN and related material. In: *GaN and Related Material*, ed. by S.J. Pearton (Golden Breach, Amsterdam 1997) p. 85
31.289 R. Dingle, D. Sell, S. Stokowski, M. Ilegems: Absorption, reflectance, and luminescence of GaN epitaxial layers, Phys. Rev. B **4**(4), 1211 (1971)
31.290 B. Monemar: Fundamental energy gap of GaN from photoluminescence excitation spectra, Phys. Rev. B **10**(2), 676 (1974)
31.291 B. Gil, O. Briot, R.-L. Aulombard: Valence-band physics and the optical properties of GaN epilayers grown onto sapphire with wurtzite symmetry, Phys. Rev. B **52**(24), R17028 (1995)
31.292 W. Shan, B. Little, A. Fischer, J. Song, B. Goldenberg, W. Perry, M. Bremser, R. Davis: Binding energy for the intrinsic excitons in wurtzite GaN, Phys. Rev. B **54**(23), 16369 (1996)
31.293 M. Smith, G.D. Chen, J.Z. Li, J.Y. Lin, H.X. Jiang, A. Salvador, W.K. Kim, O. Aktas, A. Botchkarev, H. Morkoç: Excitonic recombination in GaN grown by molecular beam epitaxy, Appl. Phys. Lett. **67**(23), 3387 (1995)
31.294 R. Kucharski, M. Rudziński, M. Zajac, R. Doradziński, J. Garczyński, L. Sierzputowski, R. Kudrawiec, J. Serafińczuk, W. Strupiński, R. Dwiliński: Nonpolar GaN substrates grown by ammonothermal method, Appl. Phys. Lett. **95**(13), 131119 (2009)
31.295 R. Dwilinski, R. Doradzinski, J. Garczynski, L.P. Sierzputowski, M. Zajac, M. Rudzinski: Homoepitaxy on bulk ammonothermal GaN, J. Cryst. Growth **311**(10), 3058 (2009)
31.296 R. Kudrawiec, M. Rudziński, J. Serafinczuk, M. Zajac, J. Misiewicz: Photoreflectance study of exciton energies and linewidths for homoepitaxial and heteroepitaxial GaN layers, J. Appl. Phys. **105**(9), 093541 (2009)
31.297 K.P. Korona, A. Wysmołek, K. Pakuła, R. Stępniewski, J.M. Baranowski, I. Grzegory, B. Łucznik, M. Wróblewski, S. Porowski: Exciton region reflectance of homoepitaxial GaN layers, Appl. Phys. Lett. **69**(6), 788 (1996)
31.298 K. Kornitzer, T. Ebner, M. Grehl, K. Thonke, R. Sauer, C. Kirchner, V. Schwegler, M. Kamp, M. Leszczynski, I. Grzegory, S. Porowski: High-resolution photoluminescence and reflectance spec-

tra of homoepitaxial GaN layers, Phys. Status Solidi B **216**(1), 5 (1999)
31.299 B.J. Skromme, K.C. Palle, C.D. Poweleit, H. Yamane, M. Aoki, F.J. DiSalvo: Optical spectroscopy of bulk GaN crystals grown from a Na–Ga melt, Appl. Phys. Lett. **81**(20), 3765 (2002)
31.300 R. Kucharski, M. Zajac, R. Doradzinski, J. Garczynski, L. Sierzputowski, R. Kudrawiec, J. Serafinczuk, J. Misiewicz, R. Dwilinski: Structural and optical properties of semipolar GaN substrates obtained by ammonothermal method, Appl. Phys. Express **3**(10), 101001 (2010)
31.301 P. Paskov, T. Paskova, P. Holtz, B. Monemar: Polarized photoluminescence study of free and bound excitons in free-standing GaN, Phys. Rev. B **70**(3), 035210 (2004)
31.302 O. Aoudé, P. Disseix, J. Leymarie, A. Vasson, M. Leroux, E. Aujol, B. Beaumont, A. Trassoudaine, Y. André: Continuous-wave and ultrafast coherent reflectivity studies of excitons in bulk GaN, Phys. Rev. B **77**(4), 045206 (2008)
31.303 B. Monemar, P.P. Paskov, J.P. Bergman, A.A. Toropov, T.V. Shubina, T. Malinauskas, A. Usui: Recombination of free and bound excitons in GaN, Phys. Status Solidi B **245**(9), 1723 (2008)
31.304 M. Tchounkeu, O. Briot, B. Gil, J.P. Alexis, R.-L. Aulombard: Optical properties of GaN epilayers on sapphire, J. Appl. Phys. **80**(9), 5352 (1996)
31.305 T. Gühne, Z. Bougrioua, S. Laügt, M. Nemoz, P. Vennéguès, B. Vinter, M. Leroux: Band-edge photoluminescence and reflectivity of nonpolar $(11\bar{2}0)$ and semipolar $(11\bar{2}2)$ GaN formed by epitaxial lateral overgrowth on sapphire, Phys. Rev. B **77**(7), 075308 (2008)
31.306 P. Misra, U. Behn, O. Brandt, H.T. Grahn, B. Imer, S. Nakamura, S.P. DenBaars, J.S. Speck: Polarization anisotropy in GaN films for different nonpolar orientations studied by polarized photoreflectance spectroscopy, Appl. Phys. Lett. **88**(16), 161920 (2006)
31.307 A.V. Rodina, M. Dietrich, A. Göldner, L. Eckey, A.L. Efros, M. Rosen, A. Hoffmann, B.K. Meyer: Exciton energy structure in Wurtzite GaN, Phys. Status Solidi B **216**(1), 21 (1999)
31.308 W. Shan, T.J. Schmidt, X.H. Yang, S.J. Hwang, J.J. Song, B. Goldenberg: Temperature dependence of interband transitions in GaN grown by metalorganic chemical vapor deposition, Appl. Phys. Lett. **66**(8), 985 (1995)
31.309 F. Hamdani, A. Botchkarev, W. Kim, H. Morkoç, M. Yeadon, J.M. Gibson, S.-C.Y. Tsen, D.J. Smith, D.C. Reynolds, D.C. Look, K. Evans, C.W. Litton, W.C. Mitchel, P. Hemenger: Optical properties of GaN grown on ZnO by reactive molecular beam epitaxy, Appl. Phys. Lett. **70**(4), 467 (1997)
31.310 W. Shan, A.J. Fischer, J.J. Song, G.E. Bulman, H.S. Kong, M.T. Leonard, W.G. Perry, M.D. Bremser, R.F. Davis: Optical studies of GaN and GaN/AlGaN heterostructures on SiC substrates, Appl. Phys. Lett. **69**(6), 740 (1996)
31.311 K. Pakuła, A. Wysmołek, K.P. Korona, J.M. Baranowski, R. Stępniewski, I. Grzegory, M. Boćkowski, J. Jun, S. Krukowski, M. Wróblewski, S. Porowski: Luminescence and reflectivity in the exciton region of homoepitaxial GaN layers grown on GaN substrates, Solid State Commun. **97**(11), 919 (1996)
31.312 Y.S. Huang, F.H. Pollak, S.S. Park, K.Y. Lee, H. Morkoç: Contactless electroreflectance, in the range of 20–300 K, of freestanding wurtzite GaN prepared by hydride-vapor-phase epitaxy, J. Appl. Phys. **94**(2), 899 (2003)
31.313 L. Eckey, L. Podlowski, A. Goldner, A. Hoffmann, I. Broser, B.K. Meyer, D. Volm, T. Streibl, K. Hiramatsu, T. Detchprohm: Excitonic structure of GaN epitaxial films grown by hydride-vapor-phase epitaxy, Inst. Phys. Conf. Ser. **142**, 943 (1996)
31.314 P. Paskov, T. Paskova, P. Holtz, B. Monemar: Spin-exchange splitting of excitons in GaN, Phys. Rev. B **64**(11), 115201 (2001)
31.315 A. Shikanai, T. Azuhata, T. Sota, S. Chichibu, A. Kuramata, K. Horino, S. Nakamura: Biaxial strain dependence of exciton resonance energies in wurtzite GaN, J. Appl. Phys. **81**(1), 417 (1997)
31.316 B. Monemar, J.P. Bergman, I.A. Buyanova, H. Amano, I. Akasaki, T. Detchprohm, K. Hiramatsu, N. Sawaki: The excitonic bandgap of GaN: Dependence on substrate, Solid-State Electron. **41**(2), 239 (1997)
31.317 J. Hopfield: Theory of the contribution of excitons to the complex dielectric constant of crystals, Phys. Rev. **112**(5), 1555 (1958)
31.318 T. Steiner, M. Thewalt, E. Koteles, J. Salerno: Effect of neutral donor scattering on the time-dependent exciton-polariton photoluminescence line shape in GaAs, Phys. Rev. B **34**(2), 1006 (1986)
31.319 B. Gil, S. Clur, O. Briot: The exciton-polariton effect on the photoluminescence of GaN on sapphire, Solid State Commun. **104**(5), 267 (1997)
31.320 R. Stępniewski, K. Korona, A. Wysmołek, J. Baranowski, K. Pakuła, M. Potemski, G. Martinez, I. Grzegory, S. Porowski: Polariton effects in reflectance and emission spectra of homoepitaxial GaN, Phys. Rev. B **56**(23), 15151 (1997)
31.321 K. Torii, T. Deguchi, T. Sota, K. Suzuki, S. Chichibu, S. Nakamura: Reflectance and emission spectra of excitonic polaritons in GaN, Phys. Rev. B **60**(7), 4723 (1999)
31.322 S. Rudin, T. Reinecke, B. Segall: Temperature-dependent exciton linewidths in semiconductors, Phys. Rev. B **42**(17), 11218 (1990)
31.323 A.J. Fischer, W. Shan, J.J. Song, Y.C. Chang, R. Horning, B. Goldenberg: Temperature-dependent absorption measurements of excitons in GaN epilayers, Appl. Phys. Lett. **71**(14), 1981 (1997)
31.324 X.B. Zhang, T. Taliercio, S. Kolliakos, P. Lefebvre: Influence of electron-phonon interaction on the optical properties of III nitride semiconductors, J. Phys. Condens. Matter **13**(32), 7053 (2001)
31.325 I.A. Buyanova, J.P. Bergman, B. Monemar, H. Amano, I. Akasaki: Intrinsic optical properties of GaN epilayers grown on SiC substrates: Effect

of the built-in strain, Appl. Phys. Lett. **69**(9), 1255 (1996)
31.326 S. Permogorov: *Excitons* (North-Holland, Amsterdam 1982)
31.327 A. Viswanath, J. Lee, D. Kim, C. Lee, J. Leem: Exciton-phonon interactions, exciton binding energy, and their importance in the realization of room-temperature semiconductor lasers based on GaN, Phys. Rev. B **58**(24), 16333 (1998)
31.328 M. Smith, G.D. Chen, J.Y. Lin, H.X. Jiang, M.A. Khan, C.J. Sun, Q. Chen, J.W. Yang: Free excitonic transitions in GaN, grown by metal-organic chemical-vapor deposition, J. Appl. Phys. **79**(9), 7001 (1996)
31.329 M. Smith, G.D. Chen, J.Y. Lin, H.X. Jiang, A. Salvador, B.N. Sverdlov, A. Botchkarev, H. Morkoç: Dynamics of a band-edge transition in GaN grown by molecular beam epitaxy, Appl. Phys. Lett. **66**(25), 3474 (1995)
31.330 D.C. Reynolds, D.C. Look, W. Kim, O. Aktas, A. Botchkarev, A. Salvador, H. Morkoç, D.N. Talwar: Ground and excited state exciton spectra from GaN grown by molecular-beam epitaxy, J. Appl. Phys. **80**(1), 594 (1996)
31.331 R. Pässler: Dispersion-related assessments of temperature dependences for the fundamental band gap of hexagonal GaN, J. Appl. Phys. **90**(8), 3956 (2001)
31.332 R. Stępniewski, M. Potemski, A. Wysmołek, K. Pakuła, J. Baranowski, J. Łusakowski, I. Grzegory, S. Porowski, G. Martinez, P. Wyder: Symmetry of excitons in GaN, Phys. Rev. B **60**(7), 4438 (1999)
31.333 S. Hess, F. Walraet, R. Taylor, J. Ryan, B. Beaumont, P. Gibart: Dynamics of resonantly excited excitons in GaN, Phys. Rev. B **58**(24), R15973 (1998)
31.334 Y. Zhong, K.S. Wong, W. Zhang, D.C. Look: Radiative recombination and ultralong exciton photoluminescence lifetime in GaN freestanding film via two-photon excitation, Appl. Phys. Lett. **89**(2), 022108 (2006)
31.335 G. Neu, M. Teisseire, P. Lemasson, H. Lahreche, N. Grandjean, F. Semond, B. Beaumont, I. Grzegory, S. Porowski, R. Triboulet: Selective photoluminescence spectroscopy of shallow levels in wide band gap semiconductors, Phys. B Condens. Matter **302/303**, 39 (2001)
31.336 J.M. Baranowski, Z. Liliental-Weber, K. Korona, K. Pakuła, R. Stępniewski, A. Wysmołek, I. Grzegory, G. Nowak, S. Porowski, B. Monemar, P. Bergman: Structural and optical properties of homoepitaxial GaN layers, MRS Proc. **449**, 393 (1996)
31.337 J. Freitas, W. Moore, B. Shanabrook, G. Braga, S. Lee, S. Park, J. Han: Donor-related recombination processes in hydride-vapor-phase epitaxial GaN, Phys. Rev. B **66**(23), 233311 (2002)
31.338 A.K. Viswanath: Growth and optical properties of GaN. In: *Semiconductors and Semimetals*, Vol. 73, ed. by R. Willardson, H.S. Nalwa (Academic, San Diego 2001) p. 63
31.339 P. Dean, J. Haynes, W. Flood: New radiative recombination processes involving neutral donors and acceptors in silicon and germanium, Phys. Rev. **161**(3), 711 (1967)
31.340 M.A. Reshchikov, D. Huang, F. Yun, L. He, H. Morkoç, D.C. Reynolds, S.S. Park, K.Y. Lee: Photoluminescence of GaN grown by molecular-beam epitaxy on a freestanding GaN template, Appl. Phys. Lett. **79**(23), 3779 (2001)
31.341 W.J. Moore, J.A. Freitas, G.C.B. Braga, R.J. Molnar, S.K. Lee, K.Y. Lee, I.J. Song: Identification of Si and O donors in hydride-vapor-phase epitaxial GaN, Appl. Phys. Lett. **79**(16), 2570 (2001)
31.342 E.I. Rashba, G.E. Gurgenishvili: Edge absorption theory in semiconductors, Sov. Phys. Solid State **4**(4), 759 (1962)
31.343 B. Monemar, P. Paskov, T. Paskova, J. Bergman, G. Pozina, W. Chen, P. Hai, I. Buyanova, H. Amano, I. Akasaki: Optical characterization of III-nitrides, Mater. Sci. Eng. B **93**(1–3), 112 (2002)
31.344 G. Pozina, N.V. Edwards, J.P. Bergman, T. Paskova, B. Monemar, M.D. Bremser, R.F. Davis: Time-resolved spectroscopy of strained GaN/AlN/6H-SiC heterostructures grown by metalorganic chemical vapor deposition, Appl. Phys. Lett. **78**(8), 1062 (2001)
31.345 S. Pau, J. Kuhl, M. Khan, C. Sun: Application of femtosecond-excitation correlation to the study of emission dynamics in hexagonal GaN, Phys. Rev. B **58**(19), 12916 (1998)
31.346 T. Shubina, M. Glazov, A. Toropov, N. Gippius, A. Vasson, J. Leymarie, A. Kavokin, A. Usui, J. Bergman, G. Pozina, B. Monemar: Resonant light delay in GaN with ballistic and diffusive propagation, Phys. Rev. Lett. **100**(8), 087402 (2008)
31.347 K. Kornitzer, T. Ebner, K. Thonke, R. Sauer, C. Kirchner, V. Schwegler, M. Kamp, M. Leszczynski, I. Grzegory, S. Porowski: Photoluminescence and reflectance spectroscopy of excitonic transitions in high-quality homoepitaxial GaN films, Phys. Rev. B **60**(3), 1471 (1999)
31.348 G. Pozina, J.P. Bergman, T. Paskova, B. Monemar: Bound exciton dynamics in GaN grown by hydride vapor-phase epitaxy, Appl. Phys. Lett. **75**(26), 4124 (1999)
31.349 K.P. Korona, R. Doradziński, M. Palczewska, M. Pietras, M. Kamińska, J. Kuhl: Properties of zinc acceptor and exciton bound to zinc in ammonothermal GaN, Phys. Status Solidi B **235**(1), 40 (2003)
31.350 M. Leroux, B. Beaumont, N. Grandjean, P. Lorenzini, S. Haffouz, P. Vennéguès, J. Massies, P. Gibart: Luminescence and reflectivity studies of undoped, n- and p-doped GaN on (0001) sapphire, Mater. Sci. Eng. B **50**(1–3), 97 (1997)
31.351 M. Leroux, N. Grandjean, B. Beaumont, G. Nataf, F. Semond, J. Massies, P. Gibart: Temperature quenching of photoluminescence intensities in undoped and doped GaN, J. Appl. Phys. **86**(7), 3721 (1999)

31.352 G. Pozina, B. Monemar, P.P. Paskov, C. Hemmingsson, L. Hultman, H. Amano, I. Akasaki, T. Paskova, S. Figge, D. Hommel, A. Usui: Metastability of the UV luminescence in Mg-doped GaN layers grown by MOVPE on quasi-bulk GaN templates, Phys. B Condens. Matter **401/402**, 302 (2007)

31.353 O. Gelhausen, M. Phillips, E. Goldys, T. Paskova, B. Monemar, M. Strassburg, A. Hoffmann: Dissociation of H-related defect complexes in Mg-doped GaN, Phys. Rev. B **69**(12), 125210 (2004)

31.354 T.A.G. Eberlein, R. Jones, S. Öberg, P.R. Briddon: Shallow acceptors in GaN, Appl. Phys. Lett. **91**(13), 132105 (2007)

31.355 D.C. Reynolds, D.C. Look, B. Jogai, V.M. Phanse, R.P. Vaudo: Identification of an ionized-donor-bound-exciton transition in GaN, Solid State Commun. **103**(9), 533 (1997)

31.356 B. Šantic, C. Merz, U. Kaufmann, R. Niebuhr, H. Obloh, K. Bachem: Ionized donor bound excitons in GaN, Appl. Phys. Lett. **71**(13), 1837 (1997)

31.357 A.K. Viswanath, J.I. Lee, S. Yu, D. Kim, Y. Choi, C. Hong: Photoluminescence studies of excitonic transitions in GaN epitaxial layers, J. Appl. Phys. **84**(7), 3848 (1998)

31.358 R.A. Mair, J. Li, S.K. Duan, J.Y. Lin, H.X. Jiang: Time-resolved photoluminescence studies of an ionized donor-bound exciton in GaN, Appl. Phys. Lett. **74**(4), 513 (1999)

31.359 C. Merz, M. Kunzer, U. Kaufmann, I. Akasaki, H. Amano: Free and bound excitons in thin wurtzite GaN layers on sapphire, Semicond. Sci. Technol. **11**(5), 712 (1996)

31.360 B. Monemar, P.P. Paskov, J.P. Bergman, T. Paskova, S. Figge, J. Dennemarck, D. Hommel: The dominant shallow 0.225 eV acceptor in GaN, Phys. Status Solidi B **243**(7), 1604 (2006)

31.361 B. Monemar, P.P. Paskov, G. Pozina, C. Hemmingsson, J.P. Bergman, S. Khromov, V.N. Izyumskaya, V. Avrutin, X. Li, H. Morkoç, H. Amano, M. Iwaya, I. Akasaki: Properties of the main Mg-related acceptors in GaN from optical and structural studies, J. Appl. Phys. **115**(5), 053507 (2014)

31.362 S. Khromov, C.G. Hemmingsson, H. Amano, B. Monemar, L. Hultman, G. Pozina: Luminescence related to high density of Mg-induced stacking faults in homoepitaxially grown GaN, Phys. Rev. B **84**(7), 075324 (2011)

31.363 D. Thomas, J. Hopfield: Optical properties of bound exciton complexes in cadmium sulfide, Phys. Rev. **128**(5), 2135 (1962)

31.364 H. Saito, S. Shionoya, E. Hanamura: New luminescence line associated with the inelastic collision of excitonic molecules in CdS and CdSe, Solid State Commun. **12**(3), 227 (1973)

31.365 O. Lagerstedt: Luminescence in epitaxial GaN:Cd, J. Appl. Phys. **45**(5), 2266 (1974)

31.366 S. Fischer, C. Wetzel, E.E. Haller, B.K. Meyer: On p-type doping in GaN-acceptor binding energies, Appl. Phys. Lett. **67**(9), 1298 (1995)

31.367 J.L. Lyons, A. Janotti, C.G. van de Walle: Shallow versus deep nature of Mg acceptors in nitride semiconductors, Phys. Rev. Lett. **108**(15), 156403 (2012)

31.368 Y. Kamiura, M. Kaneshiro, J. Tamura, T. Ishiyama, Y. Yamashita, T. Mitani, T. Mukai: Enhancement of blue emission from Mg-doped GaN using remote plasma containing atomic hydrogen, Jpn. J. Appl. Phys. **44**(29), L926 (2005)

31.369 M. Reshchikov, G.-C. Yi, B. Wessels: Behavior of 2.8 and 3.2 eV photoluminescence bands in Mg-doped GaN at different temperatures and excitation densities, Phys. Rev. B **59**(20), 13176 (1999)

31.370 B. Monemar, P. Paskov, G. Pozina, C. Hemmingsson, J. Bergman, T. Kawashima, H. Amano, I. Akasaki, T. Paskova, S. Figge, D. Hommel, A. Usui: Evidence for two Mg related acceptors in GaN, Phys. Rev. Lett. **102**(23), 235501 (2009)

31.371 D. Thomas, J. Hopfield, W. Augustyniak: Kinetics of radiative recombination at randomly distributed donors and acceptors, Phys. Rev. **140**(1A), A202 (1965)

31.372 B. Monemar, S. Khromov, G. Pozina, P. Paskov, P. Bergman, C. Hemmingsson, L. Hultman, H. Amano, V. Avrutin, X. Li, H. Morkoç: Luminescence of acceptors in Mg-doped GaN, Jpn. J. Appl. Phys. **52**(8S), 08JJ03 (2013)

31.373 M.A. Reshchikov, H. Morkoç: Luminescence properties of defects in GaN, J. Appl. Phys. **97**(6), 061301 (2005)

31.374 E. Glaser, T. Kennedy, K. Doverspike, L. Rowland, D. Gaskill, J. Freitas, M. Asif Khan, D. Olson, J. Kuznia, D. Wickenden: Optically detected magnetic resonance of GaN films grown by organometallic chemical-vapor deposition, Phys. Rev. B **51**(19), 13326 (1995)

31.375 D. Demchenko, I. Diallo, M. Reshchikov: Yellow luminescence of gallium nitride generated by carbon defect complexes, Phys. Rev. Lett. **110**(8), 087404 (2013)

31.376 T. Ogino, M. Aoki: Mechanism of yellow luminescence in GaN, Jpn. J. Appl. Phys. **19**(12), 2395 (1980)

31.377 K. Saarinen, T. Laine, S. Kuisma, J. Nissilä, P. Hautojärvi, L. Dobrzynski, J. Baranowski, K. Pakula, R. Stepniewski, M. Wojdak, A. Wysmolek, T. Suski, M. Leszczynski, I. Grzegory, S. Porowski: Observation of native Ga vacancies in GaN by positron annihilation, Phys. Rev. Lett. **79**(16), 3030 (1997)

31.378 J. Neugebauer, C.G. Van de Walle: Gallium vacancies and the yellow luminescence in GaN, Appl. Phys. Lett. **69**(4), 503 (1996)

31.379 T. Mattila, R.M. Nieminen: Point-defect complexes and broadband luminescence in GaN and AlN, Phys. Rev. B **55**(15), 9571 (1997)

31.380 S.O. Kucheyev, M. Toth, M.R. Phillips, J.S. Williams, C. Jagadish, G. Li: Chemical origin of the yellow luminescence in GaN, J. Appl. Phys. **91**(9), 5867 (2002)

31.381 I. Shalish, L. Kronik, G. Segal, Y. Rosenwaks, Y. Shapira, U. Tisch, J. Salzman: Yellow luminescence and related deep levels in unintentionally doped GaN films, Phys. Rev. B **59**(15), 9748 (1999)

31.382 E. Calleja, F.J. Sánchez, D. Basak, M.A. Sánchez-García, E. Muñoz, I. Izpura, F. Calle, J.M.G. Tijero, J.L. Sánchez-Rojas, B. Beaumont, P. Lorenzini,

P. Gibart: Yellow luminescence and related deep states in undoped GaN, Phys. Rev. B **55**(7), 4689 (1997)
31.383 J.L. Lyons, A. Janotti, C.G. Van de Walle: Carbon impurities and the yellow luminescence in GaN, Appl. Phys. Lett. **97**(15), 152108 (2010)
31.384 D. Hofmann, D. Kovalev, G. Steude, B. Meyer, A. Hoffmann, L. Eckey, R. Heitz, T. Detchprom, H. Amano, I. Akasaki: Properties of the yellow luminescence in undoped GaN epitaxial layers, Phys. Rev. B **52**(23), 16702 (1995)
31.385 R. Korotkov, M. Reshchikov, B. Wessels: Transient photoluminescence of defects in undoped GaN prepared by metal organic vapor phase epitaxy, Phys. B Condens. Matter **273/274**, 80 (1999)
31.386 M.A. Reshchikov, H. Morkoç, S.S. Park, K.Y. Lee: Transient photoluminescence of defect transitions in freestanding GaN, Appl. Phys. Lett. **78**(19), 2882 (2001)
31.387 S. Pačesová, L. Jastrabík: Radiative recombination at deep impurity centres in AlN:O, Czechoslov. J. Phys. **29**(8), 913 (1979)
31.388 R.A. Youngman, J.H. Harris: Luminescence studies of oxygen-related defects in aluminum nitride, J. Am. Ceram. Soc. **73**(11), 3238 (1990)
31.389 J.H. Harris, R.A. Youngman, R.G. Teller: On the nature of the oxygen-related defect in aluminum nitride, J. Mater. Res. **5**(08), 1763 (1990)
31.390 R. Collazo, J. Xie, B.E. Gaddy, Z. Bryan, R. Kirste, M. Hoffmann, R. Dalmau, B. Moody, Y. Kumagai, T. Nagashima, Y. Kubota, T. Kinoshita, A. Koukitu, D.L. Irving, Z. Sitar: On the origin of the 265 nm absorption band in AlN bulk crystals, Appl. Phys. Lett. **100**(19), 191914 (2012)
31.391 T. Koyama, M. Sugawara, T. Hoshi, A. Uedono, J.F. Kaeding, R. Sharma, S. Nakamura, S.F. Chichibu: Relation between Al vacancies and deep emission bands in AlN epitaxial films grown by NH_3-source molecular beam epitaxy, Appl. Phys. Lett. **90**(24), 241914 (2007)
31.392 K.B. Nam, M.L. Nakarmi, J.Y. Lin, H.X. Jiang: Deep impurity transitions involving cation vacancies and complexes in AlGaN alloys, Appl. Phys. Lett. **86**(22), 222108 (2005)
31.393 B.N. Pantha, N. Nepal, T.M. Al Tahtamouni, M.L. Nakarmi, J. Li, J.Y. Lin, H.X. Jiang: Correlation between biaxial stress and free exciton transition in AlN epilayers, Appl. Phys. Lett. **91**(12), 121117 (2007)
31.394 W. Rieger, T. Metzger, H. Angerer, R. Dimitrov, O. Ambacher, M. Stutzmann: Influence of substrate-induced biaxial compressive stress on the optical properties of thin GaN films, Appl. Phys. Lett. **68**(7), 970 (1996)
31.395 E. Silveira, J. Freitas, O. Glembocki, G. Slack, L. Schowalter: Excitonic structure of bulk AlN from optical reflectivity and cathodoluminescence measurements, Phys. Rev. B **71**(4), 041201 (2005)
31.396 B. Neuschl, K. Thonke, M. Feneberg, S. Mita, J. Xie, R. Dalmau, R. Collazo, Z. Sitar: Optical identification of silicon as a shallow donor in MOVPE grown homoepitaxial AlN, Phys. Status Solidi B **249**(3), 511 (2012)
31.397 M. Feneberg, B. Neuschl, K. Thonke, R. Collazo, A. Rice, Z. Sitar, R. Dalmau, J. Xie, S. Mita, R. Goldhahn: Sharp bound and free exciton lines from homoepitaxial AlN, Phys. Status Solidi A **208**(7), 1520 (2011)
31.398 H. Ikeda, T. Okamura, K. Matsukawa, T. Sota, M. Sugawara, T. Hoshi, P. Cantu, R. Sharma, J.F. Kaeding, S. Keller, U.K. Mishra, K. Kosaka, K. Asai, S. Sumiya, T. Shibata, M. Tanaka, J.S. Speck, S.P. DenBaars, S. Nakamura, T. Koyama, T. Onuma, S.F. Chichibu: Impact of strain on free-exciton resonance energies in wurtzite AlN, J. Appl. Phys. **102**(12), 123707 (2007)
31.399 L. Chen, B.J. Skromme, R.F. Dalmau, R. Schlesser, Z. Sitar, C. Chen, W. Sun, J. Yang, M.A. Khan, M.L. Nakarmi, J.Y. Lin, H.-X. Jiang: Band-edge exciton states in AlN single crystals and epitaxial layers, Appl. Phys. Lett. **85**(19), 4334 (2004)
31.400 Y. Yamada, K. Choi, S. Shin, H. Murotani, T. Taguchi, N. Okada, H. Amano: Photoluminescence from highly excited AlN epitaxial layers, Appl. Phys. Lett. **92**(13), 131912 (2008)
31.401 E. Silveira, J.A. Freitas, M. Kneissl, D.W. Treat, N.M. Johnson, G.A. Slack, L.J. Schowalter: Near-bandedge cathodoluminescence of an AlN homoepitaxial film, Appl. Phys. Lett. **84**(18), 3501 (2004)
31.402 M. Feneberg, R.A.R. Leute, B. Neuschl, K. Thonke, M. Bickermann: High-excitation and high-resolution photoluminescence spectra of bulk AlN, Phys. Rev. B **82**(7), 075208 (2010)
31.403 T. Onuma, T. Shibata, K. Kosaka, K. Asai, S. Sumiya, M. Tanaka, T. Sota, A. Uedono, S.F. Chichibu: Free and bound exciton fine structures in AlN epilayers grown by low-pressure metalorganic vapor phase epitaxy, J. Appl. Phys. **105**(2), 023529 (2009)
31.404 H. Murotani, T. Kuronaka, Y. Yamada, T. Taguchi, N. Okada, H. Amano: Temperature dependence of excitonic transitions in *a*-plane AlN epitaxial layers, J. Appl. Phys. **105**(8), 083533 (2009)
31.405 R. Phillips, D. Lovering, G. Denton, G. Smith: Biexciton creation and recombination in a GaAs quantum well, Phys. Rev. B **45**(8), 4308 (1992)
31.406 R.A.R. Leute, M. Feneberg, R. Sauer, K. Thonke, S.B. Thapa, F. Scholz, Y. Taniyasu, M. Kasu: Photoluminescence of highly excited AlN: Biexcitons and exciton-exciton scattering, Appl. Phys. Lett. **95**(3), 031903 (2009)
31.407 M. Funato, K. Matsuda, R.G. Banal, R. Ishii, Y. Kawakami: Homoepitaxy and photoluminescence properties of (0001) AlN, Appl. Phys. Express **5**(8), 082001 (2012)
31.408 G. Rossbach, M. Feneberg, M. Röppischer, C. Werner, N. Esser, C. Cobet, T. Meisch, K. Thonke, A. Dadgar, J. Bläsing, A. Krost, R. Goldhahn: Influence of exciton-phonon coupling and strain on the anisotropic optical response of wurtzite AlN around the band edge, Phys. Rev. B **83**(19), 195202 (2011)

31.409 R. Dalmau, B. Moody, R. Schlesser, S. Mita, J. Xie, M. Feneberg, B. Neuschl, K. Thonke, R. Collazo, A. Rice, J. Tweedie, Z. Sitar: Growth and characterization of AlN and AlGaN epitaxial films on AlN single crystal substrates, J. Electrochem. Soc. **158**(5), H530 (2011)

31.410 G.M. Prinz, A. Ladenburger, M. Schirra, M. Feneberg, K. Thonke, R. Sauer, Y. Taniyasu, M. Kasu, T. Makimoto: Cathodoluminescence, photoluminescence, and reflectance of an aluminum nitride layer grown on silicon carbide substrate, J. Appl. Phys. **101**(2), 023511 (2007)

31.411 G.M. Prinz, M. Feneberg, M. Schirra, R. Sauer, K. Thonke, S.B. Thapa, F. Scholz: Silicon-doping induced strain of AlN layers: A comparative luminescence and Raman study, Phys. Status Solidi RRL–Rapid Res. Lett. **2**(5), 215 (2008)

31.412 K.B. Nam, M.L. Nakarmi, J. Li, J.Y. Lin, H.X. Jiang: Photoluminescence studies of Si-doped AlN epilayers, Appl. Phys. Lett. **83**(14), 2787 (2003)

31.413 M. Yoshikawa, M. Kunzer, J. Wagner, H. Obloh, P. Schlotter, R. Schmidt, N. Herres, U. Kaufmann: Band-gap renormalization and band filling in Si-doped GaN films studied by photoluminescence spectroscopy, J. Appl. Phys. **86**(8), 4400 (1999)

31.414 N. Nepal, M.L. Nakarmi, K.B. Nam, J.Y. Lin, H.X. Jiang: Acceptor-bound exciton transition in Mg-doped AlN epilayer, Appl. Phys. Lett. **85**(12), 2271 (2004)

31.415 K.B. Nam, J. Li, M.L. Nakarmi, J.Y. Lin, H.X. Jiang: Deep ultraviolet picosecond time-resolved photoluminescence studies of AlN epilayers, Appl. Phys. Lett. **82**(11), 1694 (2003)

31.416 S.F. Chichibu, T. Onuma, K. Hazu, A. Uedono: Major impacts of point defects and impurities on the carrier recombination dynamics in AlN, Appl. Phys. Lett. **97**(20), 201904 (2010)

31.417 Q. Guo, A. Yoshida: Temperature dependence of band gap change in InN and AlN, Jpn. J. Appl. Phys. **33**(5A), 2453 (1994)

31.418 E. Kuokstis, J. Zhang, Q. Fareed, J.W. Yang, G. Simin, M.A. Khan, R. Gaska, M. Shur, C. Rojo, L. Schowalter: Near-band-edge photoluminescence of wurtzite-type AlN, Appl. Phys. Lett. **81**(15), 2755 (2002)

31.419 X. Tang, F. Hossain, K. Wongchotigul, M.G. Spencer: Near band-edge transition in aluminum nitride thin films grown by metal organic chemical vapor deposition, Appl. Phys. Lett. **72**(12), 1501 (1998)

31.420 F. Chen, A. Cartwright, H. Lu, W.J. Schaff: Ultrafast carrier dynamics in InN epilayers, J. Cryst. Growth **269**(1), 10 (2004)

31.421 K. Osamura, S. Naka, Y. Murakami: Preparation and optical properties of $Ga_{1-x}In_xN$ thin films, J. Appl. Phys. **46**(8), 3432 (1975)

31.422 A. Wakahara, T. Tsuchiya, A. Yoshida: Epitaxial growth of indium nitride, J. Cryst. Growth **99**(1–4), 385 (1990)

31.423 K. Kubota, Y. Kobayashi, K. Fujimoto: Preparation and properties of III–V nitride thin films, J. Appl. Phys. **66**(7), 2984 (1989)

31.424 T.L. Tansley, C.P. Foley: Optical band gap of indium nitride, J. Appl. Phys. **59**(9), 3241 (1986)

31.425 S. Yamaguchi, M. Kariya, S. Nitta, T. Takeuchi, C. Wetzel, H. Amano, I. Akasaki: Anomalous features in the optical properties of $Al_{1-x}In_xN$ on GaN grown by metal organic vapor phase epitaxy, Appl. Phys. Lett. **76**(7), 876 (2000)

31.426 T. Inushima, V. Mamutin, V. Vekshin, S. Ivanov, T. Sakon, M. Motokawa, S. Ohoya: Physical properties of InN with the band gap energy of 1.1 eV, J. Cryst. Growth **227/228**, 481 (2001)

31.427 J. Wu, W. Walukiewicz, W. Shan, K. Yu, J. Ager, E. Haller, H. Lu, W. Schaff: Effects of the narrow band gap on the properties of InN, Phys. Rev. B **66**(20), 201403(R) (2002)

31.428 J. Wu, W. Walukiewicz, K.M. Yu, J.W. Ager, E.E. Haller, H. Lu, W.J. Schaff: Small band gap bowing in $In_{1-x}Ga_xN$ alloys, Appl. Phys. Lett. **80**(25), 4741 (2002)

31.429 V.Y. Davydov, A.A. Klochikhin, R.P. Seisyan, V.V. Emtsev, S.V. Ivanov, F. Bechstedt, J. Furthmüller, H. Harima, A.V. Mudryi, J. Aderhold, O. Semchinova, J. Graul: Absorption and emission of hexagonal InN. Evidence of narrow fundamental band gap, Phys. Status Solidi B **229**(3), r1 (2002)

31.430 J. Wu, W. Walukiewicz, K.M. Yu, J.W. Ager, E.E. Haller, H. Lu, W.J. Schaff, Y. Saito, Y. Nanishi: Unusual properties of the fundamental band gap of InN, Appl. Phys. Lett. **80**(21), 3967 (2002)

31.431 T. Inushima, M. Higashiwaki, T. Matsui: Optical properties of Si-doped InN grown on sapphire (0001), Phys. Rev. B **68**(23), 235204 (2003)

31.432 J. Wu, W. Walukiewicz, S.X. Li, R. Armitage, J.C. Ho, E.R. Weber, E.E. Haller, H. Lu, W.J. Schaff, A. Barcz, R. Jakiela: Effects of electron concentration on the optical absorption edge of InN, Appl. Phys. Lett. **84**(15), 2805 (2004)

31.433 V.Y. Davydov, A.A. Klochikhin, V.V. Emtsev, S.V. Ivanov, V.V. Vekshin, F. Bechstedt, J. Furthmüller, H. Harima, A.V. Mudryi, A. Hashimoto, A. Yamamoto, J. Aderhold, J. Graul, E.E. Haller: Band gap of inN and in-rich $In_xGa_{1-x}N$ alloys ($0.36 < x < 1$), Phys. Status Solidi B **230**(2), R4 (2002)

31.434 J. Thakur, Y. Danylyuk, D. Haddad, V. Naik, R. Naik, G. Auner: Influence of defects on the absorption edge of InN thin films: The band gap value, Phys. Rev. B **76**(3), 035309 (2007)

31.435 I. Mahboob, T. Veal, L. Piper, C. McConville, H. Lu, W. Schaff, J. Furthmüller, F. Bechstedt: Origin of electron accumulation at wurtzite InN surfaces, Phys. Rev. B **69**(20), 201307(R) (2004)

31.436 I. Mahboob, T. Veal, C. McConville, H. Lu, W. Schaff: Intrinsic electron accumulation at clean InN surfaces, Phys. Rev. Lett. **92**(3), 036804 (2004)

31.437 H. Lu, W.J. Schaff, L.F. Eastman, C.E. Stutz: Surface charge accumulation of InN films grown by molecular-beam epitaxy, Appl. Phys. Lett. **82**(11), 1736 (2003)

31.438 H. Lu, W.J. Schaff, L.F. Eastman, J. Wu, W. Walukiewicz, D.C. Look, R.J. Molnar: Growth of thick InN by molecular beam epitaxy, MRS Proc.

743, 317 (2002)
31.439 J. Wu, W. Walukiewicz: Band gaps of InN and group III nitride alloys, Superlattices Microstruct. **34**(1/2), 63 (2003)
31.440 K.S.A. Butcher, M. Wintrebert-Fouquet, S.K. Shrestha, H. Timmers, K.E. Prince, T.L. Tansley: A study of indium nitride films grown under conditions resulting in apparent band-gaps from 0.7 eV to 2.3 eV, MRS Proc. **743**, 707 (2002)
31.441 T. Shubina, S. Ivanov, V. Jmerik, D. Solnyshkov, V. Vekshin, P. Kop'ev, A. Vasson, J. Leymarie, A. Kavokin, H. Amano, K. Shimono, A. Kasic, B. Monemar: Mie resonances, infrared emission, and the band gap of InN, Phys. Rev. Lett. **92**(11), 117407 (2004)
31.442 O. Briot, B. Maleyre, S. Clur-Ruffenach, B. Gil, C. Pinquier, F. Demangeot, J. Frandon: The value of the direct bandgap of InN: A re-examination, Phys. Status Solidi C **1**(6), 1425 (2004)
31.443 B. Monemar, P.P. Paskov, A. Kasic: Optical properties of InN – The bandgap question, Superlattices Microstruct. **38**(1), 38 (2005)
31.444 P. Specht, R. Armitage, J. Ho, E. Gunawan, Q. Yang, X. Xu, C. Kisielowski, E. Weber: The influence of structural properties on conductivity and luminescence of MBE grown InN, J. Cryst. Growth **269**(1), 111 (2004)
31.445 S. Gwo, C.-L. Wu, C.-H. Shen, W.-H. Chang, T.M. Hsu, J.-S. Wang, J.-T. Hsu: Heteroepitaxial growth of wurtzite InN films on Si(111) exhibiting strong near-infrared photoluminescence at room temperature, Appl. Phys. Lett. **84**(19), 3765 (2004)
31.446 A. Wakahara, T. Tsuchiya, A. Yoshida: Epitaxial layers of indium nitride by microwave-excited metalorganic vapor phase epitaxy, Vacuum **41**(4–6), 1071 (1990)
31.447 F. Chen, A. Cartwright, H. Lu, W.J. Schaff: Temperature-dependent optical properties of wurtzite InN, Phys. E Low-Dimens. Syst. Nanostructures **20**(3/4), 308 (2004)
31.448 L. Görgens, O. Ambacher, M. Stutzmann, C. Miskys, F. Scholz, J. Off: Characterization of InGaN thin films using high-resolution x-ray diffraction, Appl. Phys. Lett. **76**(5), 577 (2000)
31.449 T. Peng, J. Piprek, G. Qiu, J.O. Olowolafe, K.M. Unruh, C.P. Swann, E.F. Schubert: Band gap bowing and refractive index spectra of polycrystalline $Al_xIn_{1-x}N$ films deposited by sputtering, Appl. Phys. Lett. **71**(17), 2439 (1997)
31.450 V. Darakchieva, M. Beckers, M.-Y. Xie, L. Hultman, B. Monemar, J.-F. Carlin, E. Feltin, M. Gonschorek, N. Grandjean: Effects of strain and composition on the lattice parameters and applicability of Vegard's rule in Al-rich $Al_{1-x}In_xN$ films grown on sapphire, J. Appl. Phys. **103**(10), 103513 (2008)
31.451 F. Yun, M.A. Reshchikov, L. He, T. King, H. Morkoç, S.W. Novak, L. Wei: Energy band bowing parameter in $Al_xGa_{1-x}N$ alloys, J. Appl. Phys. **92**(8), 4837 (2002)
31.452 N. Nepal, J. Li, M.L. Nakarmi, J.Y. Lin, H.X. Jiang: Temperature and compositional dependence of the energy band gap of AlGaN alloys, Appl. Phys. Lett. **87**(24), 242104 (2005)
31.453 H. Murotani, Y. Yamada, T. Taguchi, A. Ishibashi, Y. Kawaguchi, T. Yokogawa: Temperature dependence of localized exciton transitions in AlGaN ternary alloy epitaxial layers, J. Appl. Phys. **104**(5), 053514 (2008)
31.454 M.A. Khan, J.M. Van Hove, J.N. Kuznia, D.T. Olson: High electron mobility $GaN/Al_xGa_{1-x}N$ heterostructures grown by low-pressure metalorganic chemical vapor deposition, Appl. Phys. Lett. **58**(21), 2408 (1991)
31.455 Y.A. Xi, K.X. Chen, F. Mont, J.K. Kim, E.F. Schubert, W. Liu, X. Li, J.A. Smart: Comparative study of n-type AlGaN grown on sapphire by using a superlattice layer and a low-temperature AlN interlayer, J. Cryst. Growth **299**(1), 59 (2007)
31.456 W.H. Sun, J.W. Yang, J.P. Zhang, M.E. Gaevski, C.Q. Chen, J.W. Li, Z. Gong, M. Su, M. Asif Khan: n-$Al_{0.75}Ga_{0.25}N$ epilayers for 250 nm emission ultraviolet light emitting diodes, Phys. Status Solidi C **2**(7), 2083 (2005)
31.457 M.L. Nakarmi, K.H. Kim, K. Zhu, J.Y. Lin, H.X. Jiang: Transport properties of highly conductive n-type Al-rich $Al_xGa_{1-x}N$ ($x \geq 0.7$), Appl. Phys. Lett. **85**(17), 3769 (2004)
31.458 M. Stutzmann, O. Ambacher, A. Cros, M. Brandt, H. Angerer, R. Dimitrov, N. Reinacher, T. Metzger, R. Höpler, D. Brunner, F. Freudenberg, R. Handschuh, C. Deger: Properties and applications of MBE grown AlGaN, Mater. Sci. Eng. B **50**(1–3), 212 (1997)
31.459 T. Tanaka, A. Watanabe, H. Amano, Y. Kobayashi, I. Akasaki, S. Yamazaki, M. Koike: p-Type conduction in Mg-doped GaN and $Al_{0.08}Ga_{0.92}N$ grown by metalorganic vapor phase epitaxy, Appl. Phys. Lett. **65**(5), 593 (1994)
31.460 S.-R. Jeon, Z. Ren, G. Cui, J. Su, M. Gherasimova, J. Han, H.-K. Cho, L. Zhou: Investigation of Mg doping in high-Al content p-type $Al_xGa_{1-x}N$ ($0.3 < x < 0.5$), Appl. Phys. Lett. **86**(8), 082107 (2005)
31.461 W. Liu, A.A. Balandin: Thermal conduction in $Al_xGa_{1-x}N$ alloys and thin films, J. Appl. Phys. **97**(7), 073710 (2005)
31.462 F.K. Yam, Z. Hassan: InGaN: An overview of the growth kinetics, physical properties and emission mechanisms, Superlattices Microstruct. **43**(1), 1 (2008)
31.463 C.-R. Lee, S.-J. Son, I.-H. Lee, J.-Y. Leem, S.K. Noh: Characteristics of $In_xGa_{1-x}N/GaN$ grown by LPMOVPE with the variation of growth temperature, J. Cryst. Growth **182**(1), 6 (1997)
31.464 W. Van der Stricht, I. Moerman, P. Demeester, J. Crawley, E.J. Thrush: Study of GaN and InGaN films grown by metalorganic chemical vapour deposition, J. Cryst. Growth **170**(1–4), 344 (1997)
31.465 H.-C. Lin, C.-K. Shu, J. Ou, Y.-C. Pan, W.-K. Chen, W.-H. Chen, M.-C. Lee: Growth temperature effects on $In_xGa_{1-x}N$ films studied by x-ray and photoluminescence, J. Cryst. Growth **189/190**, 57 (1998)

31.466 K. Kumakura, T. Makimoto, N. Kobayashi: Activation energy and electrical activity of Mg in Mg-doped $In_xGa_{1-x}N$ ($x < 0.2$), Jpn. J. Appl. Phys. **39**(4B), L337 (2000)

31.467 M. Suzuki, J. Nishio, M. Onomura, C. Hongo: Doping characteristics and electrical properties of Mg-doped AlGaN grown by atmospheric-pressure MOCVD, J. Cryst. Growth **189/190**, 511 (1998)

31.468 B.N. Pantha, A. Sedhain, J. Li, J.Y. Lin, H.X. Jiang: Electrical and optical properties of p-type InGaN, Appl. Phys. Lett. **95**(26), 261904 (2009)

31.469 J. Xie, X. Ni, M. Wu, J.H. Leach, U. Özgür, H. Morkoç: High electron mobility in nearly lattice-matched AlInN/AlN/GaN heterostructure field effect transistors, Appl. Phys. Lett. **91**(13), 132116 (2007)

31.470 J.-F. Carlin, M. Ilegems: High-quality AlInN for high index contrast Bragg mirrors lattice matched to GaN, Appl. Phys. Lett. **83**(4), 668 (2003)

31.471 O. Ambacher, J. Majewski, C. Miskys, A. Link, M. Hermann, M. Eickhoff, M. Stutzmann, F. Bernardini, V. Fiorentini, V. Tilak, B. Schaff, L.F. Eastman: Pyroelectric properties of Al(In)GaN/GaN hetero- and quantum well structures, J. Phys. Condens. Matter **14**(13), 3399 (2002)

31.472 R.E. Jones, R. Broesler, K.M. Yu, J.W. Ager, E.E. Haller, W. Walukiewicz, X. Chen, W.J. Schaff: Band gap bowing parameter of $In_{1-x}Al_xN$, J. Appl. Phys. **104**(12), 123501 (2008)

31.473 W. Terashima, S.-B. Che, Y. Ishitani, A. Yoshikawa: Growth and characterization of AlInN ternary alloys in whole composition range and fabrication of InN/AlInN multiple quantum wells by RF molecular beam epitaxy, Jpn. J. Appl. Phys. **45**(21), L539 (2006)

31.474 K. Lorenz, N. Franco, E. Alves, S. Pereira, I.M. Watson, R.W. Martin, K.P. O'Donnell: Relaxation of compressively strained AlInN on GaN, J. Cryst. Growth **310**(18), 4058 (2008)

31.475 W.R. Bryden, T.J. Kistenmacher: Electrical transport properties of InN, GaInN and AlInN. In: *Properties of Group III Nitrides*, ed. by J.H. Edgar (Ed. Inspec, London 1995) p. 910

31.476 M.J. Lukitsch, Y.V. Danylyuk, V.M. Naik, C. Huang, G.W. Auner, L. Rimai, R. Naik: Optical and electrical properties of $Al_{1-x}In_xN$ films grown by plasma source molecular-beam epitaxy, Appl. Phys. Lett. **79**(5), 632 (2001)

31.477 S.N. Mohammad, A.A. Salvador, H. Morkoç: Emerging gallium nitride based devices, Proc. IEEE **83**(10), 1306 (1995)

31.478 S.M. Bedair, F.G. McIntosh, J.C. Roberts, E.L. Piner, K.S. Boutros, N.A. El-Masry: Growth and characterization of In-based nitride compounds, J. Cryst. Growth **178**(1/2), 32 (1997)

31.479 J.-Y. Chang, Y.-K. Kuo: Influence of polarization-matched AlGaInN barriers in blue InGaN light-emitting diodes, Opt. Lett. **37**(9), 1574 (2012)

31.480 M.F. Schubert, J. Xu, J.K. Kim, E.F. Schubert, M.H. Kim, S. Yoon, S.M. Lee, C. Sone, T. Sakong, Y. Park: Polarization-matched GaInN/AlGaInN multi-quantum-well light-emitting diodes with reduced efficiency droop, Appl. Phys. Lett. **93**(4), 041102 (2008)

31.481 M.E. Aumer, S.F. LeBoeuf, B.F. Moody, S.M. Bedair: Strain-induced piezoelectric field effects on light emission energy and intensity from AlInGaN/InGaN quantum wells, Appl. Phys. Lett. **79**(23), 3803 (2001)

31.482 M.-Y. Ryu, C.Q. Chen, E. Kuokstis, J.W. Yang, G. Simin, M.A. Khan: Luminescence mechanisms in quaternary $Al_xIn_yGa_{1-x-y}N$ materials, Appl. Phys. Lett. **80**(20), 3730 (2002)

31.483 W. Walukiewicz: Intrinsic limitations to the doping of wide-gap semiconductors, Proc. Yamada Conf. LIV 9th Int. Conf. Shallow-Level Centers in Semiconductors (2001) p. 123

31.484 N. Koide, H. Kato, M. Sassa, S. Yamasaki, K. Manabe, M. Hashimoto, H. Amano, K. Hiramatsu, I. Akasaki: Doping of GaN with Si and properties of blue m/i/n/n+ GaN LED with Si-doped n+-layer by MOVPE, J. Cryst. Growth **115**(1–4), 639 (1991)

31.485 S. Fritze, A. Dadgar, H. Witte, M. Bugler, A. Rohrbeck, J. Blasing, A. Hoffmann, A. Krost: High Si and Ge n-type doping of GaN doping: Limits and impact on stress, Appl. Phys. Lett. **100**(12), 122104 (2012)

31.486 W.J. Moore, J.A. Freitas, S.K. Lee, S.S. Park, J.Y. Han: Magneto-optical studies of free-standing hydride-vapor-phase epitaxial GaN, Phys. Rev. B **65**(8), 081201 (2002)

31.487 P.J. Dean, J.D. Cuthbert, D.G. Thomas, R.T. Lynch: Two-electron transitions in the luminescence of excitons bound to neutral donors in gallium phosphide, Phys. Rev. Lett. **18**(4), 122 (1967)

31.488 M. Wieneke, H. Witte, K. Lange, M. Feneberg, A. Dadgar, J. Blasing, R. Goldhahn, A. Krost: Ge as a surfactant in metal-organic vapor phase epitaxy growth of *a*-plane GaN exceeding carrier concentrations of 1020 cm^{-3}, Appl. Phys. Lett. **103**(1), 012103 (2013)

31.489 J.K. Sheu, G.C. Chi: The doping process and dopant characteristics of GaN, J. Phys. Condens. Matter **14**(22), 657 (2002)

31.490 H. Amano, M. Kito, K. Hiramatsu, I. Akasaki: P-type conduction in Mg-doped GaN treated with low-energy electron beam irradiation (LEEBI), Jpn. J. Appl. Phys. **28**(12A), L2112 (1989)

31.491 S. Nakamura, M. Senoh, T. Mukai: Highly p-typed Mg-doped GaN films grown with GaN buffer layers, Jpn. J. Appl. Phys. **30**(10A), L1708 (1991)

31.492 S. Nakamura, T. Mukai, M. Senoh, N. Iwasa: Thermal annealing effects on p-type Mg-doped GaN films, Jpn. J. Appl. Phys. **31**(2B), L139 (1992)

31.493 H. Amano, M. Kitoh, K. Hiramatsu, I. Akasaki: Growth and luminescence properties of Mg-doped GaN prepared by MOVPE, J. Electrochem. Soc. **137**(5), 1639 (1990)

31.494 H. Obloh, K.H. Bachem, U. Kaufmann, M. Kunzer, M. Maier, A. Ramakrishnan, P. Schlotter: Self-compensation in Mg doped p-type GaN grown by MOCVD, Proc. Metalorganic Vapour Phase Epitaxy 1998. Ninth Int. Conf (1998) p. 270

31.495 W. Kim, A.E. Botchkarev, A. Salvador, G. Popovici, H. Tang, H. Morkoc: On the incorporation of Mg and the role of oxygen, silicon, and hydrogen in GaN prepared by reactive molecular beam epitaxy, J. Appl. Phys. **82**(1), 219 (1997)

31.496 W. Kim, A. Salvador, A.E. Botchkarev, O. Aktas, S.N. Mohammad, H. Morcoç: Mg-doped p-type GaN grown by reactive molecular beam epitaxy, Appl. Phys. Lett. **69**(4), 559 (1996)

31.497 W. Gotz, N.M. Johnson, J. Walker, D.P. Bour, R.A. Street: Activation of acceptors in Mg-doped GaN grown by metalorganic chemical vapor deposition, Appl. Phys. Lett. **68**(5), 667 (1996)

31.498 I.D. Goepfert, E.F. Schubert, A. Osinsky, P.E. Norris: Demonstration of efficient p-type doping in $Al_xGa_{1-x}N$/GaN superlattice structures, Electron. Lett. **35**(13), 1109 (1999)

31.499 P. Kozodoy, H. Xing, S.P. DenBaars, U.K. Mishra, A. Saxler, R. Perrin, S. Elhamri, W.C. Mitchel: Heavy doping effects in Mg-doped GaN, J. Appl. Phys. **87**(4), 1832 (2000)

31.500 M. Suzuki, J. Nishio, M. Onomura, C. Hongo: Doping characteristics and electrical properties of Mg-doped AlGaN grown by atmospheric-pressure MOCVD, Proc. Second Int. Conf. Nitride Semiconductors (1998) p. 511

31.501 L. Sugiura, M. Suzuki, J. Nishio, K. Itaya, Y. Kokubun, M. Ishikawa: Characteristics of Mg-doped GaN and AlGaN grown by H_2-ambient and N_2-ambient metallorganic chemical vapor deposition, Jpn. J. Appl. Phys. Part 1 **37**(7), 3878 (1998)

31.502 J. Li, T.N. Oder, M.L. Nakarmi, J.Y. Lin, H.X. Jiang: Optical and electrical properties of Mg-doped p-type $Al_xGa_{1-x}N$, Appl. Phys. Lett. **80**(7), 1210 (2002), (1998) p. 511

31.503 H. Yu, E. Ulker, E. Ozbay: MOCVD growth and electrical studies of p-type AlGaN with Al fraction 0.35, J. Cryst. Growth **289**(2), 419 (2006)

31.504 S.N. Mohammed, H. Morkoc: Progress and prospects of group-III nitride semiconductors, Prog. Quantum Electron. **20**(5/6), 361 (1996)

31.505 L.K. Li, M.J. Jurkovic, W.I. Wang, J. Van Hove, P.P. Chow: Surface polarity dependence of Mg doping in GaN grown by molecular-beam epitaxy, Appl. Phys. Lett. **76**(13), 1740 (2000)

31.506 A.J. Ptak, T.H. Myers, L.T. Romano, C.G.J.E. Van de Walle: Northrup: Magnesium incorporation in GaN grown by molecular-beam epitaxy, Appl. Phys. Lett. **78**(3), 285 (2001)

31.507 S. Guha, N.A. Bojarczuk, F. Cardone: Mg in GaN: Incorporation of a volatile species at high temperatures during molecular beam epitaxy, Appl. Phys. Lett. **71**(12), 1685 (1997)

31.508 T.S. Cheng, S.V. Novikov, C.T. Foxon, J.W. Orton: Mechanisms of magnesium incorporation into GaN layers grown by molecular beam epitaxy, Solid State Commun. **109**(7), 439 (1999)

31.509 I.P. Smorchkova, E. Haus, B. Heying, P. Kozodoy, P. Fini, J.P. Ibbetson, S. Keller, S.P. DenBaars, J.S. Speck, U.K. Mishra: Mg doping of GaN layers grown by plasma-assisted molecular-beam epitaxy, Appl. Phys. Lett. **76**(6), 718 (2000)

31.510 R.Y. Korotkov, J.M. Gregie, B.W. Wessels: Codoping of wide gap epitaxial III-nitride semiconductors, Opto-Electron. Rev. **10**(4), 243 (2002)

31.511 H. Reiss, C.S. Fuller, F.J. Morin: Chemical interactions among defects in germanium and silicon, Bell Syst. Tech. J. **35**, 535 (1956)

31.512 H. Reiss, C.S. Fuller, A.J. Pietruszkiewicz: Solubility of lithium in doped and undoped silicon, evidence for compound formation, J. Chem. Phys. **25**(4), 650 (1956)

31.513 U. Kaufmann, P. Schlotter, H. Obloh, K. Kohler, M. Maier: Hole conductivity and compensation in epitaxial GaN: Mg layers, Phys. Rev. B Condens. Matter **62**(16), 10867 (2000)

31.514 Y. Qi, C. Musante, K.M. Lau, L. Smith, R. Odedra, R. Kanjolia: OMVPE growth of p-type GaN using solution Cp_2Mg, Tenth Bienn. Workshop Organomet, Vap. Phase Epitaxy **30**(11), 1382 (2001)

31.515 K.S. Kim, M.S. Han, G.M. Yang, C.J. Youn, H.J. Lee, H.K. Cho, J.Y. Lee: Codoping characteristics of Zn with Mg in GaN, Appl. Phys. Lett. **77**(8), 1123 (2000)

31.516 K.H. Ploog, O. Brandt: Doping of group III nitrides, J. Vac. Sci. Technol. A: Vac. Surf. Films **16**(3), 1609 (1998)

31.517 R.Y. Korotkov, J.M. Gregie, B.W. Wessels: Electrical properties of p-type GaN:Mg codoped with oxygen, Appl. Phys. Lett. **78**(2), 222 (2001)

31.518 G. Kipshidze, V. Kuryatkov, B. Borisov, Y. Kudryavtsev, R. Asomoza, S. Nikishin, H. Temkin: Mg and O codoping in p-type GaN and $Al_xGa_{1-x}N$ ($0 \leq x \leq 0.08$), Appl. Phys. Lett. **80**(16), 2910 (2002)

31.519 E.F. Schubert, W. Grieshaber, I.D. Goepfert: Enhancement of deep acceptor activation in semiconductors by superlattice doping, Appl. Phys. Lett. **69**(24), 3737 (1996)

31.520 P. Kozodoy, Y.P. Smorchkova, M. Hansen, H. Xing, S.P. DenBaars, U.K. Mishra: Polarization-enhanced Mg doping of AlGaN/GaN superlattices, Appl. Phys. Lett. **75**(16), 2444 (1999)

31.521 A. Saxler, W.C. Mitchel, P. Kung, M. Razeghi: Aluminum gallium nitride short-period superlattices doped with magnesium, Appl. Phys. Lett. **74**(14), 2023 (1999)

31.522 L. Hsu, W. Walukiewicz: Theoretical transport studies of p-type GaN/AlGaN modulation-doped heterostructures, Appl. Phys. Lett. **74**(17), 2405 (1999)

31.523 P. Kozodoy, M. Hansen, S.P. DenBaars, U.K. Mishra: Enhanced Mg doping efficiency in $Al_{0.2}Ga_{0.8}N$/GaN superlattices, Appl. Phys. Lett. **4**(24), 3681 (1999)

31.524 A.D. Bykhovski, B.L. Gelmont, M.S. Shur: Elastic strain relaxation and piezoeffect in GaN-AlN, GaN-AlGaN and GaN-InGaN superlattices, J. Appl. Phys. **81**(9), 6332 (1997)

31.525 E.L. Waldron, J.W. Graff, E.F. Schubert: Improved mobilities and resistivities in modulation-doped p-type AlGaN/GaN superlattices, Appl. Phys. Lett. **79**(17), 2737 (2001)

31.526 F. Bernardini, V. Fiorentini, D. Vanderbilt: Spontaneous polarization and piezoelectric constants of III–V nitrides, Phys. Rev. B **56**(16), R10024 (1997)

31.527 V. Adivarahan, J. Zhang, A. Chitnis, S. Wu, J. Sun, R. Pachipulusu, M. Shatalov, M.A. Khan: Submilliwatt power III-N light emitting diodes at 285 nm, Jpn. J. Appl. Phys. Part 2 **41**(4B), 435 (2002)

31.528 M. Stavola: To 40 years of defects in semiconductors: May the problem never be solved!, Phys. B Condens. Matter **273/274**, 1 (1999)

31.529 D.V. Lang: Space-charge spectroscopy in semiconductor. In: *Thermally Stimulated Relaxation in Solids*, Topics in Applied Physics, Vol. 37, ed. by P. Bräunlich (Springer, Berlin 1979) pp. 93–133

31.530 L.C. Kimerling: *Defects in Semiconductors*, Vol. 2 (North-Holland, Amsterdam 1981)

31.531 W. Götz, N.M. Johnson, H. Amano, I. Akasaki: Deep level defects in n-type GaN, Appl. Phys. Lett. **65**(4), 463 (1994)

31.532 P. Hacke, T. Detchprohm, K. Hiramatsu, N. Sawaki, K. Tadatomo, K. Miyake: Analysis of deep levels in n-type GaN by transient capacitance methods, J. Appl. Phys. **76**(1), 304 (1994)

31.533 W.I. Lee, T.C. Huang, J.D. Guo, M.S. Feng: Effects of column III alkyl sources on deep levels in GaN grown by organometallic vapor phase epitaxy, Appl. Phys. Lett. **67**(12), 1721 (1995)

31.534 L. Chernyak, A. Osinsky, H. Temkin, J.W. Yang, Q. Chen, M. Asif Khan: Electron beam induced current measurements of minority carrier diffusion length in gallium nitride, Appl. Phys. Lett. **69**(17), 2531 (1996)

31.535 L. Chernyak, A. Osinsky, V. Fuflyigin, E.F. Schubert: Electron beam-induced increase of electron diffusion length in p-type GaN and AlGaN/GaN superlattices, Appl. Phys. Lett. **77**(6), 875 (2000)

31.536 L. Chernyak, A. Osinsky, G. Nootz, A. Schulte, J. Jasinski, M. Benamara, Z. Liliental-Weber, D.C. Look, R.J. Molnar: Electron beam and optical depth profiling of quasibulk GaN, Appl. Phys. Lett. **77**(17), 2695 (2000)

31.537 R.W. Siegel: Positron annihilation spectroscopy, Ann. Rev. Mater. Sci. **10**, 393 (1980)

31.538 R.W. Siegel, M.J. Fluss, L.C. Smedskjaer: Application of positron annihilation in materials science, Proc. 5th Riso Int. Symp. Metallurgy and Materials Sci., Microstructural Characterization of Materials by Non-Microscopical Techniques (1984) p. 131

31.539 R. Krause-Rehberg, H.S. Leipner: Determination of absolute vacancy concentrations in semiconductors by means of positron annihilation, Appl. Phys. **A64**(5), 457 (1997)

31.540 R. Krause-Rehberg, F. Borner, F. Redmann: Positron beam studies of defects in semiconductors, Mat. Sci. Forum **363–365**, 404 (2001)

31.541 A. Laakso, K. Saarinen, P. Hautojarvi: Positron lifetime beam for defect studies in thin epitaxial semiconductor structures, Phys. B Condens. Matter **308–310**, 1157 (2001)

31.542 R.M. Gwilliam, A.P. Knights, E. Wendler, B.J. Sealy, C.P. Burrows, P.G. Coleman: Development of a novel tool for semiconductor process control, Mat. Sci. Eng. B **80**, 60 (2001)

31.543 A.H. Deng, Y.Y. Shan, S. Fung, C.D. Beling: Application of positron annihilation lifetime technique to the study of deep level transients in semiconductors, J. Appl. Phys. **91**(6), 3931 (2002)

31.544 J. Dekker, R. Aavikko, K. Saarinen: Characterization of superlattices using positron annihilation, Appl. Surf. Sci. **194**, 97 (2002)

31.545 T. Onuma, S.F. Chichibu, A. Uedono, T. Sota, P. Cantu, T.M. Katona, J.F. Keading, S. Keller, U.K. Mishra, S. Nakamura, S.P. DenBaars: Radiative and nonradiative processes in strain-free $Al_xGa_{1-x}N$ films studied by time-resolved photoluminescence and positron annihilation techniques, J. Appl. Phys. **95**(5), 2495 (2004)

31.546 G.P. Karwasz, A. Zecca, R.S. Brusa, D. Pliszka: Application of positron annihilation techniques for semiconductor studies, J. Alloys and Compounds **382**, 244 (2004)

31.547 S.F. Chichibu, A. Uedono, T. Onuma, T. Sota, B.A. Haskell, S.P. DenBaars, J.S. Speck, S. Nakamura: Limiting factors of room-temperature nonradiative photoluminescence lifetime in polar and nonpolar GaN studied by time-resolved photoluminescence and slow positron annihilation techniques, Appl. Phys. Lett. **86**(2), 21914 (2005)

31.548 P. Premchander, S. Abhaya, K. Sivaji, G. Amarendra, K. Baskar, Y.T. Lee: Production of vacancy defects in high-energy Sn-ion irradiated GaN-Positron beam Doppler broadening study, Phys. B Condens. Matter **376/377**, 507 (2006)

31.549 S.F. Chichibu, A. Uedono, T. Onuma, B.A. Haskell, A. Chakraborty, T. Koyama, P.T. Fini, S. Keller, S.P. Denbaars, J.S. Speck, U.K. Mishra, S. Nakamura, S. Yamaguchi, S. Kamiyama, H. Amano, I. Akasaki, J. Han, T. Sotas: Origin of localized excitons in In-containing three-dimensional bulk (Al, In, Ga)N alloy films probed by time-resolved photoluminescence and monoenergetic positron annihilation techniques, Philos. Mag. **87**(13), 2019 (2007)

31.550 P.G. Coleman: Defect profiles in semiconductor structures, Phys. Status Solidi C **4**(10), 3620 (2007)

31.551 A. Uedono, K. Ito, H. Nakamori, K. Mori, Y. Nakano, T. Kachi, S. Ishibashi, T. Ohdaira, R. Suzuki: Annealing properties of vacancy-type defects in ion-implanted GaN studied by monoenergetic positron beams, J. Appl. Phys. **102**(8), 084505 (2007)

31.552 X.L. Yang, W.X. Zhu, C.D. Wang, H. Fang, T.J. Yu, Z.J. Yang, G.Y. Zhang, X.B. Qin, R.S. Yu, B.Y. Wang: Positron annihilation in (Ga, Mn)N: A study of vacancy-type defects, Appl. Phys. Lett. **94**(15), 151907 (2009)

31.553 A. Uedono, S. Ishibashi, T. Ohdaira, R. Suzuki: Point defects in group-III nitride semiconductors studied by positron annihilation, J. Cryst. Growth **311**(10), 3075 (2009)

31.554 A. Uedono: Material characterization by means of positron annihilation, Actual Phys. **79**(4), 307 (2010)

31.555 A. Uedono, S. Ishibashi, S.F. Chichibu, K. Akimoto: Point defects in GaN and related group-III nitrides studied by means of positron annihilation. In: *Gallium Nitride Materials and Devices VI*,

Vol. 7939, ed. by J.-I. Chyi, Y. Nanishi, H. Morkoç, J. Piprek, E. Yoon (SPIE Press, San Francisco 2011)
31.556 A. Zubiaga, J.A. Garcia, F. Plazaola, J. Zuniga-Perez, V. Munoz-Sanjose: Determination of defect content and defect profile in semiconductor heterostructures, J. Phys. Conf. Ser. **265**, 012004 (2011)
31.557 F. Tuomisto, I. Makkonen: Defect identification in semiconductors with positron annihilation: Experiment and theory, Rev. Mod. Phys. **85**(4), 1583 (2013)
31.558 A. Uedono, S. Ishibashi, N. Oshima, R. Suzuki: Positron annihilation spectroscopy on nitride-based semiconductors, Recent Adv. Nitride Semicond. **52**(8), 08JJ02 (2013)
31.559 K. Saarinen, T. Laine, S. Kuisma, J. Nissilä, P. Hautojärvi, L. Dobrzynski, J.M. Baranowski, K. Pakula, R. Stepniewski, M. Wojdak, A. Wysmolek, T. Suski, M. Leszczynski, I. Grzegory, S. Porowski: Observation of native Ga vacancies in GaN by positron annihilation, Phys. Rev. Lett. **79**(16), 3030 (1997)
31.560 F. Tuomisto, K. Saarinen, B. Lucznik, I. Grzegory, H. Teisseyre, T. Suski, S. Porowski, P.R. Hageman, J. Likonen: Effect of growth polarity on vacancy defect and impurity incorporation in dislocation-free GaN, Appl. Phys. Lett. **86**(3), 31915 (2005)
31.561 F. Tuomisto, V. Ranki, D.C. Look, G.C. Farlow: Introduction and recovery of Ga and N sublattice defects in electron-irradiated GaN, Phys. Rev. B **76**(16), 165207 (2007)
31.562 F. Tuomisto: Vacancy Defects in III-nitrides: What does positron annihilation spectroscopy reveal?, Proc. Int. Workshop Positron Studies of Defects (PSD 08) (2011) p. 012003
31.563 S. Hautakangas, I. Makkonen, V. Ranki, M.J. Puska, K. Saarinen, X. Xu, D.C. Look: Direct evidence of impurity decoration of Ga vacancies in GaN from positron annihilation spectroscopy, Phys. Rev. B **73**(19), 193301 (2006)
31.564 J. Oila, J. Kivioja, V. Ranki, K. Saarinen, D.C. Look, R.J. Molnar, S.S. Park, S.K. Lee, J.Y. Han: Ga vacancies as dominant intrinsic acceptors in GaN grown by hydride vapor phase epitaxy, Appl. Phys. Lett. **82**(20), 3433 (2003)
31.565 K. Saarinen, S. Hautakangas, F. Tuomisto: Dominant intrinsic acceptors in GaN and ZnO, Proc. 21st Nordic Semiconductor Meet., 21NSM (2006) p. 105
31.566 F. Tuomisto, S. Hautakangas, I. Makkonen, V. Ranki, M.J. Puska, K. Saarinen, M. Bockowski, T. Suski, T. Paskova, B. Monemar, X. Xu, D.C. Look: Dissociation of VGa–ON complexes in HVPE GaN by high pressure and high temperature annealing, Phys. Status Solidi B **243**(7), 1436 (2006)
31.567 K. Saarinen, T. Suski, I. Grzegory, D.C. Look: Thermal stability of isolated and complexed Ga vacancies in GaN bulk crystals, Phys. Rev. B **64**(23), 233201 (2001)
31.568 T. Paskova, D. Hommel, P.P. Paskov, V. Darakchieva, B. Monemar, M. Bockowski, T. Suski, I. Grzegory, F. Tuomisto, K. Saarinen, N. Ashkenov, M. Schubert: Effect of high-temperature annealing on the residual strain and bending of freestanding GaN films grown by hydride vapor phase epitaxy, Appl. Phys. Lett. **88**(14), 141909 (2006)
31.569 F. Tuomisto, K. Saarinen, T. Paskova, B. Monemar, M. Bockowski, T. Suski: Thermal stability of in-grown vacancy defects in GaN grown by hydride vapor phase epitaxy, J. Appl. Phys. **99**(6), 06610511 (2006)
31.570 F. Tuomisto, T. Suski, H. Teisseyre, M. Krysko, M. Leszczynski, B. Lucznik, I. Grzegory, S. Porowski, D. Wasik, A. Witowski, W. Gebicki, P. Hageman, K. Saarinen: Polarity dependent properties of GaN layers grown by hydride vapor phase epitaxy on GaN bulk crystals, Phys. Status Solidi B **240**(2), 289 (2003)
31.571 N. Faleev, C. Honsberg, O. Jani, I. Ferguson: Crystalline perfection of GaN and AlN epitaxial layers and the main features of structural transformation of crystalline defects, J. Cryst. Growth **300**(1), 246–250 (2007)
31.572 F. Tuomisto, T. Paskova, R. Kröger, S. Figge, D. Hommel, B. Monemar, R. Kersting: Defect distribution in *a*-plane GaN on Al_2O_3, Appl. Phys. Lett. **90**(12), 121915 (2007)
31.573 J. Oila, V. Ranki, J. Kivioja, K. Saarinen, P. Hautojärvi, J. Likonen, J.M. Baranowski, K. Pakula, T. Suski, M. Leszczynski, I. Grzegory: Influence of dopants and substrate material on the formation of Ga vacancies in epitaxial GaN layers, Phys. Rev. B **63**(4), 045205 (2001)
31.574 S. Hautakangas, J. Oila, M. Alatalo, K. Saarinen, L. Liszkay, D. Seghier, H.P. Gislason: Vacancy defects as compensating centers in Mg-doped GaN, Phys. Rev. Lett. **90**(13), 137402 (2003)
31.575 S. Hautakangas, K. Saarinen, L. Liszkay, J.A. Freitas, R.L. Henry: Role of open volume defects in Mg-doped GaN films studied by positron annihilation spectroscopy, Phys. Rev. B **72**(16), 165303 (2005)
31.576 F. Tuomisto, T. Paskova, S. Figge, D. Hommel, B. Monemar: Vacancy defect distribution in heteroepitaxial-plane GaN grown by hydride vapor phase epitaxy, First Int. Symp. Growth Nitrides ISGN-1 (2007) p. 251
31.577 J. Oila, A. Kemppinen, A. Laakso, K. Saarinen, W. Egger, L. Liszkay, P. Sperr, H. Lu, W.J. Schaff: Influence of layer thickness on the formation of In vacancies in InN grown by molecular beam epitaxy, Appl. Phys. Lett. **84**(9), 1486 (2004)
31.578 F. Tuomisto, A. Pelli, K.M. Yu, W. Walukiewicz, W.J. Schaff: Compensating point defects in He^{+4}-irradiated InN, Phys. Rev. B **75**(19), 193201 (2007)
31.579 P. Asoka-Kumar, K.G. Lynn: Applications of positron annihilation spectroscopy, J Phys IV Fr. **05**(C1), C1 (1995)
31.580 A. Pelli, K. Saarinen, F. Tuomisto, S. Ruffenach, O. Briot: Influence of V/III molar ratio on the formation of in vacancies in InN grown by metal-organic vapor-phase epitaxy, Appl. Phys. Lett. **89**(1), 011911 (2006)
31.581 F. Reurings, C. Rauch, F. Tuomisto, R.E. Jones, K.M. Yu, W. Walukiewicz, W.J. Schaff: Defect redistribution in postirradiation rapid-thermal-

annealed InN, Phys. Rev. B **82**(15), 153202 (2010)
31.582 F. Tuomisto, J.-M. Mäki, T.Y. Chemekova, Y.N. Makarov, O.V. Avdeev, E.N. Mokhov, A.S. Segal, M.G. Ramm, S. Davis, G. Huminic, H. Helava, M. Bickermann, B.M. Epelbaum: Characterization of bulk AlN crystals with positron annihilation spectroscopy, Spec. Issue IWBNS-5 Int. Workshop Bulk Nitride Semicond. V (2008) p. 3998
31.583 J.-M. Mäki, F. Tuomisto, B. Bastek, F. Bertam, J. Christen, A. Dadgar, A. Krost: Effect of growth conditions on vacancy defects in MOVPE grown AlN thin layers, Phys. Status Solidi C **6**(11), 2575 (2009)
31.584 Z. Liliental-Weber, X. Ni, H. Morkoc: Structural perfection of laterally overgrown GaN layers grown in polar- and non-polar directions, J. Mater. Sci. Mater. Electron. **19**(8/9), 815 (2008)
31.585 S.K. Mathis, A.E. Romanov, L.F. Chen, G.E. Beltz, W. Pompe, J.S. Speck: Modeling of threading dislocation reduction in growing GaN layers, Phys. Status Solidi A **179**(1), 125 (2000)
31.586 T. Sugahara, H. Sato, M. Hao, Y. Naoi, S. Kurai, S. Tottori, K. Yamashita, K. Nishino, L.T. Romano, S. Sakai: Direct evidence that dislocations are non-radiative recombination centers in GaN, Jpn. J. Appl. Phys. Part 2 Lett. **37**(4A), 398 (1998)
31.587 A. Chakraborty, B.A. Haskell, S. Keller, J.S. Speck, S.P. Denbaars, S. Nakamura, U.K. Mishra: Demonstration of nonpolar *m*-plane InGaN/GaN light-emitting diodes on free-standing *m*-plane GaN substrates, Jpn. J. Appl. Phys. Part 2 Lett. **44**, 173 (2005)
31.588 A. Sakai, H. Sunakawa, A. Usui: Defect structure in selectively grown GaN films with low threading dislocation density, Appl. Phys. Lett. **71**(16), 2259 (1997)
31.589 Z. Liliental-Weber, M. Benamara, W. Swider, J. Washburn, J. Park, P.A. Grudowski, C.J. Eiting, R.D. Dupuis: TEM study of defects in laterally overgrown GaN layers, MRS Internet J. Nitride Semicond. Res. **4S1**, 6d (1999)
31.590 Z. Liliental-Weber, D. Cherns: Microstructure of laterally overgrown GaN layers, J. Appl. Phys. **89**(12), 7833 (2001)
31.591 T.S. Zheleva, S.A. Smith, D.B. Thomson, K.J. Linthicum, P. Rajagopal, R.P. Davis: Pendeo-epitaxy: A new approach for lateral growth of gallium nitride films, J. Electron. Mater. **28**(4), 5 (1999)
31.592 P. Fini, H. Marchand, J.P. Ibbetson, B. Moran, L. Zhao, S.P. Denbaars, J.S. Speck, U.K. Mishra: Maskless lateral epitaxial overgrowth of GaN on sapphire, Proc. 1999 MRS Spring Meet. – Symp. Wide-Bandgap Semiconductors for High-Power, High Frequency and High-Temperature Applications (1999) p. 315
31.593 R.F. Davis, T. Gehrke, K.J. Linthicum, T.S. Zheleva, E.A. Preble, P. Rajagopal, C.A. Zorman, M. Mehregany: Pendeo-epitaxial growth of thin films of gallium nitride and related materials and their characterization, Proc. Twelfth Am. Conf. Crystal Growth and Epitaxy (2001) p. 134
31.594 S. Ruvimov, Z. Liliental-Weber, T. Suski, J.W. Ager III, J. Washburn, J. Krueger, C. Kisielowski, E.R. Weber, H. Amano, I. Akasaki: Effect of Si doping on the dislocation structure of GaN grown on the A-face of sapphire, Appl. Phys. Lett. **69**(7), 990 (1996)
31.595 M.E. Hawkridge, Z. Liliental-Weber, K.M. Yu, L.A. Reichertz, W.J. Schaff, J.W. Ager, W. Walukiewicz: Stacking faults and phase changes in Mg-doped InGaN grown on Si, Phys. Status Solidi C **6**(S2), 421 (2009)
31.596 D. Zubia, S.H. Zaidi, S.R.J. Brueck, S.D. Hersee: Nanoheteroepitaxial growth of GaN on Si by organometallic vapor phase epitaxy, Appl. Phys. Lett. **76**(7), 858 (2000)
31.597 I. Vurgaftman, J.R. Meyer, N. Tansu, L.J. Mawst: (In)GaAsN-based type-II W quantum-well lasers for emission at $\lambda = 1.55$ μm, Appl. Phys. Lett. **83**(14), 2742 (2003)
31.598 Y. Ikenaga, T. Miyamoto, S. Makino, T. Kageyama, M. Arai, F. Koyama, K. Iga: 1.4 μm GaInNAs/GaAs quantum well laser grown by chemical beam epitaxy, Jpn. J. Appl. Phys. **41**(2A), 664 (2002)
31.599 T. Kitatani, K. Nakahara, M. Kondow, K. Uomi, T. Tanaka: A 1.3 μm GaInNAs/GaAs single-quantum-well laser diode with a high characteristic temperature over 200 K, Jpn. J. Appl. Phys. **39**(2A), L86 (2000)
31.600 M. Leroux, N. Grandjean, J. Massies, B. Gil, P. Lefebvre, P. Bigenwald: Barrier-width dependence of group-III nitrides quantum-well transition energies, Phys. Rev. B Condens. Matter **60**(3), 1496 (1999)
31.601 G. Cheng, A. Kolmakov, Y. Zhang, M. Moskovits, R. Munden, M.A. Reed, G. Wang, D. Moses, J. Zhang: Current rectification in a single GaN nanowire with a well-defined p-n junction, Appl. Phys. Lett. **83**(8), 1578 (2003)
31.602 W. Guo, M. Zhang, P. Bhattacharya, J. Heo: Auger recombination in III-nitride nanowires and its effect on nanowire light-emitting diode characteristics, Nano Lett. **11**(4), 1434 (2011)
31.603 T. Kuykendall, P. Ulrich, S. Aloni, P. Yang: Complete composition tunability of InGaN nanowires using a combinatorial approach, Nat. Mater. **6**(12), 951 (2007)
31.604 F. Qian, S. Gradecak, Y. Li, C.-Y. Wen, C.M. Lieber: Core/multishell nanowire heterostructures as multicolor, high-efficiency light-emitting diodes, Nano Lett. **5**(11), 2287 (2005)
31.605 T. Kuykendall, P.J. Pauzauskie, Y. Zhang, J. Goldberger, D. Sirbuly, J. Denlinger, P. Yang: Crystallographic alignment of high-density gallium nitride nanowire arrays, Nat. Mater. **3**(8), 524 (2004)
31.606 H.P.T. Nguyen, S. Zhang, K. Cui, X. Han, S. Fathololoumi, M. Couillard, G.A. Botton, Z. Mi: p-Type modulation doped InGaN/GaN dot-in-a-wire white-light-emitting diodes monolithically grown on Si(111), Nano Lett. **11**(5), 1919 (2011)
31.607 B. Daudin: Polar and nonpolar GaN quantum dots, J. Phys. Condens. Matter **20**(47), 473201

(2008)
31.608 N. Grandjean, J. Massies: GaN/$Al_xGa_{1-x}N$ quantum wells grown by molecular beam epitaxy with thickness control at the monolayer scale, Appl. Phys. Lett. **73**(9), 1260 (1998)
31.609 J. Seo Im, H. Kollmer, J. Off, A. Sohmer, F. Scholz, A. Hangleiter: Reduction of oscillator strength due to piezoelectric fields in GaN/$Al_xGa_{1-x}N$ quantum wells, Phys. Rev. B **57**(16), R9435–R9438 (1998)
31.610 P. Lefebvre, J. Allègre, B. Gil, H. Mathieu, N. Grandjean, M. Leroux, J. Massies, P. Bigenwald: Time-resolved photoluminescence as a probe of internal electric fields in GaN-(GaAl)N quantum wells, Phys. Rev. B **59**(23), 15363 (1999)
31.611 M. Zamfirescu, B. Gil, N. Grandjean, G. Malpuech, A. Kavokin, P. Bigenwald, J. Massies: Extremely sharp dependence of the exciton oscillator strength on quantum-well width in the GaN/$Al_xGa_{1-x}N$ system: The polarization field effect, Phys. Rev. B **64**(12), 121304 (2001)
31.612 F. Yang, M. Wilkinson, E.J. Austin, K.P. O'Donnell: Origin of the stokes shift a geometrical model of exciton spectra in 2-D semiconductors_I, Phys. Rev. Lett. **71**(8), 1287 (1993)
31.613 F. Yang, M. Wilkinson, E.J. Austin, K.P. O'Donnell: Origin of the stokes shift a geometrical model of exciton spectra in 2-D semiconductors, Phys. Rev. Lett. **70**(3), 323 (1993)
31.614 F. Bernardini, V. Fiorentini, D. Vanderbilt: Spontaneous polarization and piezoelectric constants of III–V nitrides, Phys. Rev. B **56**(6), R10024 (1997)
31.615 J. Seo Im, H. Kollmer, J. Off, A. Sohmer, F. Scholz, A. Hangleiter: Effects of piezoelectric fields in GaInN/GaN and GaN/AlGaN heterostructures and quantum wells, Proc. Nitride Semicond. Symp (1997) p. 513
31.616 J.C. Harris, T. Someya, S. Kako, K. Hoshino, Y. Arakawa: Time-resolved photoluminescence of GaN/$Al_{0.5}Ga_{0.5}N$ quantum wells, Appl. Phys. Lett. **77**(7), 1005 (2000)
31.617 P. Waltereit, O. Brandt, A. Trampert, H.T. Grahn, J. Menniger, M. Ramsteiner, M. Reiche, K.H. Ploog: Nitride semiconductors free of electrostatic fields for efficient white light-emitting diodes, Nature **406**(6798), 865 (2000)
31.618 E. Kuokstis, C.Q. Chen, M.E. Gaevski, W.H. Sun, J.W. Yang, G. Simin, M. Asif Khan, H.P. Maruska, D.W. Hill, M.C. Chou, J.J. Gallagher, B. Chai: Polarization effects in photoluminescence of *c*- and *m*-plane GaN/AlGaN multiple quantum wells, Appl. Phys. Lett. **81**(22), 4130 (2002)
31.619 H.M. Ng: Molecular-beam epitaxy of GaN/$Al_xGa_{1-x}N$ multiple quantum wells on *r*-plane $(10\bar{1}2)$ sapphire substrates, Appl. Phys. Lett. **80**(23), 4369 (2002)
31.620 M.D. Craven, S.H. Lim, F. Wu, J.S. Speck, S.P. DenBaars: Structural characterization of nonpolar $(11\bar{2}0)$ *a*-plane GaN thin films grown on $(1\bar{1}02)$ *r*-plane sapphire, Appl. Phys. Lett. **81**(3), 469 (2002)
31.621 H.M. Ng, A. Bell, F.A. Ponce, S.N.G. Chu: Structural and optical characterization of nonpolar GaN/AlN quantum wells, Appl. Phys. Lett. **83**(4), 653 (2003)
31.622 A. Chakraborty, B.A. Haskell, S. Keller, J.S. Speck, S.P. DenBaars, S. Nakamura, U.K. Mishra: Nonpolar InGaN/GaN emitters on reduced-defect lateral epitaxially overgrown a -plane GaN with drive-current-independent electroluminescence emission peak, Appl. Phys. Lett. **85**(22), 5143 (2004)
31.623 C.Q. Chen, M.E. Gaevski, W.H. Sun, E. Kuokstis, J.P. Zhang, R.S.Q. Fareed, H.M. Wang, J.W. Yang, G. Simin, M.A. Khan, H.-P. Maruska, D.W. Hill, M.M.C. Chou, B. Chai: GaN homoepitaxy on free-standing $(11\bar{0}0)$ oriented GaN substrates, Appl. Phys. Lett. **81**(17), 3194 (2002)
31.624 S. Juršènas, E. Kuokštis, S. Miasojedovas, G. Kurilčik, A. Žukauskas, C.Q. Chen, J.W. Yang, V. Adivarahan, M. Asif Khan: Increase of free carrier lifetime in nonpolar *a*-plane GaN grown by epitaxial lateral overgrowth, Appl. Phys. Lett. **85**(5), 771 (2004)
31.625 M.D. Craven, P. Waltereit, J.S. Speck, S.P. DenBaars: Well-width dependence of photoluminescence emission from *a*-plane GaN/AlGaN multiple quantum wells, Appl. Phys. Lett. **84**(4), 496 (2004)
31.626 L. Jiawei, Y. Zhizhen, N.M. Nasser: GaN-based quantum dots, Phys. E Low-Dimens. Syst. Nanostructures **16**(2), 244 (2003)
31.627 K. Kato, T. Kusunoki, C. Takenaka, T. Tanahashi, K. Nakajima: Reduction of dislocations in InGaAs layer on GaAs using epitaxial lateral overgrowth, J. Cryst. Growth **115**(1–4), 174 (1991)
31.628 Y. Miyamoto, A. Sano, E. Yoshida, T. Sakano: Ultra-high-capacity digital coherent optical transmission technology, NTT Tech. Rev. **9**(8), 1 (2011)
31.629 M. Kondow, T. Kitatani, S. Nakatsuka, M.C. Larson, K. Nakahara, Y. Yazawa, M. Okai, K. Uomi: GaInNAs: A novel material for long-wavelength semiconductor lasers, IEEE J. Sel. Top. Quant. **3**(3), 719 (1997)
31.630 J.S. Harris: Tunable long-wavelength vertical-cavity lasers: The engine of next generation optical networks?, Sel. Top. Quantum Electron. IEEE J. **6**(6), 1145 (2000)
31.631 K. Yang, C.P. Hains, J. Cheng: Efficient continuous-wave lasing operation of a narrow-stripe oxide-confined GaInNAs-GaAs multiquantum-well laser grown by MOCVD, Photonics Technol. Lett. IEEE **12**(1), 7 (2000)
31.632 M. Kondow, T. Kitatani, K. Nakahara, T. Tanaka: A 1.3 μm GaInNAs Laser Diode with a lifetime of over 1000 hours, Jpn. J. Appl. Phys. **38**(12A), L1355–L1356 (1999)
31.633 N.Y. Li, C.P. Hains, K. Yang, J. Lu, J. Cheng, P.W. Li: Organometallic vapor phase epitaxy growth and optical characteristics of almost 1.2 μm GaInNAs three-quantum-well laser diodes, Appl. Phys. Lett. **75**(8), 1051 (1999)
31.634 S. Sato, S. Satoh: Room-temperature pulsed operation of strained GaInNAs/GaAs double quantum well laser diode grown by metal organic chemical vapour deposition, Electron. Lett. **34**(15), 1495 (1998)
31.635 T. Kitatani, M. Kondow, K. Nakahara, M.C. Larson, K. Uomi: Temperature dependence of the thresh-

old current and the lasing wavelength in 1.3 μm GaInNAs/GaAs single quantum well laser diode, Opt. Rev. **5**(2), 69 (1998)

31.636 S. Nakatsuka, M. Kondow, T. Kitatani, Y. Yazawa, M. Okai: Index-guide GaInNAs laser diode for optical communications, Jpn. J. Appl. Phys. **37**(3B), 1380 (1998)

31.637 A. Ougazzaden, S. Bouchoule, A. Mereuta, E.V.K. Rao, J. Decobert: Room temperature laser operation of bulk InGaAsN/GaAs structures grown by AP-MOVPE using N_2 as carrier gas, Electron. Lett. **35**(6), 474 (1999)

31.638 M. Fischer, M. Reinhardt, A. Forchel: High temperature operation of GaInAsN laserdiodes in the 1.3 μm regime, Proc. Device Res. Conf., Conf. Digest. 58th DRC (2000) p. 119

31.639 H. Shimizu, K. Kumada, S. Uchiyama, A. Kasukawa: High performance CW 1.26 μm GaInNAsSb-SQW and 1.20 μm GaInAsSb-SQW ridge lasers, Electron. Lett. **36**(20), 1701 (2000)

31.640 C. Setiagung, H. Shimizu, Y. Ikenaga, K. Kumada, A. Kasukawa: Very low threshold current density of 1.3 μm-range GaInNAsSb-GaNAs3 and 5 QWs lasers, Sel. Top. Quantum Electron. IEEE J. **9**(5), 1209 (2003)

31.641 S.R. Bank, H.B. Yuen, H. Bae, M.A. Wistey, J.S. Harris: Overannealing effects in GaInNAs(Sb) alloys and their importance to laser applications, Appl. Phys. Lett. **88**(22), 221115 (2006)

31.642 D. Gollub, S. Moses, M. Fischer, A. Forchel: 1.42 μm continuous-wave operation of GaInNAs laser diodes, Electron. Lett. **39**(10), 777 (2003)

31.643 N. Tansu, L.J. Mawst: Low-threshold strain-compensated InGaAs(N) ($\lambda = 1.19$–1.31 μm) quantum-well lasers, Photonics Technol. Lett. IEEE **14**(4), 444 (2002)

31.644 C.S. Peng, T. Jouhti, P. Laukkanen, E.-M. Pavelescu, J. Konttinen, W. Li, M. Pessa: 1.32 μm GaInNAs-GaAs laser with a low threshold current density, Photonics Technol. Lett. IEEE **14**(3), 275 (2002)

31.645 X. Yang, M.J. Jurkovic, J.B. Heroux, W.I. Wang: Molecular beam epitaxial growth of InGaAsN:Sb/GaAs quantum wells for long-wavelength semiconductor lasers, Appl. Phys. Lett. **75**(2), 178 (1999)

31.646 X. Yang, J.B. Héroux, L.F. Mei, W.I. Wang: InGaAsNSb/GaAs quantum wells for 1.55 μm lasers grown by molecular-beam epitaxy, Appl. Phys. Lett. **78**(26), 4068 (2001)

31.647 S.R. Bank, M.A. Wistey, L.L. Goddard, H.B. Yuen, V. Lordi, J.S. Harris: Low-threshold continuous-wave 1.5 μm GaInNAsSb lasers grown on GaAs, Quantum Electron. IEEE J. Of **40**(6), 656 (2004)

31.648 M.O. Fischer, M. Reinhardt, A. Forchel: Room-temperature operation of GaInAsN-GaAs laser diodes in the 1.5 μm range, Sel. Top. Quantum Electron. IEEE J. **7**(2), 149 (2001)

31.649 D.A. Livshits, A.Y. Egorov, H. Riechert: 8 W continuous wave operation of InGaAsN lasers at 1.3 μm, Electron. Lett. **36**(16), 1381 (2000)

31.650 F. Bugge, G. Erbert, J. Fricke, S. Gramlich, R. Staske, H. Wenzel, U. Zeimer, M. Weyers: 12 W continuous-wave diode lasers at 1120 nm with InGaAs quantum wells, Appl. Phys. Lett. **79**(13), 1965 (2001)

31.651 S. Okur, R. Shimada, F. Zhang, S.D.A. Hafiz, J. Lee, V. Avrutin, H. Morkoc, A. Franke, F. Bertram, J. Christen, U. Özgür: GaN-based vertical cavities with all dielectric reflectors by epitaxial lateral overgrowth, Jpn. J. Appl. Phys. **52**(8), 08JH03 (2013)

31.652 E. Wong: Energy efficient passive optical networks with low power VCSELs, Proc. Annu. Conf. Wirel. Optical Commun. (WOCC) (2012) p. 48

31.653 M.A. Khan, J.N. Kuznia, J.M. van Hove, D.T. Olson: Reflective filters based on single-crystal GaN/$Al_xGa_{1-x}N$ multilayers deposited using low-pressure metalorganic chemical vapor deposition, Appl. Phys. Lett. **59**(12), 1449 (1991)

31.654 T. Someya, Y. Arakawa: Highly reflective GaN/$Al_{0.34}Ga_{0.66}N$ quarter-wave reflectors grown by metal organic chemical vapor deposition, Appl. Phys. Lett. **73**(25), 3653 (1998)

31.655 N. Nakada, H. Ishikawa, T. Egawa, T. Jimbo, M. Umeno: MOCVD growth of high reflective GaN/AlGaN distributed Bragg reflectors, 13th Int, Conf. Cryst. Growth Conj. Unction Elev. Int. Conf. Vap. Growth Epitaxy **237–239**, 961 (2002)

31.656 M. Arita, M. Nishioka, Y. Arakawa: InGaN vertical microcavity LEDs with a Si-doped AlGaN/GaN distributed Bragg reflector, Physica Status Solidi C **194**, 403 (2002)

31.657 R. Langer, A. Barski, J. Simon, N.T. Pelekanos, O. Konovalov, R.A. Le Si Dang: High-reflectivity GaN/GaAlN Bragg mirrors at blue/green wavelengths grown by molecular beam epitaxy, Appl. Phys. Lett. **74**(24), 3610 (1999)

31.658 R. Butté, G. Christmann, E. Feltin, A. Castiglia, J. Levrat, G. Cosendey, A. Altoukhov, J.-F. Carlin, N. Grandjean: Room temperature polariton lasing in III-nitride microcavities, a comparison with blue GaN-based vertical cavity surface emitting lasers, Proc. Gallium Nitride Mat. and Devices IV (2009) p. 721619

31.659 C. Holder, D. Feezell, J.S. Speck, S.P. DenBaars, S. Nakamura: Demonstration of nonpolar GaN-based vertical-cavity surface-emitting lasers, Proc. SPIE Vert.-Cavity Surf.-Emitting Lasers XVII, Vol. 8639 (2013) p. 863906

31.660 D. Kasahara, D. Morita, T. Kosugi, K. Nakagawa, J. Kawamata, Y. Higuchi, H. Matsumura, T. Mukai: Demonstration of blue and green GaN-based vertical-cavity surface-emitting lasers by current injection at room temperature, Appl. Phys. Express **4**(7), 072103 (2011)

31.661 H.-Y. Ryu: Extraction efficiency in GaN nanorod light-emitting diodes investigated by finite-difference time-domain simulation, J. Korean Phys. Soc. **58**(4), 878 (2011)

31.662 Y.-J. Liu, C.-C. Huang, T.-Y. Chen, C.-S. Hsu, J.-K. Liou, W.-C. Liu: Improved performance of an InGaN-based light-emitting diode With a p-

GaN/n-GaN barrier junction, IEEE J. Quantum Electron. **47**(6), 755 (2011)
31.663 L.-B. Chang, M.-J. Lai, R.-M. Lin, C.-H. Huang: Effect of electron leakage on efficiency droop in wide-well InGaN-based light-emitting diodes, Appl. Phys. Express **4**(1), 012106 (2011)
31.664 C.-H. Chiu, C.-C. Lin, D.-M. Deng, D.-W. Lin, J.-C. Li, Z.-Y. Li, G.-W. Shu, T.-C. Lu, J.-L. Shen, H.-C. Kuo, K.-M. Lau: Optical and electrical properties of GaN-based light emitting diodes grown on micro- and nano-scale patterned Si substrate, IEEE J. Quantum Electron. **47**(7), 899 (2011)
31.665 S. Gradecak, F. Qian, Y. Li, H.-G. Park, C.M. Lieber: GaN nanowire lasers with low lasing thresholds, Appl. Phys. Lett. **87**(17), 173111 (2005)
31.666 H.-G. Park, F. Qian, C.J. Barrelet, Y. Li: Microstadium single-nanowire laser, Appl. Phys. Lett. **91**(25), 251115 (2007)
31.667 M.A. Zimmler, F. Capasso, S. Muller, C. Ronning: Optically pumped nanowire lasers: Invited review, Semicond. Sci. Technol. **25**(2), 024001 (2010)
31.668 S. Arafin, X. Liu, Z. Mi: Review of recent progress of III-nitride nanowire lasers, J. Nanophotonics **7**(1), 074599 (2013)
31.669 H.-J. Choi, J.C. Johnson, R. He, S.-K. Lee, F. Kim, P. Pauzauskie, J. Goldberger, R.J. Saykally, P. Yang: Self-organized GaN quantum wire UV lasers, J. Phys. Chem. B **107**(34), 8721 (2003)
31.670 D. Saxena, S. Mokkapati, P. Parkinson, N. Jiang, Q. Gao, H.H. Tan, C. Jagadish: Optically pumped room-temperature GaAs nanowire lasers, Nat. Photonics **7**(12), 963 (2013)
31.671 B. Mayer, D. Rudolph, J. Schnell, S. Morkötter, J. Winnerl, J. Treu, K. Müller, G. Bracher, G. Abstreiter, G. Koblmüller, J.J. Finley: Lasing from individual GaAs-AlGaAs core-shell nanowires up to room temperature, Nat. Commun. **4**, 2931 (2013) doi:10.1038/ncomms3931
31.672 J. Treu, M. Bormann, H. Schmeiduch, M. Döblinger, S. Morkötter, S. Matich, P. Wiecha, K. Saller, B. Mayer, M. Bichler, M.-C. Amann, J.J. Finley, G. Abstreiter, G. Koblmüller: Enhanced luminescence properties of InAs–InAsP core-shell nanowires, Nano Lett. **13**(12), 6070 (2013)
31.673 S.J. Chang, S.M. Wang, P.C. Chang, C.H. Kuo, S.J. Young, T.P. Chen, S.L. Wu, B.R. Huang: GaN Schottky barrier photodetectors, IEEE Sens. J. **10**(10), 1609 (2010)
31.674 K.P. Korona: Nitride-based photodetectors containing quantum wells in tunable electric fields, J. Optoelectron. Adv. Mater. **11**(9), 1108 (2009)
31.675 H.C. Lee, Y.K. Su, J.C. Lin, Y.C. Cheng, S.L. Wu, Y.D. Jhou: AlInGaN metal-insulator-semiconductor photodetectors at UV-C 280 nm, Electrochem. Solid-State Lett. **12**(10), 357 (2009)
31.676 E. Muñoz: (Al,In,Ga)N-based photodetectors. Some materials issues, Phys. Status Solidi B **244**(8), 2859 (2007)
31.677 M.-H. Kim, M.F. Schubert, Q. Dai, J.K. Kim, E.F. Schubert, J. Piprek, Y. Park: Origin of efficiency droop in GaN-based light-emitting diodes, Appl. Phys. Lett. **91**(18), 183507 (2007)
31.678 X. Ni, X. Li, J. Lee, S. Liu, V. Avrutin, U. Ozgur, H. Morkoç, A. Matulionis, T. Paskova, G. Mulholland, K.R. Evans: The effect of ballistic and quasi-ballistic electrons on the efficiency droop of InGaN light emitting diodes, Phys. Status Solidi RRL Rapid Res. Lett. **4**(8/9), 194 (2010)
31.679 J. Xie, X. Ni, Q. Fan, R. Shimada, U. Ozgur, H. Morkoç: On the efficiency droop in InGaN multiple quantum well blue light emitting diodes and its reduction with p-doped quantum well barriers, Appl. Phys. Lett. **93**(12), 121107 (2008)
31.680 C. Guthy, C.-Y. Nam, J.E. Fischer: Unusually low thermal conductivity of gallium nitride nanowires, J. Appl. Phys. **103**(6), 064319 (2008)
31.681 J. Zou: Lattice thermal conductivity of freestanding gallium nitride nanowires, J. Appl. Phys. **108**(3), 034324 (2010)
31.682 Z. Wang, X. Zu, F. Gao, W.J. Weber, J.-P. Crocombette: Atomistic simulation of the size and orientation dependences of thermal conductivity in GaN nanowires, Appl. Phys. Lett. **90**(16), 161923 (2007)
31.683 M.L. Colussi, R.J. Baierle, R.H. Miwa: Doping effects of C, Si and Ge in wurtzite [0001] GaN, AlN, and InN nanowires, J. Appl. Phys. **110**(3), 033709 (2011)
31.684 S. Zhao, S. Fathololoumi, K.H. Bevan, D.P. Liu, M.G. Kibria, Q. Li, G.T. Wang, H. Guo, Z. Mi: Tuning the surface charge properties of epitaxial InN nanowires, Nano Lett. **12**(6), 2877 (2012)
31.685 A. Miehr, O. Ambacher, W. Rieger, T. Metzger, E. Born, R.A. Fischer: The first monomeric, volatile bis-azide single-source precursor to Gallium nitride thin films, Chem. Vap. Depos. **2**(2), 51 (1996)
31.686 A.C. Jones, S.A. Rushworth, D.J. Houlton, J.S. Roberts, V. Roberts, C.R. Whitehouse, G.W. Critchlow: Deposition of aluminum nitride thin films by MOCVD from the trimethylaluminum–ammonia adduct, Chem. Vap. Depos. **2**(1), 5 (1996)
31.687 M. Yoshizawa, A. Kikuchi, N. Fujita, K. Kushi, H. Sasamoto, K. Kishino: Self-organization of GaN/$Al_{0.18}Ga_{0.82}N$ multi-layer nano-columns on (0 0 0 1) Al_2O_3 by RF molecular beam epitaxy for fabricating GaN quantum disks, J. Cryst. Growth **189/190**(0), 138–141 (1998)
31.688 T. Karabacak, P.-I. Wang, G.-C. Wang, T.-M. Lu: Growth of single crystal tungsten nanorods by oblique angle sputter deposition, Proc. Contin. Nanophase and Nanostructured Mat. (2003) p. 75
31.689 Y.Y. Kim, B.H. Kong, H.K. Cho: Vertically arrayed Ga-doped ZnO nanorods grown by magnetron sputtering: The effect of Ga contents and microstructural evaluation, J. Cryst. Growth **330**(1), 17 (2011)
31.690 F. Shi: GaN nanowires fabricated by magnetron sputtering deposition. In: *Nanowires – Fundamental Research*, ed. by A. Hashim (InTech 2011) doi:10.5772/16355, available from https://www.intechopen.com/books/nanowires-fundamental-research/gan-nanowires-fabricated-bymagnetron-sputtering-deposition

31.691 M. Junaid: *Magnetron Sputter Epitaxy of GaN Epilayers and Nanorods*, Ph.D. Thesis (Linköping Studies in Science and Technology, Sweden 2012)

31.692 H. Zhuang, S. Xue: GaN nanowires synthesized on Si(111) substrates by magnetron sputtering and ammonification technique, Jpn. J. Appl. Phys. **46**(10R), 6886 (2007)

31.693 R.S. Wagner, W.C. Ellis: Vapour–liquid–solid mechanism of single crystal growth, Appl. Phys. Lett. **4**(5), 89 (1964)

31.694 M. He, I. Minus, P. Zhou, S.N. Mohammed, J.B. Halpern, R. Jacobs, W.L. Sarney, L. Salamanca-Riba, R.D. Vispute: Growth of large-scale GaN nanowires and tubes by direct reaction of Ga with NH_3, Appl. Phys. Lett. **77**(23), 3731 (2000)

31.695 C.C. Tang, S.S. Fan, H.Y. Dang, P. Li, Y.M. Liu: Simple and high-yield method for synthesizing single-crystal GaN nanowires, Appl. Phys. Lett. **77**(13), 1961 (2000)

31.696 C.H. Liang, L.C. Chen, J.S. Hwang, K.H. Chen, Y.T. Hung, Y.F. Chen: Selective-area growth of indium nitride nanowires on gold-patterned Si(100) substrates, Appl. Phys. Lett. **81**(1), 22 (2002)

31.697 H. Parala, A. Devi, F. Hipler, E. Maile, A. Birkner, H.W. Becker, R.A. Fischer: Investigations on InN whiskers grown by chemical vapour deposition, J. Cryst. Growth **231**(1/2), 68 (2001)

31.698 Z.R. Dai, Z.W. Pan, Z.L. Wang: Novel nanostructures of functional oxides synthesized by thermal evaporation, Adv. Funct. Mater. **13**(1), 9 (2003)

31.699 R.-Q. Zhang, Y. Lifshitz, S.-T. Lee: Oxide-assisted growth of semiconducting nanowires, Adv. Mater. **15**(7/8), 635 (2003)

31.700 C.M. Balkaş, R.F. Davis: Synthesis routes and characterization of high-purity, single-phase gallium nitride powders, J. Am. Ceram. Soc. **79**(9), 2309 (1996)

31.701 Y. Zhang, J. Liu, R. He, Q. Zhang, X. Zhang, J. Zhu: Synthesis of aluminum nitride nanowires from carbon nanotubes, Chem. Mater. **13**(11), 3899 (2001)

31.702 J. Liu, X. Zhang, Y. Zhang, R. He, J. Zhu: Novel synthesis of AlN nanowires with controlled diameters, J. Mater. Res. **16**(11), 3133 (2001)

31.703 C.R. Martin: Nanomaterials: A membrane-based synthetic approach, Science **266**(5193), 1961 (1994)

31.704 J. Zhang, X. Peng, X. Wang, Y. Wang, L. Zhang: Micro-Raman investigation of GaN nanowires prepared by direct reaction Ga with NH_3, Chem. Phys. Lett. **345**(5/6), 372 (2001)

31.705 F. Qian, Y. Li, S. Gradecak, D. Wang, C.J. Barrelet, C.M. Lieber: Gallium nitride-based nanowire radial heterostructures for nanophotonics, Nano Lett. **4**(10), 1975 (2004)

31.706 Y.-L. Chang, J.L. Wang, F. Li, Z. Mi: High efficiency green, yellow, and amber emission from InGaN/GaN dot-in-a-wire heterostructures on Si(111), Appl. Phys. Lett. **96**(1), 013106 (2010)

31.707 R. Armitage, K. Tsubaki: Multicolour luminescence from InGaN quantum wells grown over GaN nanowire arrays by molecular-beam epitaxy, Nanotechnology **21**(19), 195202 (2010)

31.708 Y. Kawakami, S. Suzuki, A. Kaneta, M. Funato, A. Kikuchi, K. Kishino: Origin of high oscillator strength in green-emitting InGaN/GaN nanocolumns, Appl. Phys. Lett. **89**(16), 163124 (2006)

31.709 H.-M. Kim, H. Lee, S.I. Kim, S.R. Ryu, T.W. Kang, K.S. Chung: Formation of InGaN nanorods with indium mole fractions by hydride vapor phase epitaxy, Proc. 5th Int. Symp. Blue Laser and Light Emitting Diodes (2004) p. 2802

31.710 H.-M. Kim, W.C. Lee, T.W. Kang, K.S. Chung, C.S. Yoon, C.K. Kim: InGaN nanorods grown on (111) silicon substrate by hydride vapor phase epitaxy, Chem. Phys. Lett. **380**(1/2), 181 (2003)

31.711 A. Kikuchi, M. Kawai, M. Tada, K. Kishino: InGaN/GaN multiple quantum disk nanocolumn light-emitting diodes grown on (111) Si substrate, Jpn. J. Appl. Phys. **43**(12A), L1524 (2004)

31.712 H.-W. Lin, Y.-J. Lu, H.-Y. Chen, H.M. Lee, S. Gwo: InGaN/GaN nanorod array white light-emitting diode, Appl. Phys. Lett. **97**(7), 073101 (2010)

31.713 M. Zervos, D. Tsokkou, M. Pervolaraki, A. Othonos: Low temperature growth of In203 and InN nanocrystals on Si(111) via chemical vapour deposition based on the sublimation of NH_4Cl in In, Nanoscale Res. Lett. **4**(6), 491 (2009)

31.714 J. Grandal, M.A. Sánchez-García, E. Calleja, E. Luna, A. Trampert: Accommodation mechanism of InN nanocolumns grown on Si(111) substrates by molecular beam epitaxy, Appl. Phys. Lett. **91**(2), 021902 (2007)

31.715 I. Shalish, G. Seryogin, W. Yi, J.M. Bao, M.A. Zimmler, E. Likovich, D.C. Bell, F. Capasso, V. Narayanamurti: Epitaxial catalyst-free growth of InN nanorods on *c*-plane sapphire, Nanoscale Res. Lett. **4**(6), 532 (2009)

31.716 K.R. Wang, S.J. Lin, L.W. Tu, M. Chen, Q.Y. Chen, T.H. Chen, M.L. Chen, H.W. Seo, N.H. Tai, S.C. Chang, I. Lo, D.P. Wang, W.K. Chu: InN nanotips as excellent field emitters, Appl. Phys. Lett. **92**(12), 123105 (2008)

31.717 M.-S. Hu, G.-M. Hsu, K.-H. Chen, C.-J. Yu, H.-C. Hsu, L.-C. Chen, J.-S. Hwang, L.-S. Hong, Y.-F. Chen: Infrared lasing in InN nanobelts, Appl. Phys. Lett. **90**(12), 123109 (2007)

31.718 P.J. Pauzauskie, P. Yang: Nanowire photonics, Mater. Today **9**(10), 36 (2006)

31.719 R. Yan, D. Gargas, P. Yang: Nanowire photonics, Nat. Photonics **3**(10), 569 (2009)

31.720 D. Saxena, S. Mokkapati, C. Jagadish: Semiconductor nanolasers, IEEE Photonics J **4**(2), 582 (2012)

31.721 D. Vanmaekelbergh, L.K. van Vugt: ZnO nanowire lasers, Nanoscale **3**(7), 2783 (2011)

31.722 C. Ning: Semiconductor nanowire lasers. In: *Advances in Semiconductor Lasers*, ed. by J.J. Coleman, A.C. Bryce, C. Jagadish (Academic, Burlington 2012) p. 455

31.723 L.C. Chen, K.H. Chen, C.C. Chen: Group III- and group IV-nitride nanorods and nanowires. In: *Nanowires and Nanobelts – Materials Properties and Devices: Metal and Semiconductor Nanostructures*, Vol. 1, ed. by Z.L. Wang (Springer, New

York 2003) p. 257
31.724 M. Law, J. Goldberger, P. Yang: Semiconductor nanowires and nanotubes, Annu. Rev. Mater. Res. **34**, 83 (2004)
31.725 G.T. Wang, A.A. Talin, D.J. Werder, J.R. Creighton, E. Lai, R.J. Anderson, I. Arslan: Highly aligned, template-free growth and characterization of vertical GaN nanowires on sapphire by metal–organic chemical vapour deposition, Nanotechnology **17**(23), 5773 (2006)
31.726 L. Cerutti, J. Ristic, S. Fernandez-Garrido, E. Calleja, A. Trampert, K.H. Ploog, S. Lazic, J.M. Calleja: Wurtzite GaN nanocolumns grown on Si(001) by molecular beam epitaxy, Appl. Phys. Lett. **88**(21), 213114 (2006)
31.727 H.-M. Kim, T.W. Kang, K.S. Chung: Nanoscale ultraviolet-light-emitting diodes using wide-bandgap gallium nitride nanorods, Adv. Mater. **15**(7/8), 567 (2003)
31.728 B. Ha, S.H. Seo, J.H. Cho, C.S. Yoon, J. Yoo, G.-C. Yi, C.Y. Park, C.J. Lee: Optical and field emission properties of thin single-crystalline GaN nanowires, J. Phys. Chem. B **109**(22), 11095 (2005)
31.729 T. Kuykendall, P. Pauzauskie, S. Lee, Y. Zhang, J. Goldberger, P. Yang: Metalorganic chemical vapor deposition route to GaN nanowires with triangular cross sections, Nano Lett. **3**(8), 1063 (2003)
31.730 K. Kishino, A. Kikuchi: Highly efficient blue to red emissions of InGaN/GaN nano-disks integrated into GaN nanocolumns, Proc. IEEE LEOS Annu. Meet. (2005) p. 631
31.731 P. Schlotter, R. Schmidt, J. Schneider: Luminescence conversion of blue light emitting diodes, Appl. Phys. Mater. Sci. Process. **64**(4), 417 (1997)
31.732 H.P.T. Nguyen, M. Djavid, K. Cui, Z. Mi: Temperature-dependent nonradiative recombination processes in GaN-based nanowire white-light-emitting diodes on silicon, Nanotechnology **23**(19), 194012 (2012)
31.733 H.P.T. Nguyen, K. Cui, S. Zhang, M. Djavid, A. Korinek, G.A. Botton, Z. Mi: Controlling electron overflow in phosphor-free InGaN/GaN nanowire white light-emitting diodes, Nano Lett. **12**(3), 1317 (2012)
31.734 J.C. Johnson, H.-J. Choi, K.P. Knutsen, R.D. Schaller, P. Yang, R.J. Saykally: Single gallium nitride nanowire lasers, Nat. Mater. **1**(2), 106 (2002)
31.735 F. Qian, Y. Li, S. Gradecak, H.-G. Park, Y. Dong, Y. Ding, Z.L. Wang, C.M. Lieber: Multi-quantum-well nanowire heterostructures for wavelength-controlled lasers, Nat. Mater. **7**(9), 701 (2008)
31.736 Q. Li, J.B. Wright, W.W. Chow, T.S. Luk, I. Brener, L.F. Lester, G.T. Wang: Single-mode GaN nanowire lasers, Opt. Express **20**(16), 17873 (2012)
31.737 H. Xu, J.B. Wright, T.-S. Luk, J.J. Figiel, K. Cross, L.F. Lester, G. Balakrishnan, G.T. Wang, I. Brener, Q. Li: Single-mode lasing of GaN nanowire-pairs, Appl. Phys. Lett. **101**(11), 113106 (2012)
31.738 H. Xu, J.B. Wright, A. Hurtado, Q. Li, T.-S. Luk, J.J. Figiel, K. Cross, G. Balakrishnan, L.F. Lester, I. Brener, G.T. Wang: Gold substrate-induced single-mode lasing of GaN nanowires, Appl. Phys. Lett. **101**(22), 221114 (2012)
31.739 C.-Y. Wu, C.-T. Kuo, C.-Y. Wang, C.-L. He, M.-H. Lin, H. Ahn, S. Gwo: Plasmonic green nanolaser based on a metal-oxide-semiconductor structure, Nano Lett. **11**(10), 4256 (2011)
31.740 Y.-J. Lu, J. Kim, H.-P. Chen, C. Wu, N. Dabidian, C.E. Sanders, C.-Y. Wang, M.-Y. Lu, B.-H. Li, X. Qiu, W.-R. Chang, L.-J. Chen, G. Shvets, C.-K. Shih, S. Gwo: Plasmonic nanolaser using epitaxially grown silver film, Science **337**(6093), 450 (2012)
31.741 J. Heo, W. Guo, P. Bhattacharya: Monolithic single GaN nanowire laser with photonic crystal microcavity on silicon, Appl. Phys. Lett. **98**(2), 021110 (2011)
31.742 P.J. Pauzauskie, D.J. Sirbuly, P. Yang: Semiconductor nanowire ring resonator laser, Phys. Rev. Lett. **96**(14), 143903 (2006)

32. Electron Transport Within III-V Nitride Semiconductors

Stephen K. O'Leary, Poppy Siddiqua, Walid A. Hadi, Brian E. Foutz, Michael S. Shur, Lester F. Eastman†

The III-V nitride semiconductors, gallium nitride, aluminum nitride, and indium nitride, have been recognized as promising materials for novel electronic and optoelectronic device applications for some time now. Since informed device design requires a firm grasp of the material properties of the underlying electronic materials, the electron transport that occurs within these III–V nitride semiconductors has been the focus of considerable study over the years. In an effort to provide some perspective on this rapidly evolving field, in this paper we review analyses of the electron transport within these III–V nitride semiconductors. In particular, we discuss the evolution of the field, compare and contrast results obtained by different researchers, and survey the more recent literature. In order to narrow the scope of this chapter, we will primarily focus on the electron transport within bulk wurtzite gallium nitride, aluminum nitride, and indium nitride for the purposes of this review. Most of our discussion will focus on results obtained from our ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within these materials, our results conforming with state-of-the-art III–V nitride semiconductor orthodoxy. Steady-state and transient electron transport results are presented. We conclude our discussion by presenting some recent developments on the electron transport within these materials.

32.1 **Electron Transport Within Semiconductors and the Monte Carlo Simulation Approach** 830
32.1.1 The Boltzmann Transport Equation..... 831
32.1.2 Our Ensemble Semi-Classical Monte Carlo Simulation Approach 832
32.1.3 Parameter Selections for Bulk Wurtzite GaN, AlN, and InN 832
32.2 **Steady-State and Transient Electron Transport Within Bulk Wurtzite GaN, AlN, and InN**..... 834
32.2.1 Steady-State Electron Transport Within Bulk Wurtzite GaN..... 835
32.2.2 Steady-State Electron Transport: A Comparison of the III–V Nitride Semiconductors with GaAs 836
32.2.3 Influence of Temperature on the Electron Drift Velocities Within GaN and GaAs..... 836
32.2.4 Influence of Doping on the Electron Drift Velocities Within GaN and GaAs..... 839
32.2.5 Electron Transport in AlN 841
32.2.6 Electron Transport in InN..... 842
32.2.7 Transient Electron Transport..... 843
32.2.8 Electron Transport: Conclusions..... 845
32.3 **Electron Transport Within III–V Nitride Semiconductors: A Review**..... 845
32.3.1 Evolution of the Field 846
32.3.2 Developments in the 21st Century..... 848
32.3.3 Future Perspectives..... 849
32.4 **Conclusions** 850
References..... 850

The III–V nitride semiconductors, gallium nitride (GaN), aluminum nitride (AlN), and indium nitride (InN), have been known as promising materials for novel electronic and optoelectronic device applications for some time now [32.1–4]. In terms of electronics, their wide energy gaps, large breakdown fields, high thermal conductivities, and favorable electron transport characteristics, make GaN, AlN, and InN, and alloys of these materials, ideally suited for novel high-power and high-frequency electron device applications. On the optoelectronics front, the direct nature of the energy gaps associated with GaN, AlN, and InN, make this family of materials, and its alloys, well suited for novel optoelectronic device applications in the visible and ultraviolet frequency range. While initial efforts to study these materials were hindered by growth diffi-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_32

culties, recent improvements in material quality have made the realization of a number of III–V nitride semiconductor-based electronic [32.5–9] and optoelectronic [32.9–12] devices possible. These developments have fuelled considerable interest in these III–V nitride semiconductors.

In order to analyze and improve the design of III–V nitride semiconductor-based devices, an understanding of the electron transport that occurs within these materials is necessary. Electron transport within bulk GaN, AlN, and InN has been examined extensively over the years [32.13–32]. Unfortunately, uncertainty in the material parameters associated with GaN, AlN, and InN remains a key source of ambiguity in the analysis of the electron transport within these materials [32.32]. In addition, some experimental [32.33] and theoretical [32.34] developments have cast doubt upon the validity of widely accepted notions upon which our understanding of the electron transport mechanisms within the III–V nitride semiconductors, GaN, AlN, and InN, has evolved. Another confounding matter is the sheer volume of research activity being performed on the electron transport within these materials, presenting the researcher with a dizzying array of seemingly disparate approaches and results. Clearly, at this critical juncture at least, our understanding of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, remains in a state of flux.

In order to provide some perspective on this rapidly evolving field, we aim to review analyses of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, within this chapter. In particular, we will discuss the evolution of the field and survey the more recent literature. In order to narrow the scope of this review, we will primarily focus on the electron transport within bulk wurtzite GaN, AlN, and InN for the purposes of this chapter. Most of our discussion will focus upon results obtained from our ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within these materials, our results conforming with state-of-the-art III–V nitride semiconductor orthodoxy. We hope that researchers in the field will find this review useful and informative.

We begin our review with the Boltzmann transport equation, which underlies most analyses of the electron transport within semiconductors. The ensemble semi-classical three-valley Monte Carlo simulation approach that we employ in order to solve this Boltzmann transport equation is then discussed. The material parameters corresponding to bulk wurtzite GaN, AlN, and InN are then presented. We then use these material parameter selections and our ensemble semi-classical three-valley Monte Carlo simulation approach to determine the nature of the steady-state and transient electron transport within the III–V nitride semiconductors. Finally, we present some developments on the electron transport within these materials.

This chapter is organized in the following manner. In Sect. 32.1, we present the Boltzmann transport equation and our ensemble semi-classical three-valley Monte Carlo simulation approach that we employ in order to solve this equation for the III–V nitride semiconductors, GaN, AlN, and InN. The material parameters, corresponding to bulk wurtzite GaN, AlN, and InN, are also presented in Sect. 32.1, these material parameters being updated for the specific case of wurtzite InN. Then, in Sect. 32.2, using results obtained from our ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within these III–V nitride semiconductors, we study the nature of the steady-state electron transport that occurs within these materials. Transient electron transport within the III–V nitride semiconductors is also discussed in Sect. 32.2. A review of the III–V nitride semiconductor electron transport literature, in which the evolution of the field is discussed and a survey of the more recent literature is presented, is then featured in Sect. 32.3. Finally, conclusions are provided in Sect. 32.4.

32.1 Electron Transport Within Semiconductors and the Monte Carlo Simulation Approach

The electrons within a semiconductor are in a perpetual state of motion. In the absence of an applied electric field, this motion arises as a result of the thermal energy that is present, and is referred to as thermal motion. From the perspective of an individual electron, thermal motion may be viewed as a series of trajectories, interrupted by a series of random scattering events. Scattering may arise as a result of interactions with the lattice atoms, impurities, other electrons, and defects. As these interactions lead to electron trajectories in all possible directions, i. e., there is no preferred direction, while individual electrons will move from one location to another, when taken as an ensemble, and assuming that the electrons are in thermal equilibrium, the overall electron distribution will remain static. Accordingly, no net current flow occurs.

With the application of an applied electric field $\boldsymbol{E}$ each electron in the ensemble will experience a force,

$-q\boldsymbol{E}$, where q denotes the magnitude of the electron charge. While this force may have a negligible impact upon the motion of any given individual electron, taken as an ensemble, the application of such a force will lead to a net aggregate motion of the electron distribution. Accordingly, a net current flow will occur, and the overall electron ensemble will no longer be in thermal equilibrium. This movement of the electron ensemble in response to an applied electric field, in essence, represents the fundamental issue at stake when we study the electron transport within a semiconductor.

In this section, we provide a brief tutorial on the issues at stake in our analysis of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN. We begin our analysis with an introduction to the Boltzmann transport equation. This equation describes how the electron distribution function evolves under the action of an applied electric field, and underlies the electron transport within bulk semiconductors. We then introduce the Monte Carlo simulation approach to solving this Boltzmann transport equation, focusing on the ensemble semi-classical three-valley Monte Carlo simulation approach used in our simulations of the electron transport within the III–V nitride semiconductors. Finally, we present the material parameters corresponding to bulk wurtzite GaN, AlN, and InN.

This section is organized in the following manner. In Sect. 32.1.1, the Boltzmann transport equation is introduced. Then, in Sect. 32.1.2, our ensemble semi-classical three-valley Monte Carlo simulation approach to solving this Boltzmann transport equation is presented. Finally, in Sect. 32.1.3, our material parameter selections, corresponding to bulk wurtzite GaN, AlN, and InN, are presented.

32.1.1 The Boltzmann Transport Equation

An electron ensemble may be characterized by its distribution function $f(\boldsymbol{r},\boldsymbol{p},t)$, where $\boldsymbol{r}$ denotes the position, $\boldsymbol{p}$ represents the momentum, and t indicates time. The response of this distribution function to an applied electric field $\boldsymbol{E}$ is the issue at stake when one investigates the electron transport within a semiconductor. When the dimensions of the semiconductor are large, and quantum effects are negligible, the ensemble of electrons may be treated as a continuum, so the corpuscular nature of the individual electrons within the ensemble, and the attendant complications which arise, may be neglected. In such a circumstance, the evolution of the distribution function $f(\boldsymbol{r},\boldsymbol{p},t)$ may be determined using the Boltzmann transport equation. In contrast, when the dimensions of the semiconductor are small, and quantum effects are significant, then the Boltzmann transport equation, and its continuum description of the electron ensemble, is no longer valid. In such a case, it is necessary to adopt quantum transport methods in order to study the electron transport within the semiconductor [32.35].

For the purposes of this analysis, we will focus on the electron transport within bulk semiconductors, i. e., semiconductors of sufficient dimensions so that the Boltzmann transport equation is valid. *Ashcroft* and *Mermin* [32.36] demonstrated that this equation can be expressed as

$$\frac{\partial f}{\partial t} = -\dot{\boldsymbol{p}}\cdot\nabla_{\mathrm{p}} f - \dot{\boldsymbol{r}}\cdot\nabla_{\mathrm{r}} f + \left.\frac{\partial f}{\partial t}\right|_{\mathrm{scat}} . \tag{32.1}$$

The first term on the right-hand side of (32.1) represents the change in the distribution function due to external forces applied to the system. The second term on the right-hand side of (32.1) accounts for the electron diffusion which occurs. The final term on the right-hand side of (32.1) describes the effects of scattering.

Owing to its fundamental importance in the analysis of the electron transport within semiconductors, a number of techniques have been developed over the years in order to solve the Boltzmann transport equation. Approximate solutions to the Boltzmann transport equation, such as the displaced Maxwellian distribution function approach of *Ferry* [32.14] and *Das* and *Ferry* [32.15] and the nonstationary charge transport analysis of *Sandborn* et al. [32.37], have proven useful. Low-field approximate solutions have also proven elementary and insightful [32.17, 20, 38]. A number of these techniques have been applied to the analysis of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN [32.14, 15, 17, 20, 38, 39]. Alternatively, more sophisticated techniques have been developed which solve the Boltzmann transport equation directly. These techniques, while allowing for a rigorous solution of the Boltzmann transport equation, are rather involved, and require intense numerical analysis. They are further discussed by *Nag* [32.40].

For studies of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, by far the most common approach to solving the Boltzmann transport equation has been the ensemble semi-classical Monte Carlo simulation approach. Of the III–V nitride semiconductors, the electron transport within GaN has been studied the most extensively using this ensemble Monte Carlo simulation approach [32.13, 16, 18, 19, 21, 22, 27, 29, 32], with AlN [32.24, 25, 29] and InN [32.23, 28, 29, 31] less so. The Monte Carlo simulation approach has also been used to study the electron transport within the two-dimensional electron gas of the AlGaN/GaN interface which occurs in high electron

mobility AlGaN/GaN field-effect transistors [32.41, 42].

At this point, it should be noted that the complete solution of the Boltzmann transport equation requires the resolution of both steady-state and transient responses. Steady-state electron transport refers to the electron transport that occurs long after the application of an applied electric field, i. e., once the electron ensemble has settled to a new equilibrium state (we are not necessarily referring to thermal equilibrium here, since thermal equilibrium is only achieved in the absence of an applied electric field). As the distribution function is difficult to visualize quantitatively, researchers typically study the dependence of the electron drift velocity (the average electron velocity determined by statistically averaging over the entire electron ensemble) on the applied electric field in the analysis of steady-state electron transport; in other words, they determine the velocity–field characteristic. Transient electron transport, by way of contrast, refers to the transport that occurs while the electron ensemble is evolving into its new equilibrium state. Typically, it is characterized by studying the dependence of the electron drift velocity on the time elapsed, or the distance displaced, since the electric field was initially applied. Both steady-state and transient electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, are reviewed within this chapter.

32.1.2 Our Ensemble Semi-Classical Monte Carlo Simulation Approach

For the purposes of our analysis of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, we employ ensemble semi-classical Monte Carlo simulations. A three-valley model for the conduction band is employed. Nonparabolicity is considered in the lowest conduction band valley, this nonparabolicity being treated through the application of the Kane model [32.43].

In the Kane model, the energy band of the Γ valley is assumed to be nonparabolic, spherical, and of the form

$$\frac{\hbar^2 k^2}{2m^*} = E\,(1+\alpha E)\,, \tag{32.2}$$

where $\hbar k$ denotes the magnitude of the crystal momentum, E represents the energy above the minimum, m^* is the effective mass, and the nonparabolicity coefficient α is given by

$$\alpha = \frac{1}{E_g}\left(1-\frac{m^*}{m_e}\right)^2\,, \tag{32.3}$$

where m_e and E_g denote the free electron mass and the energy gap, respectively [32.43].

The scattering mechanisms considered in our analysis are:

1. Ionized impurity
2. Polar optical phonon
3. Piezoelectric [32.44, 45]
4. Acoustic deformation potential.

Intervalley scattering is also considered. Piezoelectric scattering is treated using the well established zinc blende scattering rates, and so a suitably transformed piezoelectric constant, e_{14}, must be selected. This may be achieved through the transformation suggested by *Bykhovski* et al. [32.44, 45]. We also assume that all donors are ionized and that the free electron concentration is equal to the dopant concentration. The motion of three thousand electrons is examined in our steady-state electron transport simulations, while the motion of ten thousand electrons is considered in our transient electron transport simulations. The crystal temperature is set to 300 K and the doping concentration is set to $10^{17}\,\mathrm{cm}^{-3}$ in all cases, unless otherwise specified. Electron degeneracy effects are accounted for by means of the rejection technique of *Lugli* and *Ferry* [32.46]. Electron screening is also accounted for following the Brooks–Herring method [32.47]. Further details of our approach are discussed in the literature [32.16, 21–24, 29, 32, 48].

32.1.3 Parameter Selections for Bulk Wurtzite GaN, AlN, and InN

The material parameter selections, used for our simulations of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, are tabulated in Table 32.1. These parameter selections are the same as those employed by *Foutz* et al. [32.29], with the exception of the case of InN, whose material parameters have been updated [32.49, 50]; following the lead of *Polyakov* et al. [32.50], the parameters corresponding to InN are as set by *Hadi* et al. [32.49]. While the band structures corresponding to bulk wurtzite GaN, AlN, and InN are still not agreed upon, the band structures of *Lambrecht* and *Segall* [32.51] are adopted for the purposes of this analysis except for the case of InN; for the case of InN, following the lead of *Polyakov* et al. [32.50], the band structure is as set by *Hadi* et al. [32.49]. For the case of bulk wurtzite GaN, the analysis of *Lambrecht* and *Segall* [32.51] suggests that the lowest point in the conduction band is located at the center of the Brillouin zone, at the Γ point, the first upper conduction band valley minimum also occurring at the Γ point, 1.9 eV above the lowest point

Table 32.1 The material parameter selections corresponding to bulk wurtzite GaN, AlN, and InN. (These parameter selections are from *Foutz* et al. [32.29] and *Hadi* et al. [32.49])

Parameter	GaN	AlN	InN
Mass density (g/cm^3)	6.15	3.23	6.81
Longitudinal sound velocity (cm/s)	6.56×10^5	9.06×10^5	6.24×10^5
Transverse sound velocity (cm/s)	2.68×10^5	3.70×10^5	2.55×10^5
Acoustic deformation potential (eV)	8.3	9.5	7.1
Static dielectric constant	8.9	8.5	15.3
High-frequency dielectric constant	5.35	4.77	8.4
Effective mass (Γ_1 valley)	$0.20\,m_e$	$0.48\,m_e$	$0.04\,m_e$
Piezoelectric constant, e_{14} (C/cm^2)	3.75×10^{-5}	9.2×10^{-5}	3.75×10^{-5}
Direct energy gap (eV)	3.39	6.2	0.7
Optical phonon energy (meV)	91.2	99.2	73.0
Intervalley deformation potentials (eV/cm)	10^9	10^9	10^9
Intervalley phonon energies (meV)	91.2	99.2	73.0

Table 32.2 The band structure parameter selections corresponding to bulk wurtzite GaN, AlN, and InN. (These parameter selections are from *Foutz* et al. [32.29] and *Hadi* et al. [32.49]. These parameters were originally determined from the band structural calculations of *Polyakov* et al. [32.50] and *Lambrecht* and *Segall* [32.51])

		Valley number		
		1	2	3
GaN	Valley location	Γ_1	Γ_2	L–M
	Valley degeneracy	1	1	6
	Effective mass	$0.2\,m_e$	m_e	m_e
	Intervalley energy separation (eV)	–	1.9	2.1
	Energy gap (eV)	3.39	5.29	5.49
	Nonparabolicity (eV^{-1})	0.189	0.0	0.0
AlN	Valley location	Γ_1	L-M	K
	Valley degeneracy	1	6	2
	Effective mass	$0.48\,m_e$	m_e	m_e
	Intervalley energy separation (eV)	–	0.7	1.0
	Energy gap (eV)	6.2	6.9	7.2
	Nonparabolicity (eV^{-1})	0.044	0.0	0.0
InN	Valley location	Γ_1	Γ_2	L–M
	Valley degeneracy	1	1	6
	Effective mass	$0.04\,m_e$	$0.25\,m_e$	m_e
	Intervalley energy separation (eV)	–	1.775	2.709
	Energy gap (eV)	0.7	2.475	3.409
	Nonparabolicity (eV^{-1})	1.43	0.0	0.0

in the conduction band, the second upper conduction band valley minima occurring along the symmetry lines between the L and M points, 2.1 eV above the lowest point in the conduction band; see Table 32.2. The other materials considered in this analysis have suitably modified band structures, as described in Table 32.2. We ascribe an effective mass equal to the free electron mass m_e to all of the upper conduction band valleys, with the exception of the case of InN, the values suggested by *Polyakov* et al. [32.50] and *Hadi* et al. [32.49] being employed for the specific case of this material. The nonparabolicity coefficient α corresponding to each upper conduction band valley is set to zero, so the upper conduction band valleys are assumed to be completely parabolic. For our simulations of the electron transport within gallium arsenide (GaAs), the material parameters employed are mostly from *Littlejohn* et al. [32.52], although it should be noted that the mass density, the energy gap, and the sound velocities are from *Blakemore* [32.53].

It should be noted that the energy gap associated with InN has been the subject of some controversy since 2002. The pioneering experimental results of *Tansley* and *Foley* [32.54], reported in 1986, suggested that InN has an energy gap of 1.89 eV. This value has been used extensively in Monte Carlo simulations of the electron transport within this material since that time [32.23, 28, 29, 31]; typically, the influence of the energy gap on the electron transport occurs through its impact on the nonparabolicity coefficient α. In 2002,

Davydov et al. [32.55], *Wu* et al. [32.56], and *Matsuoka* et al. [32.57], presented experimental evidence which instead suggests a considerably smaller energy gap for InN, around 0.7 eV. For the purposes of this analysis, the revised value for the InN energy gap is employed; the sensitivity of the velocity–field characteristic associated with bulk wurtzite GaN to variations in the nonparabolicity coefficient α has been explored, in detail, by *O'Leary* et al. [32.32].

The band structure associated with bulk wurtzite GaN has also been the focus of some controversy. In particular, *Brazel* et al. [32.58] employed ballistic electron emission microscopy measurements in order to demonstrate that the first upper conduction band valley occurs only 340 meV above the lowest point in the conduction band for this material. This contrasts rather dramatically with more traditional results, such as the calculation of *Lambrecht* and *Segall* [32.51], which instead suggest that the first upper conduction band valley minimum within wurtzite GaN occurs about 2 eV above the lowest point in the conduction band. Clearly, this will have a significant impact upon the results. While the results of *Brazel* et al. [32.58] were reported in 1997, electron transport simulations adopted the more traditional intervalley energy separation of about 2 eV until relatively recently. Accordingly, we have adopted the more traditional intervalley energy separation for the purposes of our present analysis. The sensitivity of the velocity–field characteristic associated with bulk wurtzite GaN to variations in the intervalley energy separation has been explored, in detail, by *O'Leary* et al. [32.32].

32.2 Steady-State and Transient Electron Transport Within Bulk Wurtzite GaN, AlN, and InN

The current interest in the III–V nitride semiconductors, GaN, AlN, and InN, is primarily being fuelled by the tremendous potential of these materials for novel electronic and optoelectronic device applications. With the recognition that informed electronic and optoelectronic device design requires a firm understanding of the nature of the electron transport within these materials, electron transport within the III–V nitride semiconductors has been the focus of intensive investigation over the years. The literature abounds with studies on steady-state and transient electron transport within these materials [32.13–34, 38, 39, 41, 42, 48]. As a result of this intense flurry of research activity, novel III–V nitride semiconductor-based devices are starting to be deployed in today's commercial products. Future developments in the III–V nitride semiconductor field will undoubtedly require an even deeper understanding of the electron transport mechanisms within these materials.

In the previous section, we presented details of the Monte Carlo simulation approach that we employ for the analysis of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN. In this section, an overview of the steady-state and transient electron transport results we obtained from these Monte Carlo simulations is provided. In the first part of this section, we focus upon bulk wurtzite GaN. In particular, the velocity–field characteristic associated with this material will be examined in detail. Then, an overview of our steady-state electron transport results, corresponding to the three III–V nitride semiconductors under consideration in this analysis, i. e., GaN, AlN, and InN, will be given, and a comparison with the more conventional III–V compound semiconductor, GaAs, will be presented. A comparison between the temperature dependence of the velocity–field characteristics associated with GaN and GaAs will then be provided, and our Monte Carlo results will be used to account for the differences in behavior. A similar analysis will be presented for the doping dependence. Next, detailed simulation results for AlN and InN will be presented. Finally, the transient electron transport that occurs within the III–V nitride semiconductors, GaN, AlN, and InN, is determined and compared with that in GaAs.

This section is organized in the following manner. In Sect. 32.2.1, the velocity–field characteristic associated with bulk wurtzite GaN is presented and analyzed. Then, in Sect. 32.2.2, the velocity-field characteristics associated with the III–V nitride semiconductors under consideration in this analysis will be compared and contrasted with that of GaAs. The sensitivity of the velocity–field characteristic associated with bulk wurtzite GaN to variations in the crystal temperature will then be examined in Sect. 32.2.3, and a comparison with that corresponding to GaAs presented. In Sect. 32.2.4, the sensitivity of the velocity–field characteristic associated with bulk wurtzite GaN to variations in the doping concentration level will be explored, and a comparison with that corresponding to GaAs presented. The velocity–field characteristics associated with AlN and InN will then be examined in

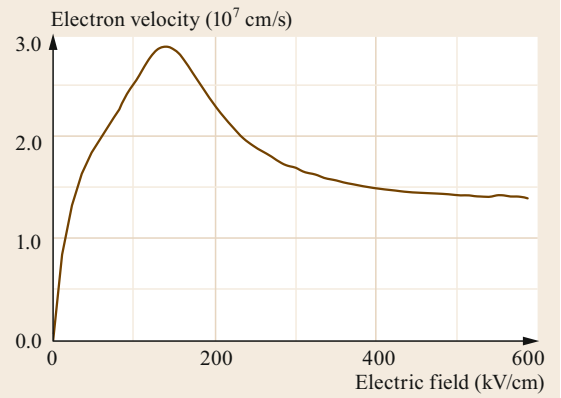


Fig. 32.1 The velocity–field characteristic associated with bulk wurtzite GaN. Like many other compound semiconductors, the electron drift velocity reaches a peak, and at higher applied electric fields it decreases until it saturates

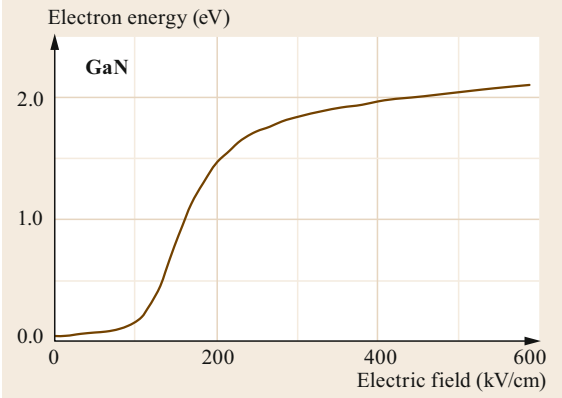


Fig. 32.2 The average electron energy as a function of the applied electric field for bulk wurtzite GaN. Initially, the average electron energy remains low, only slightly higher than the thermal energy, $\frac{3}{2}k_{\mathrm{B}}T$. At 100 kV/cm, however, the average electron energy increases dramatically. This increase is due to the fact that the polar optical phonon scattering mechanism can no longer absorb all of the energy gained from the applied electric field

Sect. 32.2.5 and Sect. 32.2.6, respectively. Our transient electron transport analysis results are then presented in Sect. 32.2.7. Finally, the conclusions of this electron transport analysis are summarized in Sect. 32.2.8.

32.2.1 Steady-State Electron Transport Within Bulk Wurtzite GaN

Our examination of results begins with GaN, the most commonly studied III–V nitride semiconductor. The velocity–field characteristic associated with this material is presented in Fig. 32.1. This result was obtained through our Monte Carlo simulations of the electron transport within this material for the bulk wurtzite GaN parameter selections specified in Table 32.1 and Table 32.2; the crystal temperature was set to 300 K and the doping concentration to $10^{17}\,\mathrm{cm}^{-3}$. We see that for applied electric fields in excess of 140 kV/cm, the electron drift velocity decreases, eventually saturating at 1.4×10^7 cm/s for high applied electric fields. By examining the results of our Monte Carlo simulation further, an understanding of this result becomes clear.

First, we discuss the results at low applied electric fields, i. e., applied electric fields of less than 30 kV/cm. This is referred to as the linear regime of electron transport as the electron drift velocity is well characterized by the low-field electron drift mobility μ in this regime, i. e., a linear low-field electron drift velocity dependence on the applied electric field, i. e., $v_{\mathrm{d}} = \mu \boldsymbol{E}$, applies in this regime. Examining the distribution function for this regime, we find that it is very similar to the zero-field distribution function with a slight shift in the direction opposite to the applied electric field. In this regime, the average electron energy remains relatively low, with most of the energy gained from the applied electric field being transferred into the lattice through polar optical phonon scattering.

If we examine the average electron energy as a function of the applied electric field, shown in Fig. 32.2, we see that there is a sudden increase at around 100 kV/cm. In order to understand why this increase occurs, we note that the dominant energy loss mechanism for many of the III–V compound semiconductors, including GaN, is polar optical phonon scattering. When the applied electric field is less than 100 kV/cm, all of the energy that the electrons gain from the applied electric field is lost through polar optical phonon scattering. The other scattering mechanisms, i. e., ionized impurity scattering, piezoelectric scattering, and acoustic deformation potential scattering, do not remove energy from the electron ensemble: they are elastic scattering mechanisms. However, beyond a certain critical applied electric field strength, the polar optical phonon scattering mechanism can no longer remove all of the energy gained from the applied electric field. Other scattering mechanisms must start to play a role if the electron ensemble is to remain in equilibrium. The average electron energy increases until intervalley scattering begins and an energy balance is re-established.

As the applied electric field is increased beyond 100 kV/cm, the average electron energy increases un-

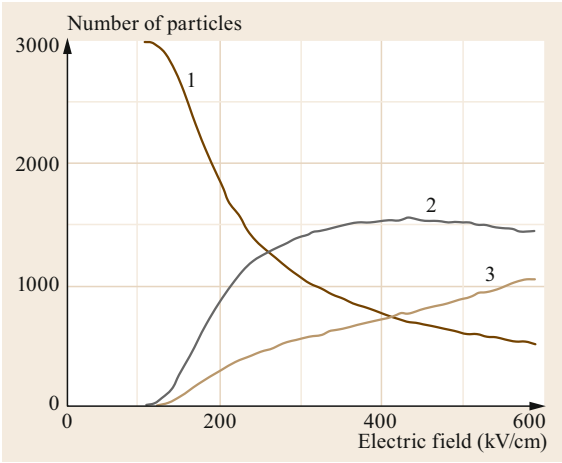


Fig. 32.3 The valley occupancy as a function of the applied electric field for the case of bulk wurtzite GaN. Soon after the average electron energy increases, electrons begin to transfer to the upper valleys of the conduction band. Three thousand electrons were employed for this simulation. The valleys are labeled 1, 2, and 3, in accordance with their energy minima; the lowest energy valley is valley 1, the next higher energy valley is valley 2, and the highest energy valley is valley 3

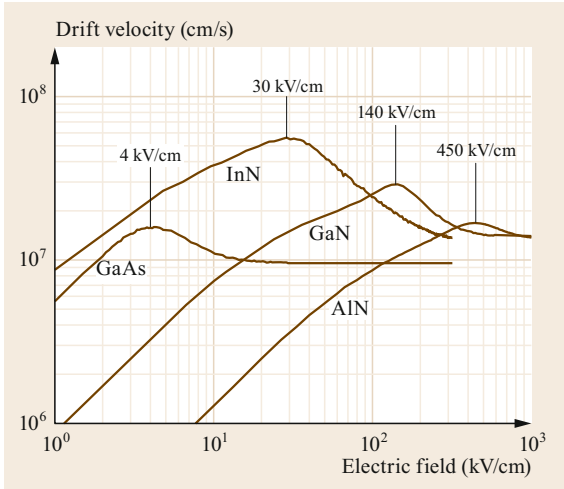


Fig. 32.4 A comparison of the velocity–field characteristics associated with the III–V nitride semiconductors, GaN, AlN, and InN, with that associated with GaAs. (After [32.29], with permission from AIP)

til a substantial fraction of the electrons have acquired enough energy in order to transfer into the upper valleys. As the effective mass of the electrons in the upper valleys is greater than that in the lowest valley, the electrons in the upper valleys will be slower. As more electrons transfer to the upper valleys (Fig. 32.3), the electron drift velocity decreases. This accounts for the negative differential mobility observed in the velocity–field characteristic depicted in Fig. 32.1.

Finally, at high applied electric fields, the number of electrons in each valley saturates. It can be shown that in the high-field limit the number of electrons in each valley is proportional to the product of the density of states of that particular valley and the corresponding valley degeneracy. At this point, the electron drift velocity stops decreasing and achieves saturation.

Thus far, electron transport results corresponding to bulk wurtzite GaN have been presented and discussed qualitatively. It should be noted, however, that the same phenomenon that occurs in the velocity–field characteristic associated with GaN also occurs for the other III–V nitride semiconductors, AlN and InN. The importance of polar optical phonon scattering when determining the nature of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, will become even more apparent later, as it will be used to account for much of the electron transport behavior within these materials.

32.2.2 Steady-State Electron Transport: A Comparison of the III–V Nitride Semiconductors with GaAs

Setting the crystal temperature to 300 K and the level of doping to $10^{17}\,\mathrm{cm}^{-3}$, the velocity–field characteristics associated with the III–V nitride semiconductors under consideration in this analysis – GaN, AlN, and InN – are contrasted with that of GaAs in Fig. 32.4. We see that each of these III–V compound semiconductors achieves a peak in its velocity–field characteristic. InN achieves the highest steady-state peak electron drift velocity, $5.6\times10^7\,\mathrm{cm/s}$ at an applied electric field of 30 kV/cm. This contrasts with the case of GaN, $2.9\times10^7\,\mathrm{cm/s}$ at 140 kV/cm, and that of AlN, $1.7\times10^7\,\mathrm{cm/s}$ at 450 kV/cm. For GaAs, the peak electron drift velocity of $1.6\times10^7\,\mathrm{cm/s}$ occurs at a much lower applied electric field than that for the III–V nitride semiconductors (only 4 kV/cm).

32.2.3 Influence of Temperature on the Electron Drift Velocities Within GaN and GaAs

The temperature dependence of the velocity–field characteristic associated with bulk wurtzite GaN is now examined. Figure 32.5a shows how the velocity–field characteristic associated with bulk wurtzite GaN varies as the crystal temperature is increased from 100 to 700 K, in increments of 200 K. The upper limit, 700 K, is chosen as it is the highest operating temperature that may be expected for AlGaN/GaN power devices. To

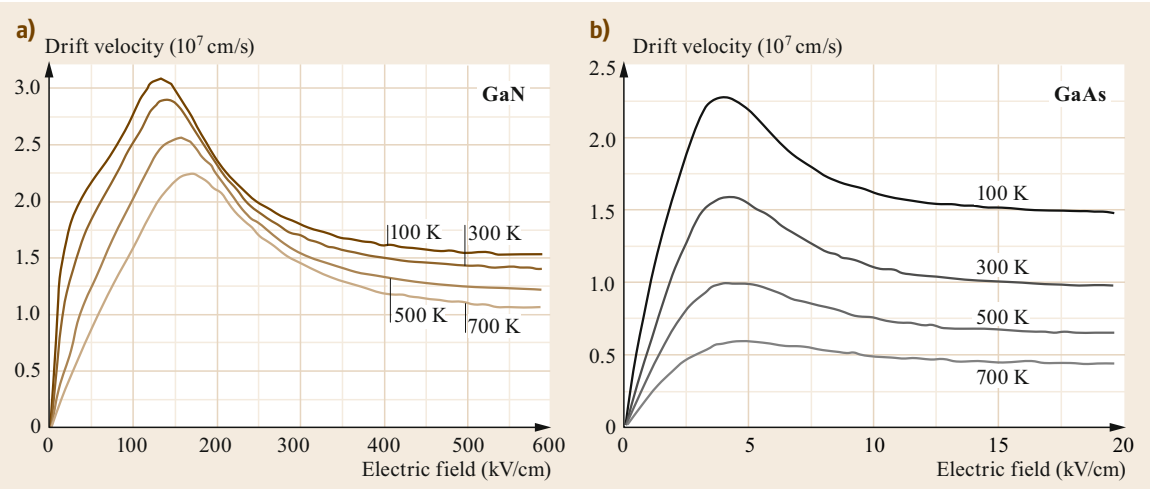


Fig. 32.5a,b A comparison of the temperature dependence of the velocity–field characteristics associated with **(a)** GaN and **(b)** GaAs. GaN maintains a higher electron drift velocity with increased temperatures than GaAs does

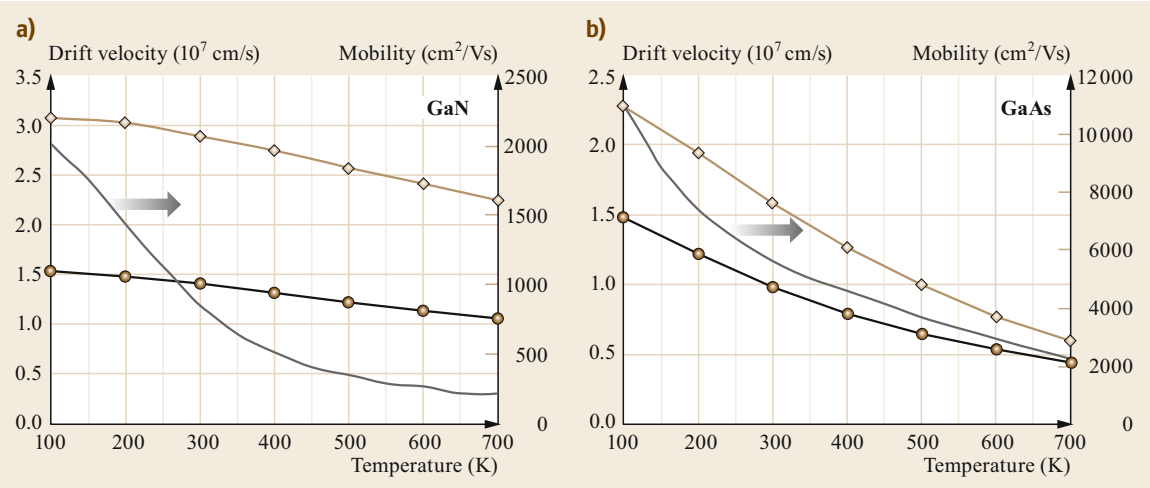


Fig. 32.6a,b A comparison of the temperature dependence of the low-field electron drift mobility (*solid lines*), the peak electron drift velocity (*diamonds*), and the saturation electron drift velocity (*solid points*) for **(a)** GaN and **(b)** GaAs. The low-field electron drift mobility of GaN drops quickly with increasing temperature, but its peak and saturation electron drift velocities are less sensitive to increases in temperature than GaAs

highlight the difference between the III–V nitride semiconductors with more conventional III–V compound semiconductors, such as GaAs, Monte Carlo simulations of the electron transport within GaAs have also been performed under the same conditions as GaN. Figure 32.5b shows the results of these simulations. Note that the electron drift velocity for the case of GaN is much less sensitive to changes in temperature than that associated with GaAs.

To quantify this dependence further, the low-field electron drift mobility, the peak electron drift velocity, and the saturation electron drift velocity are plotted as a function of the crystal temperature in Fig. 32.6, these results being determined from our Monte Carlo simulations of the electron transport within these materials. For both GaN and GaAs, it is found that all of these electron transport metrics diminish as the crystal temperature is increased. As may be seen through an inspection of Fig. 32.5, the peak and saturation electron drift velocities do not drop as much in GaN as they do in GaAs in response to increases in the crystal temperature. The low-field electron drift mobility in GaN, however, is seen to fall quite rapidly with temperature, this drop being particularly severe for tem-

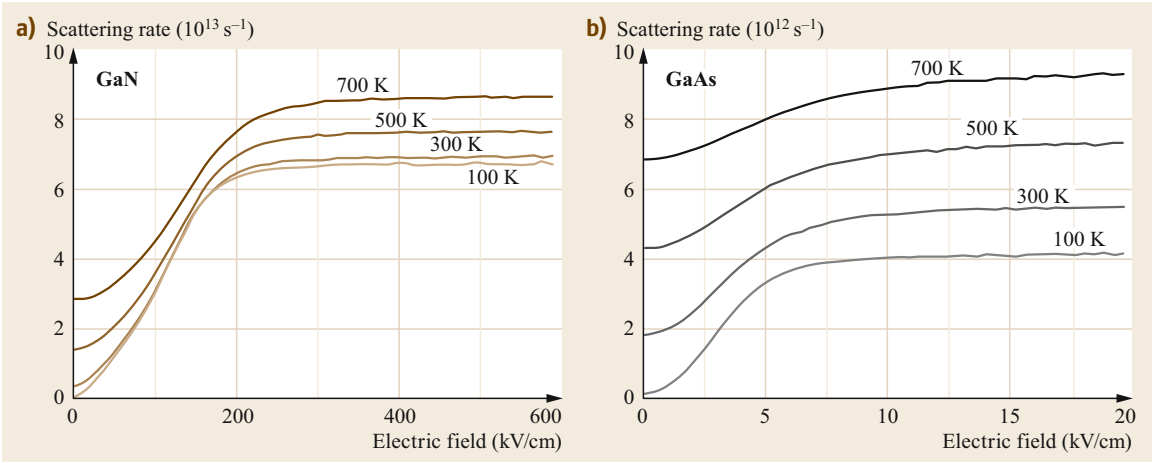


Fig. 32.7a,b A comparison of the polar optical phonon scattering rates as a function of the applied electric field strength for various crystal temperatures for **(a)** GaN and **(b)** GaAs. Polar optical phonon scattering is seen to increase much more quickly with temperature in GaAs

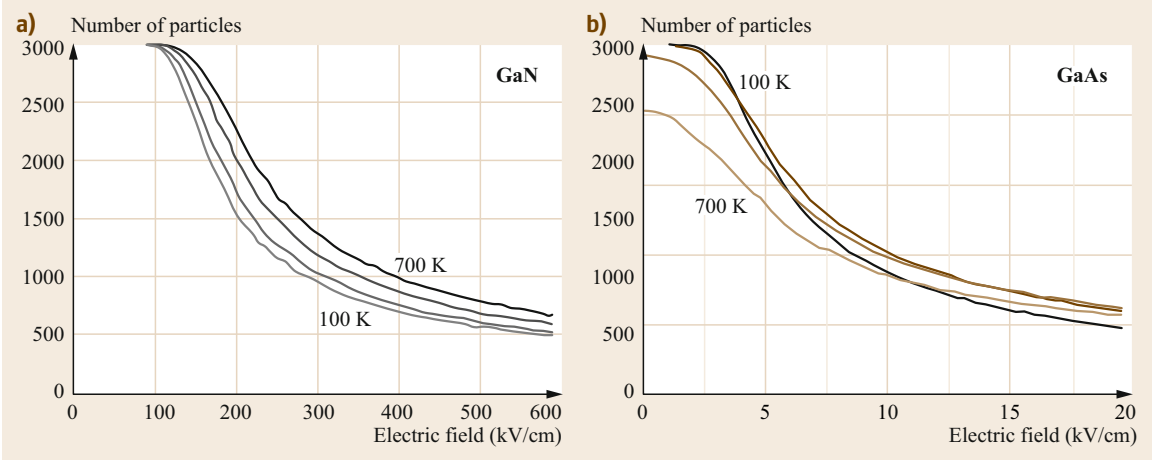


Fig. 32.8a,b A comparison of the number of particles in the lowest energy valley of the conduction band, the Γ valley, as a function of the applied electric field for various crystal temperatures, for the cases of **(a)** GaN and **(b)** GaAs. In GaAs, the electrons begin to occupy the upper valleys much more quickly, causing the electron drift velocity to drop as the crystal temperature is increased. Three thousand electrons were employed for these steady-state electron transport simulations

peratures at and below room temperature. This property will likely have an impact on high-power device performance.

Delving deeper into our Monte Carlo results yields clues into the reason for this variation in temperature dependence. First, we examine the polar optical phonon scattering rate as a function of the applied electric field strength. Figure 32.7 shows that the scattering rate only increases slightly with temperature for the case of GaN, from $6.7\times10^{13}\,\mathrm{s}^{-1}$ at 100 K to $8.6\times10^{13}\,\mathrm{s}^{-1}$ at 700 K, for high applied electric field strengths. Contrast this with the case of GaAs, where the rate increases from $4.0\times10^{12}\,\mathrm{s}^{-1}$ at 100 K to more than twice that amount at 700 K, $9.2\times10^{12}\,\mathrm{s}^{-1}$, at high applied electric field strengths. This large increase in the polar optical phonon scattering rate for the case of GaAs is one reason for the large drop in the electron drift velocity with increasing temperature for the case of GaAs.

A second reason for the variation in temperature dependence of the two materials is the occupancy of the upper valleys, shown in Fig. 32.8. In the case of GaN, the upper valleys begin to become occupied at roughly

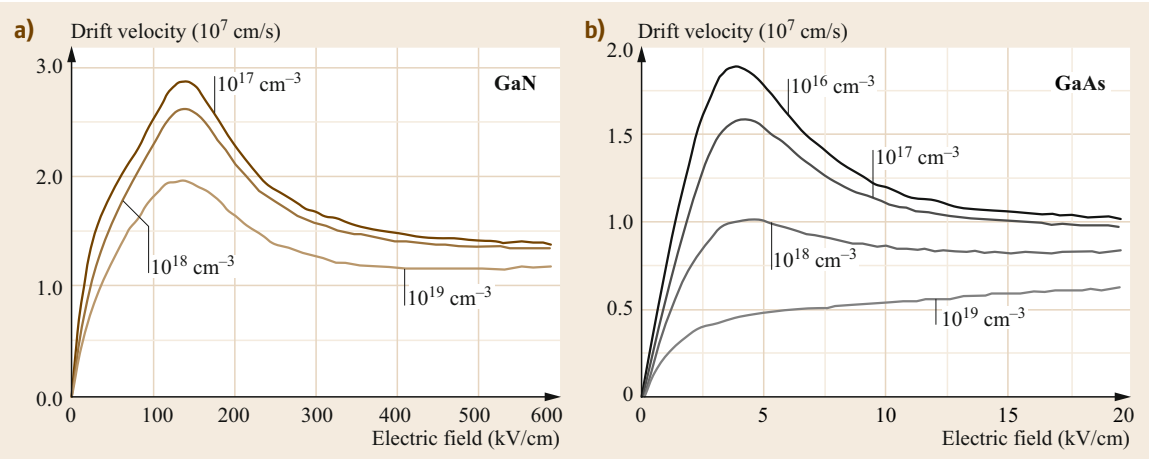


Fig. 32.9a,b A comparison of the dependence of the velocity–field characteristics associated with **(a)** GaN and **(b)** GaAs on the doping concentration. GaN maintains a higher electron drift velocity with increased doping levels than GaAs does

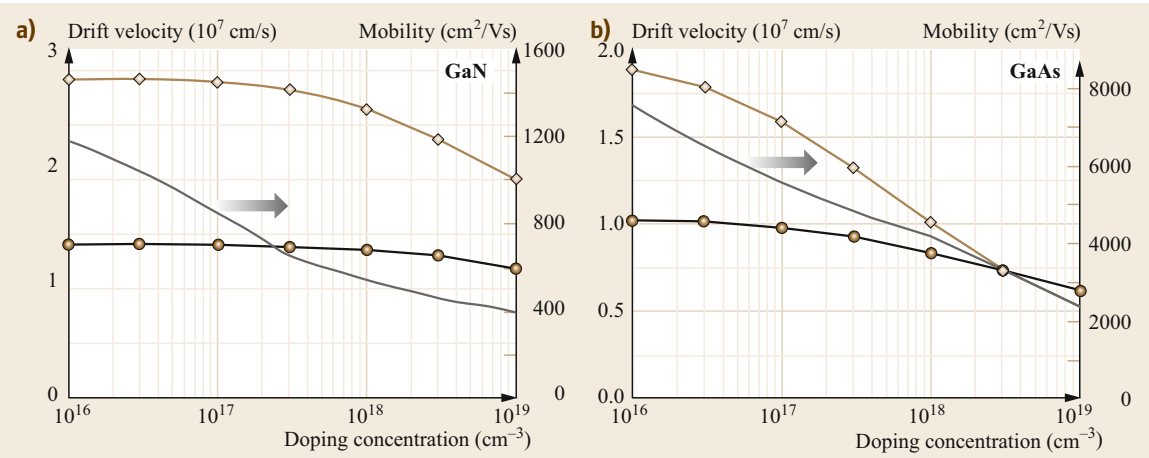


Fig. 32.10a,b A comparison of the low-field electron drift mobility (*solid lines*), the peak electron drift velocity (*diamonds*), and the saturation electron drift velocity (*solid points*) for **(a)** GaN and **(b)** GaAs as a function of the doping concentration. These parameters are more insensitive to increases in doping in GaN than in GaAs

the same applied electric field strength, 100 kV/cm, independent of temperature. For the case of GaAs, however, the upper valleys are at a much lower energy than those in GaN. In particular, while the first upper conduction band valley minimum is 1.9 eV above the lowest point in the conduction band in GaN, the first upper conduction band valley is only 290 meV above the bottom of the conduction band in GaAs [32.53]. As the upper conduction band valleys are so close to the bottom of the conduction band for the case of GaAs, the thermal energy (at 700 K, $k_B T \cong 60$ meV) is enough in order to allow for a small fraction of the electrons to transfer into the upper valleys even before an electric field is applied. When electrons occupy the upper valleys, intervalley scattering and the upper valleys' larger effective masses reduce the overall electron drift velocity. This is another reason why the velocity–field characteristic associated with GaAs is more sensitive to variations in crystal temperature than that associated with GaN.

32.2.4 Influence of Doping on the Electron Drift Velocities Within GaN and GaAs

One parameter that can be readily controlled during the fabrication of semiconductor devices is the doping concentration. Understanding the effect of doping on the resultant electron transport is also important.

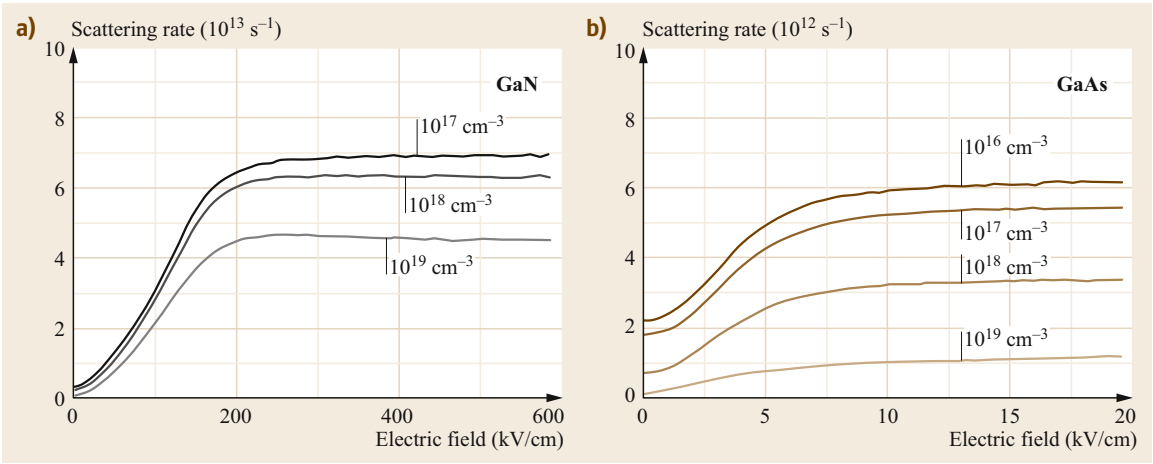


Fig. 32.11a,b A comparison of the polar optical phonon scattering rates as a function of the applied electric field, for both **(a)** GaN and **(b)** GaAs, for various doping concentrations

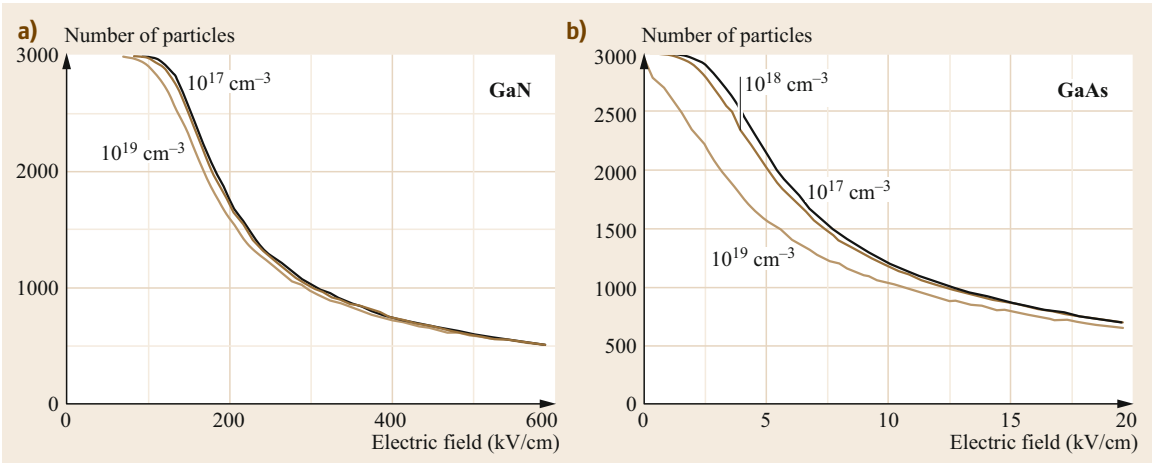


Fig. 32.12a,b A comparison of the number of particles in the lowest valley of the conduction band, the Γ valley, as a function of the applied electric field, for both **(a)** GaN and **(b)** GaAs, for various doping concentration levels. Three thousand electrons were employed for these steady-state electron transport simulations

In Fig. 32.9, the velocity–field characteristic associated with GaN is presented for a number of different doping concentration levels. Once again, three important electron transport metrics are influenced by the doping concentration level: the low-field electron drift mobility, the peak electron drift velocity, and the saturation electron drift velocity; see Fig. 32.10. Our simulation results suggest that for doping concentrations of less than $10^{17}\,\mathrm{cm}^{-3}$, there is very little effect on the velocity–field characteristic for the case of GaN. However, for doping concentrations above $10^{17}\,\mathrm{cm}^{-3}$, the peak electron drift velocity diminishes considerably, from $2.9\times10^{7}\,\mathrm{cm/s}$ for the case of $10^{17}\,\mathrm{cm}^{-3}$ doping to $2.0\times10^{7}\,\mathrm{cm/s}$ for the case of $10^{19}\,\mathrm{cm}^{-3}$ doping. The saturation electron drift velocity within GaN is found to only decrease slightly in response to increases in the doping concentration. The effect of doping on the low-field electron drift mobility is also shown. It is seen that this mobility drops significantly in response to increases in the doping concentration level, from $1200\,\mathrm{cm}^2/(\mathrm{V\,s})$ at $10^{16}\,\mathrm{cm}^{-3}$ doping to $400\,\mathrm{cm}^2/(\mathrm{V\,s})$ at $10^{19}\,\mathrm{cm}^{-3}$ doping.

As we did for temperature, we compare the sensitivity of the velocity-field characteristic associated with GaN to doping with that associated with GaAs. Figure 32.10 shows this comparison. For the case of GaAs, it is seen that the electron drift velocities decrease much more with increased doping than those associated with GaN. In particular, for the case of GaAs, the peak electron drift velocity decreases from

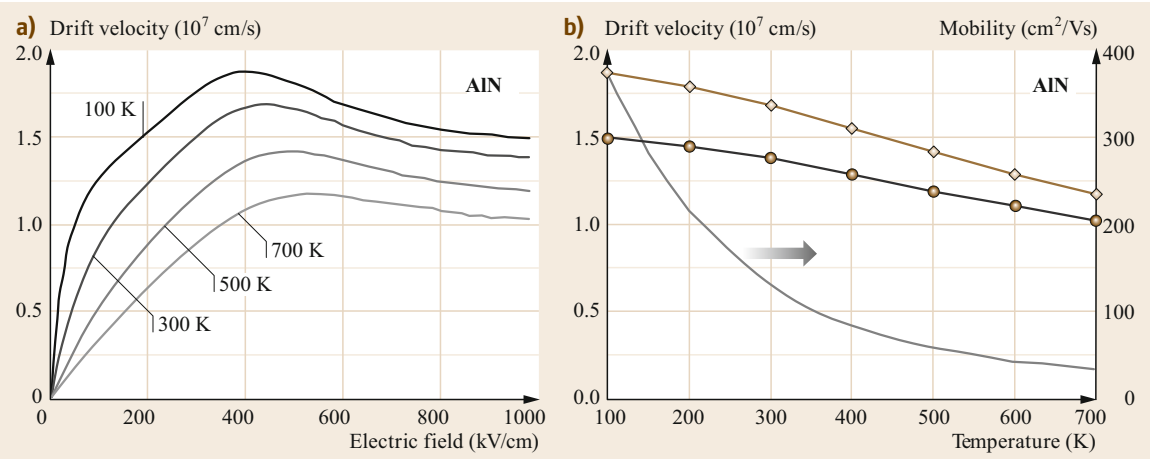


Fig. 32.13 The velocity–field characteristic associated with AlN **(a)** for various crystal temperatures. The trends in the low-field mobility (*solid line*), the peak electron drift velocity (*diamonds*), and the saturation electron drift velocity (*solid points*), are also shown. AlN exhibits its peak electron drift velocity at very high applied electric fields. AlN has the lowest peak electron drift velocity and the lowest low-field electron drift mobility of the III–V nitride semiconductors considered in this analysis **(b)**

1.8×10^7 cm/s at 10^{16} cm^{-3} doping to 0.6×10^7 cm/s at 10^{19} cm^{-3} doping. For GaAs, at the higher doping levels, the peak in the velocity-field characteristic disappears completely for sufficiently high doping concentrations. The saturation electron drift velocity decreases from 1.0×10^7 cm/s at 10^{16} cm^{-3} doping to 0.6×10^7 cm/s at 10^{19} cm^{-3} doping. The low-field electron drift mobility also diminishes dramatically with increased doping, dropping from 7800 cm^2/(V s) at 10^{16} cm^{-3} doping to 2200 cm^2/(V s) at 10^{19} cm^{-3} doping; see Fig. 32.10.

Once again, it is interesting to determine why the doping dependence in GaAs is so much more pronounced than it is in GaN. Again, we examine the polar optical phonon scattering rate and the occupancy of the upper valleys. Figure 32.11 shows the polar optical phonon scattering rates as a function of the applied electric field, for both GaN and GaAs. In this case, however, due to screening effects, the rate drops when the doping concentration is increased. The decrease, however, is much more pronounced for the case of GaAs than for GaN. It is believed that this drop in the polar optical phonon scattering rate allows for upper valley occupancy to occur more quickly in GaAs rather than in GaN (Fig. 32.12). For GaN, electrons begin to occupy the upper valleys at roughly the same applied electric field strength, independent of the doping level. However, for the case of GaAs, the upper valleys are occupied more quickly with greater doping. When the upper valleys are occupied, the electron drift velocity decreases due to intervalley scattering and the larger effective mass of the electrons within the upper valleys.

32.2.5 Electron Transport in AlN

AlN has the largest effective mass of the III–V nitride semiconductors considered in this analysis. Accordingly, it is not surprising that this material exhibits the lowest electron drift velocity and the lowest low-field electron drift mobility. The sensitivity of the velocity–field characteristic associated with AlN to variations in the crystal temperature may be examined by considering Fig. 32.13. As with the case of GaN, the velocity–field characteristic associated with AlN is extremely robust to variations in the crystal temperature. In particular, its peak electron drift velocity, which is 1.8×10^7 cm/s at 100 K, only decreases to 1.2×10^7 cm/s at 700 K. Similarly, its saturation electron drift velocity, which is 1.5×10^7 cm/s at 100 K, only decreases to 1.0×10^7 cm/s at 700 K. The low-field electron drift mobility associated with AlN also diminishes in response to increases in the crystal temperature, from 375 cm^2/(V s) at 100 K to 40 cm^2/(V s) at 700 K.

The sensitivity of the velocity–field characteristic associated with AlN to variations in the doping concentration may be examined by considering Fig. 32.14. It is noted that the variations in the velocity–field characteristic associated with AlN in response to variations in the doping concentration are not as pronounced as those which occur in response to variations in the crystal temperature. Quantitatively, the peak electron drift velocity drops from 1.7×10^7 cm/s at 10^{17} cm^{-3} doping to 1.3×10^7 cm/s at 10^{19} cm^{-3} doping. Similarly, its saturation electron drift velocity drops from 1.4×10^7 cm/s at 10^{17} cm^{-3} doping to 1.2×10^7 cm/s

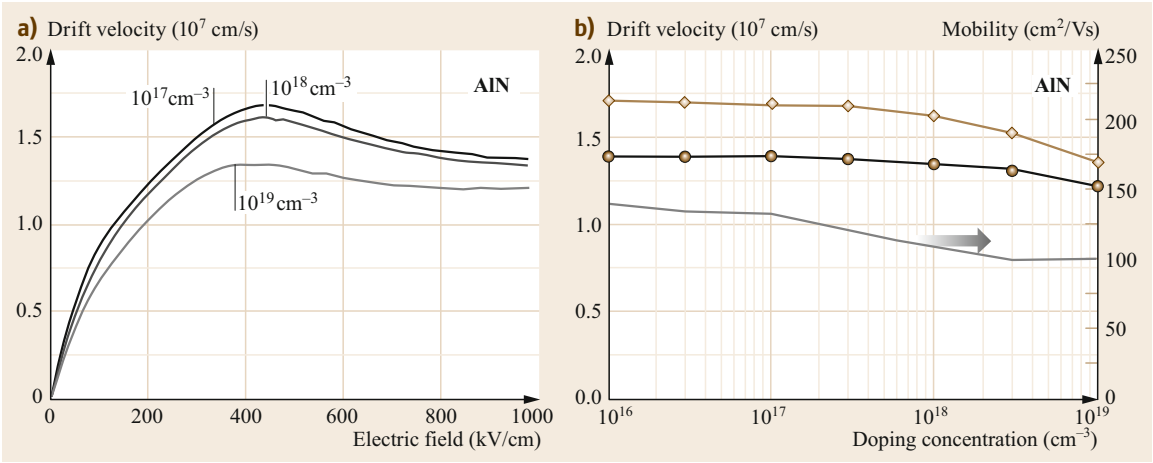


Fig. 32.14 The velocity–field characteristic associated with AlN for various doping concentrations **(a)**. The trends in the low-field electron drift mobility (*solid line*), the peak electron drift velocity (*diamonds*), and the saturation electron drift velocity (*solid points*), are also shown **(b)**

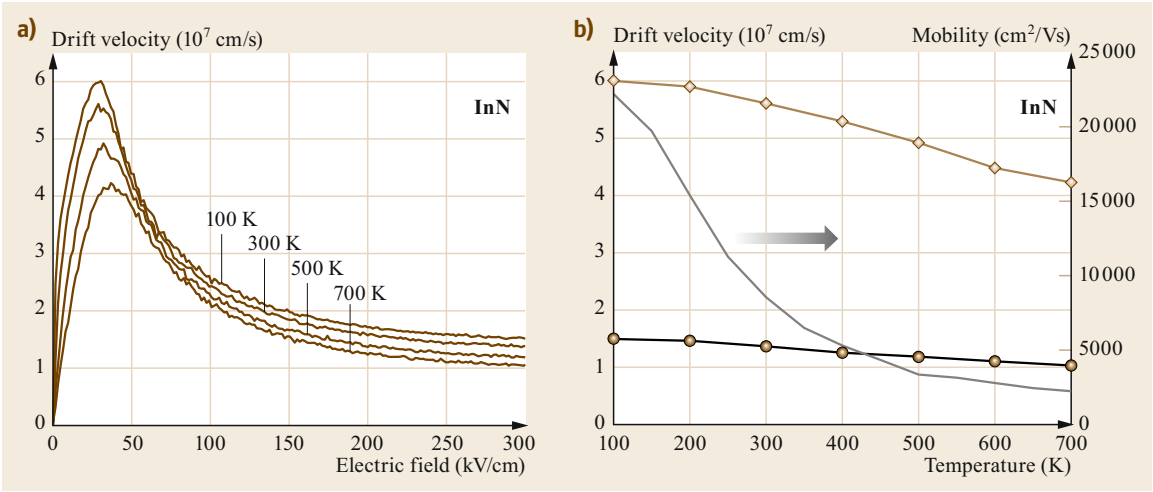


Fig. 32.15 The velocity–field characteristic associated with InN for various crystal temperatures **(a)**. The trends in the low-field electron drift mobility (*solid line*), the peak electron drift velocity (*diamonds*), and the saturation electron drift velocity (*solid points*), are also shown **(b)**. InN has the highest peak electron drift velocity and the highest low-field electron drift mobility of the III–V nitride semiconductors considered in this analysis

at $10^{19}\,\mathrm{cm}^{-3}$ doping. The influence of doping on the low-field electron drift mobility associated with AlN is also observed to be not as pronounced as for the case of crystal temperature. Figure 32.14b shows that the low-field electron drift mobility associated with AlN decreases from $140\,\mathrm{cm}^2/(\mathrm{V\,s})$ at $10^{16}\,\mathrm{cm}^{-3}$ doping to $100\,\mathrm{cm}^2/(\mathrm{V\,s})$ at $10^{19}\,\mathrm{cm}^{-3}$ doping.

32.2.6 Electron Transport in InN

InN has the smallest effective mass of the three III–V nitride semiconductors considered in this analysis. Accordingly, it is not surprising that it exhibits the highest electron drift velocity and the highest low-field electron drift mobility. The sensitivity of the velocity-field characteristic associated with InN to variations in the crystal temperature may be examined by considering Fig. 32.15. As with the cases of GaN and AlN, the velocity–field characteristic associated with InN is extremely robust to increases in the crystal temperature. In particular, its peak electron drift velocity, which is 6.0×10^7 cm/s at 100 K, only decreases to 4.2×10^7 cm/s at 700 K. Similarly, its saturation electron drift velocity, which is 1.5×10^7 cm/s at 100 K, only decreases to

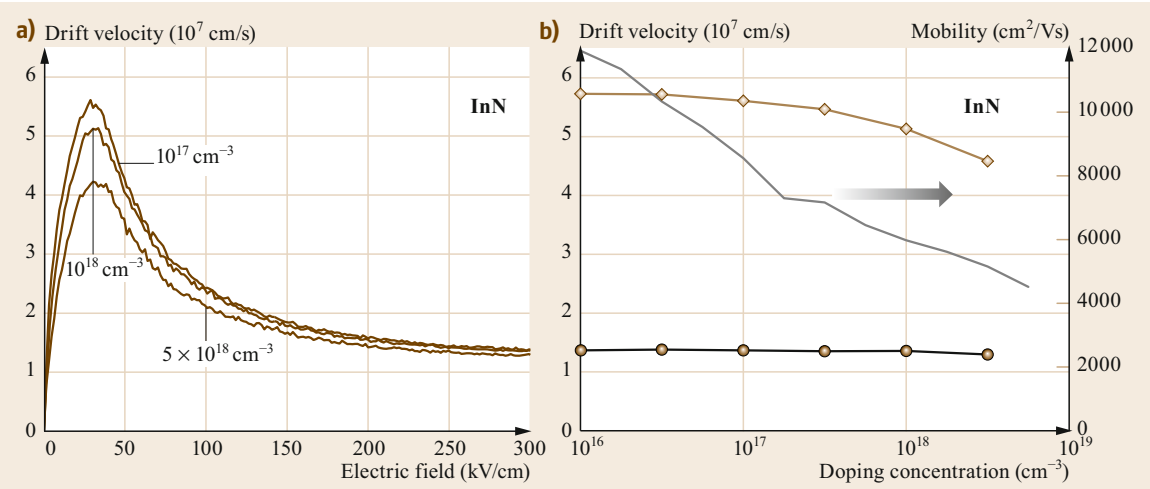


Fig. 32.16 The velocity–field characteristic associated with InN for various doping concentrations **(a)**. The trends in the low-field electron drift mobility (*solid line*), the peak electron drift velocity (*diamonds*), and the saturation electron drift velocity (*solid points*), are also shown **(b)**

1.0×10^7 cm/s at 700 K. The low-field electron drift mobility associated with InN also diminishes in response to increases in the crystal temperature, from about 22 000 cm^2/(V s) at 100 K to below 2300 cm^2/(V s) at 700 K.

The sensitivity of the velocity–field characteristic associated with InN to variations in the doping concentration may be examined by considering Fig. 32.16. These results suggest a similar robustness to the doping concentration for the case of InN. In particular, it is noted that for doping concentrations below 10^{17} cm^{-3}, the velocity–field characteristic associated with InN exhibits very little dependence on the doping concentration. When the doping concentration is increased above 10^{17} cm^{-3}, however, the peak electron drift velocity diminishes. Quantitatively, the peak electron drift velocity decreases from 5.6×10^7 cm/s at 10^{17} cm^{-3} doping to 4.2×10^7 cm/s at 5×10^{18} cm^{-3} doping; a numerical instability occurs for the case of InN at large doping concentrations, owing to its small electron effective mass, and this prevented the determination of a 10^{19} cm^{-3} doping concentration result. The saturation electron drift velocity only drops slightly, however, from 1.4×10^7 cm/s at 10^{17} cm^{-3} doping to 1.3×10^7 cm/s at 5×10^{18} cm^{-3} doping. The low-field electron drift mobility, however, drops significantly with doping, from 12 000 cm^2/(V s) at 10^{16} cm^{-3} doping to 4900 cm^2/(V s) at 5×10^{18} cm^{-3} doping.

32.2.7 Transient Electron Transport

Steady-state electron transport is the dominant electron transport mechanism in devices with larger dimensions. For devices with smaller dimensions, however, transient electron transport must also be considered when evaluating device performance. *Ruch* [32.59] demonstrated, for both silicon and GaAs, that the transient electron drift velocity may exceed the corresponding steady-state electron drift velocity by a considerable margin for appropriate selections of the applied electric field. *Shur* and *Eastman* [32.60] explored the device implications of transient electron transport, and demonstrated that substantial improvements in the device performance can be achieved as a consequence. *Heiblum* et al. [32.61] made the first direct experimental observation of transient electron transport within GaAs. Since then there have been a number of experimental investigations into the transient electron transport within the III–V compound semiconductors [32.62–64].

Thus far, very little research has been invested into the study of transient electron transport within the III–V nitride semiconductors, GaN, AlN, and InN. *Foutz* et al. [32.21] examined transient electron transport within both the wurtzite and zinc blende phases of GaN. In particular, they examined how electrons, initially in thermal equilibrium, respond to the sudden application of a constant electric field. In devices with dimensions greater than 0.2 μm, they found that steady-state electron transport is expected to dominate device performance. For devices with smaller dimensions, however, upon the application of a sufficiently high electric field, they found that the transient electron drift velocity can considerably overshoot the corresponding steady-state electron drift velocity. This velocity overshoot was found to be comparable with that which occurs within GaAs.

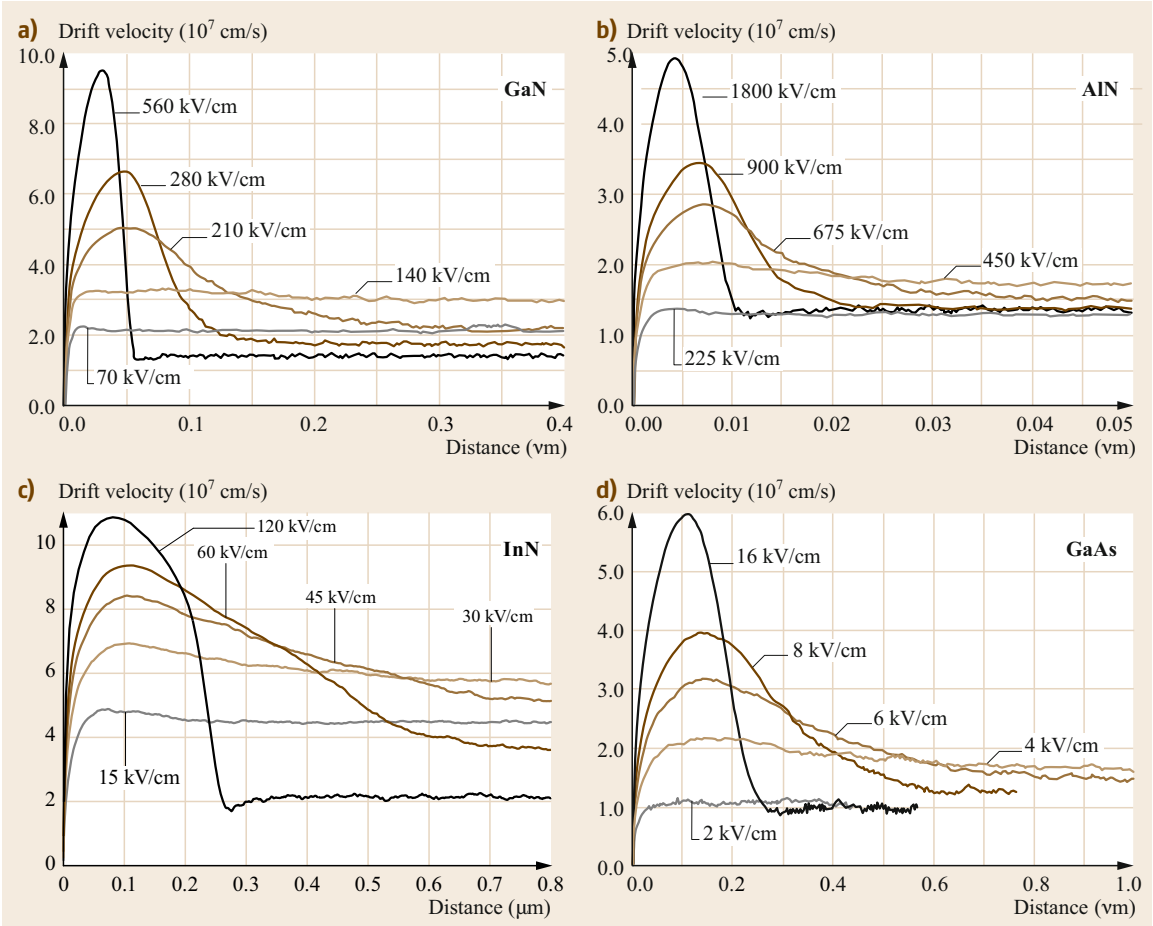


Fig. 32.17a–d The electron drift velocity as a function of the distance displaced since the application of the electric field for various applied electric field strengths, for the cases of **(a)** GaN, **(b)** AlN, **(c)** InN, and **(d)** GaAs. In all cases, we have assumed an initial zero field electron distribution, a crystal temperature of 300 K, and a doping concentration of 10^{17} cm^{-3}. (After [32.29] with permission, copyright AIP)

Foutz et al. [32.29] performed a subsequent analysis in which the transient electron transport within all of the III–V nitride semiconductors under consideration in this analysis were compared with that which occurs within GaAs. In particular, following the approach of *Foutz* et al. [32.21], they examined how electrons, initially in thermal equilibrium, respond to the sudden application of a constant electric field. A key result of this study, presented in Fig. 32.17, plots the transient electron drift velocity as a function of the distance displaced since the electric field was initially applied for a number of applied electric field strengths and for each of the materials considered in this analysis. The InN results have been updated owing to the revisions in the material parameter selections corresponding to this particular material; see Figure 32.17c.

Focusing initially on the case of GaN (Fig. 32.17a), we note that the electron drift velocity for the applied electric field strengths 70 kV/cm and 140 kV/cm reaches steady-state very quickly, with little or no velocity overshoot. In contrast, for applied electric field strengths above 140 kV/cm, significant velocity overshoot occurs. This result suggests that in GaN, 140 kV/cm is a critical field for the onset of velocity overshoot effects. As mentioned in Sect. 32.2.2, 140 kV/cm also corresponds to the peak in the velocity-field characteristic associated with GaN; recall Fig. 32.4. Steady-state Monte Carlo simulations suggest that this is the point at which significant upper valley occupation begins to occur; recall Fig. 32.3. This suggests that velocity overshoot is related to the transfer of electrons to the upper valleys. Similar results are

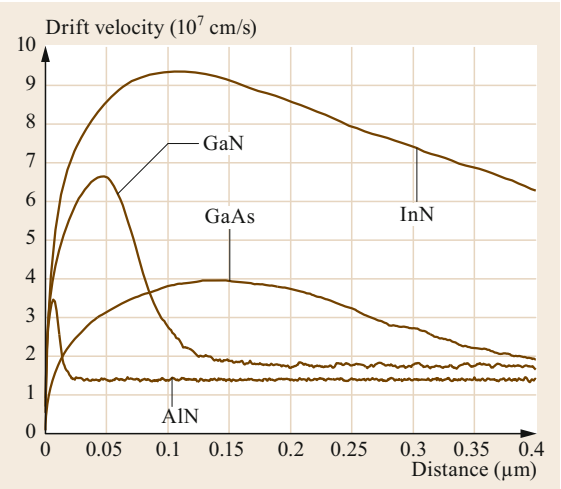


Fig. 32.18 A comparison of the velocity overshoot amongst the III–V nitride semiconductors and GaAs. The applied electric field strength chosen corresponds to twice the critical applied electric field strength at which the peak in the steady-state velocity–field characteristic occurs (Fig. 32.4), i. e., 280 kV/cm for the case of GaN, 900 kV/cm for the case of AlN, 60 kV/cm for the case of InN, and 8 kV/cm for the case of GaAs. (After [32.29] with permission, copyright AIP)

found for the other III–V nitride semiconductors, AlN and InN, and GaAs; see Figs. 32.17b–d.

We now compare the transient electron transport characteristics for the materials. From Fig. 32.17, it is clear that certain materials exhibit higher peak overshoot velocities and longer overshoot relaxation times. It is not possible to fairly compare these different semiconductors by applying the same applied electric field strength to each of the materials, as the transient effects occur over such a disparate range of applied electric field strengths in each material. In order to facilitate such a comparison, we choose a field strength equal to twice the critical applied electric field strength for each material. Figure 32.18 shows a comparison of the velocity overshoot effects amongst the four materials considered in this analysis, i. e., GaN, AlN, InN, and GaAs. It is clear that among the three III–V nitride semiconductors considered, InN exhibits superior transient electron transport characteristics. In particular, InN has the largest overshoot velocity and the distance over which this overshoot occurs, 0.6 μm, is longer than in either GaN or AlN. GaAs exhibits a longer overshoot relaxation distance, approximately 0.7 μm, but the electron drift velocity exhibited by InN is greater than that of GaAs for all distances.

32.2.8 Electron Transport: Conclusions

In this section, steady-state and transient electron transport results, corresponding to the III–V nitride semiconductors, GaN, AlN, and InN, were presented, these results being obtained from our Monte Carlo simulations of the electron transport within these materials. Steady-state electron transport was the dominant theme of our analysis. In order to aid in the understanding of these electron transport characteristics, a comparison was made between GaN and GaAs. Our simulations showed that GaN is more robust to variations in crystal temperature and doping concentration than GaAs, and an analysis of our Monte Carlo simulation results showed that polar optical phonon scattering plays the dominant role in accounting for these differences in behavior. This analysis was also performed for the other III–V nitride semiconductors considered in this analysis – AlN and InN – and similar results were obtained. Finally, we presented some key transient electron transport results. These results indicated that the transient electron transport that occurs within InN is the most pronounced of all of the materials under consideration in this review (GaN, AlN, InN, and GaAs).

32.3 Electron Transport Within III–V Nitride Semiconductors: A Review

Pioneering investigations into the material properties of the III–V nitride semiconductors, GaN, AlN, and InN, were performed during the earlier half of the 20th Century [32.65–67]. The III–V nitride semiconductor materials available at the time, small crystals and powders, were of poor quality, and completely unsuitable for device applications. Thus, it was not until the late 1960s, when *Maruska* and *Tietjen* [32.68] employed chemical vapor deposition to fabricate GaN, that interest in the III–V nitride semiconductors experienced a renaissance. Since that time, interest in the III–V nitride semiconductors has been growing, the material properties of these semiconductors improving considerably over the years. As a result of this research effort, there are currently a number of commercial devices available that employ the III–V nitride semiconductors. More III–V nitride semiconductor-based device applications are currently under development, and these should become available in the near future.

In this section, we present a brief overview of the III–V nitride semiconductor electron transport field. We start with a survey describing the evolution of the field. In particular, the sequence of critical developments that have occurred that contribute to our current understanding of the electron transport mechanisms within the III–V nitride semiconductors, GaN, AlN, and InN, is chronicled. Then, some of the more recent literature related to the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, is presented, with particular emphasis being placed on the most recent developments in the field that have occurred in the 21st Century, and how such developments have shaped our understanding of the electron transport mechanisms within these materials. Finally, frontiers for further research and investigation are presented.

This section is organized in the following manner. In Sect. 32.3.1, we present a brief survey describing the evolution of the field. Then, in Sect. 32.3.2, the developments that have occurred in the 21st Century are discussed. Finally, frontiers for further research and investigation are presented in Sect. 32.3.3.

32.3.1 Evolution of the Field

The favorable electron transport characteristics of the III–V nitride semiconductors, GaN, AlN, and InN, have long been recognized. As early as the 1970s, *Littlejohn* et al. [32.13] pointed out that the large polar optical phonon energy characteristic of GaN, in conjunction with its large intervalley energy separation, suggests a high saturation electron drift velocity for this material. As high-frequency electron device performance is, to a large degree, determined by this saturation electron drift velocity [32.14], the recognition of this fact ignited enhanced interest in this material and its III–V nitride semiconductor compatriots, AlN and InN. This enhanced interest, and the developments which have transpired as a result of it, are responsible for the III–V nitride semiconductor industry of today.

In 1975, *Littlejohn* et al. [32.13] were the first to report results obtained from semi-classical Monte Carlo simulations of the steady-state electron transport within bulk wurtzite GaN. A one-valley model for the conduction band was adopted in their analysis. Steady-state electron transport, for both parabolic and nonparabolic band structures, was considered in their analysis, nonparabolicity being treated through the application of the Kane model [32.43]. The primary focus of their investigation was the determination of the velocity-field characteristic associated with GaN. All donors were assumed to be ionized, and the free electron concentration was taken to be equal to the dopant concentration. The scattering mechanisms considered were:

1. Ionized impurity
2. Polar optical phonon
3. Piezoelectric
4. Acoustic deformation potential.

For the case of the parabolic band, in the absence of ionized impurities, they found that the electron drift velocity monotonically increases with the applied electric field strength, saturating at a value of about $2.5\times10^7\,\mathrm{cm/s}$ for the case of high applied electric fields. In contrast, for the case of the nonparabolic band, and in the absence of ionized impurities, a region of negative differential mobility was found, the electron drift velocity achieving a maximum of about $2\times10^7\,\mathrm{cm/s}$ at an applied electric field strength of about $100\,\mathrm{kV/cm}$, with further increases in the applied electric field strength resulting in a slight decrease in the corresponding electron drift velocity. The role of ionized impurity scattering was also investigated by *Littlejohn* et al. [32.13].

In 1993, *Gelmont* et al. [32.16] reported on ensemble semi-classical two-valley Monte Carlo simulations of the electron transport within bulk wurtzite GaN, this analysis improving upon the analysis of *Littlejohn* et al. [32.13] by incorporating intervalley scattering into the simulations. They found that the negative differential mobility found in bulk wurtzite GaN is much more pronounced than that found by *Littlejohn* et al. [32.13], and that intervalley transitions are responsible for this. For a doping concentration of $10^{17}\,\mathrm{cm^{-3}}$, *Gelmont* et al. [32.16] demonstrated that the electron drift velocity achieves a peak value of about $2.8\times10^7\,\mathrm{cm/s}$ at an applied electric field of about $140\,\mathrm{kV/cm}$. The impact of intervalley transitions on the electron distribution function was also determined and shown to be significant. The impact of doping and compensation on the velocity-field characteristic associated with bulk wurtzite GaN was also examined.

Since these pioneering investigations, ensemble Monte Carlo simulations of the electron transport within GaN have been performed numerous times. In particular, in 1995 *Mansour* et al. [32.18] reported the use of such an approach in order to determine how the crystal temperature influences the velocity-field characteristic associated with bulk wurtzite GaN. Later that year, *Kolník* et al. [32.19] reported on employing full-band Monte Carlo simulations of the electron transport within bulk wurtzite GaN and bulk zinc blende GaN, finding that bulk zinc blende GaN exhibits a much higher low-field electron drift mobility than bulk wurtzite GaN. The peak electron drift velocity corresponding to bulk zinc blende GaN was found to be only marginally greater than that exhibited by bulk wurtzite GaN, however. In 1997, *Bhapkar* and

Shur [32.22] reported on employing ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within bulk and confined wurtzite GaN. Their simulations demonstrated that the two-dimensional electron gas within a confined wurtzite GaN structure will exhibit a higher low-field electron drift mobility than bulk wurtzite GaN, by almost an order of magnitude, this being in agreement with experiment. In 1998, *Albrecht* et al. [32.27] reported on employing ensemble semi-classical five-valley Monte Carlo simulations of the electron transport within bulk wurtzite GaN, with the aim of determining elementary analytical expressions for a number of electron transport metrics corresponding to bulk wurtzite GaN, for the purposes of device modeling.

Electron transport within the other III–V nitride semiconductors, AlN and InN, has also been studied using ensemble semi-classical Monte Carlo simulations of the electron transport. In particular, by employing ensemble semi-classical three-valley Monte Carlo simulations, the velocity–field characteristic associated with bulk wurtzite AlN was studied and reported by *O'Leary* et al. [32.24] in 1998. They found that AlN exhibits the lowest peak and saturation electron drift velocities of the III–V nitride semiconductors considered in this analysis. Similar simulations of the electron transport within bulk wurtzite AlN were also reported by *Albrecht* et al. [32.25] in 1998. The results of *O'Leary* et al. [32.24] and *Albrecht* et al. [32.25] were found to be quite similar. The first known simulation of the electron transport within bulk wurtzite InN was the semi-classical three-valley Monte Carlo simulation of *O'Leary* et al. [32.23], reported in 1998. InN was demonstrated to have the highest peak and saturation electron drift velocities of the III–V nitride semiconductors. The subsequent ensemble full-band Monte Carlo simulations of *Bellotti* et al. [32.28], reported in 1999, produced results similar to those of *O'Leary* et al. [32.23].

The first known study of transient electron transport within the III–V nitride semiconductors was that performed by *Foutz* et al. [32.21], reported in 1997. In this study, ensemble semi-classical three-valley Monte Carlo simulations were employed in order to determine how the electrons within wurtzite and zinc blende GaN, initially in thermal equilibrium, respond to the sudden application of a constant electric field. The velocity overshoot that occurs within these materials was examined. It was found that the electron drift velocities that occur within the zinc blende phase of GaN are slightly greater than those exhibited by the wurtzite phase owing to the slightly the smaller effective mass of the electrons within the zinc blende phase of this material. A comparison with the transient electron transport that occurs within GaAs was made. Using the results from this analysis, a determination of the minimum transit time as a function of the distance displaced since the application of the applied electric field was performed for all three materials considered in this study: wurtzite GaN, zinc blende GaN, and GaAs. For distances in excess of 0.1 μm, both phases of GaN were shown to exhibit superior performance (reduced transit time) when contrasted with that associated with GaAs.

A more general analysis, in which transient electron transport within GaN, AlN, and InN was studied, was performed by *Foutz* et al. [32.29], and reported in 1999. As with their previous study, *Foutz* et al. [32.29] determined how electrons, initially in thermal equilibrium, respond to the sudden application of a constant electric field. For GaN, AlN, InN, and GaAs, it was found that the electron drift velocity overshoot only occurs when the applied electric field exceeds a certain critical applied electric field strength unique to each material. The critical applied electric field strengths, 140 kV/cm for the case of wurtzite GaN, 450 kV/cm for the case of AlN, 65 kV/cm for the case of InN (this was the critical applied electric field strength corresponding to our previous InN parameter selections [32.29]), and 4 kV/cm for the case of GaAs, were shown to correspond to the peak electron drift velocity in the velocity-field characteristic associated with each of these materials; recall Fig. 32.4. It was found that InN exhibits the highest peak overshoot velocity, and that this overshoot lasts over prolonged distances compared with AlN, InN, and GaAs. A comparison with the results of experiment was performed.

In addition to Monte Carlo simulations of the electron transport within these materials, a number of other types of electron transport studies have been performed. In 1975, for example, *Ferry* [32.14] reported on the determination of the velocity-field characteristic associated with wurtzite GaN using a displaced Maxwellian distribution function approach. For high applied electric fields, *Ferry* [32.14] found that the electron drift velocity associated with GaN monotonically increases with the applied electric field strength (it does not saturate), reaching a value of about 2.5×10^7 cm/s at an applied electric field strength of 300 kV/cm. The device implications of this result were further explored by *Das* and *Ferry* [32.15]. In 1994, *Chin* et al. [32.17] reported on a detailed study of the dependence of the low-field electron drift mobilities associated with the III–V nitride semiconductors, GaN, AlN, and InN, on crystal temperature and doping concentration. An analytical expression for the low-field electron drift mobility μ determined using a variational principle, was used for the purposes of this analysis. The results obtained were contrasted with those

from experiment. A subsequent mobility study was reported in 1997 by *Look* et al. [32.38]. Then, in 1998, *Weimann* et al. [32.26] reported on a model for determining how the scattering of electrons by the threading dislocations within bulk wurtzite GaN influence the low-field electron drift mobility. They demonstrated why the experimentally measured low-field electron drift mobility associated with this material is much lower than that predicted from Monte Carlo analyses: threading dislocations were not taken into account in the Monte Carlo simulations of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN.

While the negative differential mobility exhibited by the velocity-field characteristics associated with the III–V nitride semiconductors, GaN, AlN, and InN, is widely attributed to intervalley transitions, and while direct experimental evidence confirming this has been presented [32.69], *Krishnamurthy* et al. [32.34] suggest that the inflection points in the bands located in the vicinity of the Γ valley are primarily responsible for the negative differential mobility exhibited by wurtzite GaN instead. The relative importance of these two mechanisms (intervalley transitions and inflection point considerations) were evaluated by *Krishnamurthy* et al. [32.34], for both bulk wurtzite GaN and an AlGaN alloy.

32.3.2 Developments in the 21st Century

Thus far, the electron transport results that have been presented have focused on developments in our understanding of the nature of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN that were acquired in the latter part of the 20th Century. There have been a number of interesting more recent developments in the study of the electron transport within the III–V nitride semiconductors which have influenced the direction of thought in this field. On the experimental front, in 2000 *Wraback* et al. [32.33] reported on the use of a femtosecond optically detected time-of-flight experimental technique in order to experimentally determine the velocity–field characteristic associated with bulk wurtzite GaN. They found that the peak electron drift velocity, 1.9×10^7 cm/s, is achieved at an applied electric field strength of 225 kV/cm. No discernible negative differential mobility was observed. *Wraback* et al. [32.33] suggested that the large defect density characteristic of the GaN samples they employed, which were not taken into account in Monte Carlo simulations of the electron transport within this material, accounts for the difference between this experimental result and that obtained using simulation. They also suggested that decreasing the intervalley energy separation from about 2 eV to 340 meV, as suggested by the experimental results of *Brazel* et al. [32.58], may also account for these observations.

The determination of the electron drift velocity from experimental measurements of the unity gain cutoff frequency f_t has been pursued by a number of researchers. The key challenge in these analyses is the de-embedding of the parasitics from the experimental measurements so that the true intrinsic saturation electron drift velocity may be obtained. *Eastman* et al. [32.70] present experimental evidence which suggests that the saturation electron drift velocity within bulk wurtzite GaN is about 1.2×10^7–1.3×10^7 cm/s. A more recent report, by *Oxley* and *Uren* [32.71], suggests a value of 1.1×10^7 cm/s. The role of self-heating was also probed by *Oxley* and *Uren* [32.71] and shown to be relatively insignificant. A completely satisfactory explanation for the discrepancy between these results and those from the Monte Carlo simulations has yet to be provided.

Wraback et al. [32.72] performed a subsequent study on the transient electron transport within wurtzite GaN. In particular, using their femtosecond optically detected time-of-flight experimental technique in order to experimentally determine the velocity overshoot that occurs within bulk wurtzite GaN, they observed substantial velocity overshoot within this material. In particular, a peak transient electron drift velocity of 7.25×10^7cm/s was observed within the first 200 fs after photoexcitation for an applied electric field strength of 320 kV/cm. These experimental results were shown to be consistent with the theoretical predictions of *Foutz* et al. [32.29].

On the theoretical front, there have been a number of recent developments. In 2001, *O'Leary* et al. [32.30] presented an elementary, one-dimensional analytical model for the electron transport within the III–V compound semiconductors, and applied it to the cases of wurtzite GaN and GaAs. The predictions of this analytical model were compared with those of Monte Carlo simulations and were found to be in satisfactory agreement. Hot-electron energy relaxation times within the III–V nitride semiconductors were studied by *Matulionis* et al. [32.73] and reported in 2002. *Bulutay* et al. [32.74] studied the electron momentum and energy relaxation times within the III–V nitride semiconductors and reported the results of this study in 2003. It is particularly interesting to note that their arguments add considerable credence to the earlier inflection point argument of *Krishnamurthy* et al. [32.34]. In 2004, *Brazis* and *Raguotis* [32.75] reported on the results of a Monte Carlo study involving additional phonon modes and a smaller intervalley energy separation for bulk wurtzite GaN. Their results were found to

be much closer to the experimental results of *Wraback* et al. [32.33] than those found previously.

The influence of hot-phonons on the electron transport mechanisms within the III–V nitride semiconductors, GaN, AlN, and InN, has been the focus of considerable investigation. In particular, in 2004 *Silva* and *Nascimento* [32.76], *Gökden* [32.77], and *Ridley* et al. [32.78], to name just three, presented results related to this research focus. These results suggest that hot-phonon effects play a significant role in influencing the nature of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN. In particular, *Ridley* et al. [32.78] point out that the saturation electron drift velocity and the applied electric field strength at which the peak in the velocity–field characteristic occurs are both influenced by hot-phonon effects. The role that hot-phonons play in influencing device performance was studied by *Matulionis* and *Liberis* [32.79]. Research into the role that hot-phonons play in influencing the electron transport mechanisms within the III–V nitride semiconductors, GaN, AlN, and InN, seems likely to continue into the foreseeable future.

Over the past few years, there have been a number of developments that have occurred that have further enriched our understanding of electron transport within GaN. In 2011, *Ilgaz* et al. [32.80] studied the energy relaxation of hot electrons within AlGaN/GaN/GaN heterostructures. Then, in 2012, *Naylor* et al. [32.81] examined the steady-state and transient electron transport that occurs within bulk wurtzite GaN using an analytical band structure that more accurately reflects the nature of the actual bandstructure. In 2012, *Naylor* et al. [32.82] also examined the electron transport that occurs within dilute GaN_xAs_{1-x} samples. In 2013, *Bellotti* et al. [32.83] employed a full-band model in order to determine the velocity-field characteristics associated with AlGaN alloys. Also in 2013, *Dasgupta* et al. [32.84] estimated the hot electron relaxation time associated with wurtzite GaN using a series of electrical measurements. In 2013, *Zhang* et al. [32.85] determined the hot-electron relaxation time within lattice-matched InAlN/AlN/GaN heterostructures. The potential for electron device structures was then explored. Clearly, the study of electron transport within wurtzite GaN remains an area of active inquiry. The same situation applies for the cases of AlN and InN, and alloys of these materials [32.50, 86–95]. Further information, regarding our current understanding of the nature of the electron transport within these materials, may be found in the recently published review articles of *Hadi* et al. [32.96] and *Siddiqua* et al. [32.97].

32.3.3 Future Perspectives

It is clear that our understanding of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, is, at present at least, in a state of flux. A complete understanding of the electron transport mechanisms within these materials has yet to be achieved, and is the subject of intense current research. Most troubling is the discrepancy between the results of experiment and those of simulation. There are a two principal sources of uncertainty in our analysis of the electron transport mechanisms within these materials: (1) uncertainty in the material properties, and (2) uncertainty in the underlying physics. We discuss each of these subsequently.

Uncertainty in the material parameters associated with the III–V nitride semiconductors, GaN, AlN, and InN, remains a key source of ambiguity in the analysis of the electron transport with these materials [32.32]. Even for bulk wurtzite GaN, the most studied of the III–V nitride semiconductors considered in this analysis, uncertainty in the band structure remains an issue [32.58]. The energy gap associated with InN and the effective mass associated with this material continue to fuel debate; see, for example, *Davydov* et al. [32.55], *Wu* et al. [32.56], and *Matsuoka* et al. [32.57]. Variations in the experimentally determined energy gap associated with InN, observed from sample to sample, further confound matters. *Shubina* et al. [32.98], for example, suggested that nonstoichiometry within InN may be responsible for these variations in the energy gap. Further research will have to be performed in order to confirm this. Given this uncertainty in the band structures associated with the III–V nitride semiconductors, it is clear that new simulations of the electron transport will have to be performed once researchers have settled on the appropriate band structures. We thus view our present results as a baseline, the sensitivity analysis of *O'Leary* et al. [32.32] providing some insights into how variations in the band structures will impact upon the results.

Uncertainty in the underlying physics is considerable. The source of the negative differential mobility remains a matter to be resolved. The presence of hot-phonons within these materials, and how such phonons impact upon the electron transport mechanisms within these materials, remains another point of contention. It is clear that a deeper understanding of these electron transport mechanisms will have to be achieved in order for the next generation of III–V nitride semiconductor-based devices to be properly designed.

32.4 Conclusions

In this chapter, we reviewed analyses of the electron transport within the III–V nitride semiconductors GaN, AlN, and InN. In particular, we have discussed the evolution of the field, surveyed the more recent literature, and presented frontiers for further investigation and analysis. In order to narrow the scope of this review, we focused on the electron transport within bulk wurtzite GaN, AlN, and InN for the purposes of this paper. Most of our discussion focused upon results obtained from our ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within these materials, our results conforming with state-of-the-art III–V nitride semiconductor orthodoxy.

We began our review with the Boltzmann transport equation, since this equation underlies most analyses of the electron transport within semiconductors. A brief description of our ensemble semi-classical three-valley Monte Carlo simulation approach to solving the Boltzmann transport equation was then provided. The material parameters, corresponding to bulk wurtzite GaN, AlN, and InN, were then presented. We then used these material parameter selections, and our ensemble semi-classical three-valley Monte Carlo simulation approach, to determine the nature of the steady-state and transient electron transport within the III–V nitride semiconductors. Finally, we presented some recent developments on the electron transport within these materials, and pointed to fertile frontiers for further research and investigation.

Acknowledgments. The authors gratefully acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada. The work at Rensselaer Polytechnic Institute (M. S. Shur) was supported by the US Army Cooperative Research Agreement (Project Monitor Dr. Meredith Reed).

References

32.1 S. Strite, H. Morkoç: J. Vac. Sci. Technol. B **10**, 1237 (1992)
32.2 H. Morkoç, S. Strite, G.B. Gao, M.E. Lin, B. Sverdlov, M. Burns: J. Appl. Phys. **76**, 1363 (1994)
32.3 S.N. Mohammad, H. Morkoç: Prog. Quantum Electron. **20**, 361 (1996)
32.4 S.J. Pearton, J.C. Zolper, R.J. Shul, F. Ren: J. Appl. Phys. **86**, 1 (1999)
32.5 M.A. Khan, J.W. Yang, W. Knap, E. Frayssinet, X. Hu, G. Simin, P. Prystawko, M. Leszczynski, I. Grzegory, S. Porowski, R. Gaska, M.S. Shur, B. Beaumont, M. Teisseire, G. Neu: Appl. Phys. Lett. **76**, 3807 (2000)
32.6 X. Hu, J. Deng, N. Pala, R. Gaska, M.S. Shur, C.Q. Chen, J. Yang, G. Simin, M.A. Khan, J.C. Rojo, L.J. Schowalter: Appl. Phys. Lett. **82**, 1299 (2003)
32.7 W. Lu, V. Kumar, E.L. Piner, I. Adesida: IEEE Trans. Electron Devices **50**, 1069 (2003)
32.8 A. Jiménez, Z. Bougrioua, J.M. Tirado, A.F. Braña, E. Calleja, E. Muñoz, I. Moerman: Appl. Phys. Lett. **82**, 4827 (2003)
32.9 A.A. Burk Jr., M.J. O'Loughlin, R.R. Siergiej, A.K. Agarwal, S. Sriram, R.C. Clarke, M.F. MacMillan, V. Balakrishna, C.D. Brandt: Solid-State Electron. **43**, 1459 (1999)
32.10 M. Umeno, T. Egawa, H. Ishikawa: Mater. Sci. Semicond. Process. **4**, 459 (2001)
32.11 A. Krost, A. Dadgar: Phys. Status Solidi A **194**, 361 (2002)
32.12 C.L. Tseng, M.J. Youh, G.P. Moore, M.A. Hopkins, R. Stevens, W.N. Wang: Appl. Phys. Lett. **83**, 3677 (2003)
32.13 M.A. Littlejohn, J.R. Hauser, T.H. Glisson: Appl. Phys. Lett. **26**, 625 (1975)
32.14 D.K. Ferry: Phys. Rev. B **12**, 2361 (1975)
32.15 P. Das, D.K. Ferry: Solid-State Electron. **19**, 851 (1976)
32.16 B. Gelmont, K. Kim, M. Shur: J. Appl. Phys. **74**, 1818 (1993)
32.17 V.W.L. Chin, T.L. Tansley, T. Osotchan: J. Appl. Phys. **75**, 7365 (1994)
32.18 N.S. Mansour, K.W. Kim, M.A. Littlejohn: J. Appl. Phys. **77**, 2834 (1995)
32.19 J. Kolník, I.H. Oğuzman, K.F. Brennan, R. Wang, P.P. Ruden, Y. Wang: J. Appl. Phys. **78**, 1033 (1995)
32.20 M. Shur, B. Gelmont, M.A. Khan: J. Electron. Mater. **25**, 777 (1996)
32.21 B.E. Foutz, L.F. Eastman, U.V. Bhapkar, M.S. Shur: Appl. Phys. Lett. **70**, 2849 (1997)
32.22 U.V. Bhapkar, M.S. Shur: J. Appl. Phys. **82**, 1649 (1997)
32.23 S.K. O'Leary, B.E. Foutz, M.S. Shur, U.V. Bhapkar, L.F. Eastman: J. Appl. Phys. **83**, 826 (1998)
32.24 S.K. O'Leary, B.E. Foutz, M.S. Shur, U.V. Bhapkar, L.F. Eastman: Solid State Commun. **105**, 621 (1998)
32.25 J.D. Albrecht, R.P. Wang, P.P. Ruden, M. Farahmand, K.F. Brennan: J. Appl. Phys. **83**, 1446 (1998)
32.26 N.G. Weimann, L.F. Eastman, D. Doppalapudi, H.M. Ng, T.D. Moustakas: J. Appl. Phys. **83**, 3656 (1998)
32.27 J.D. Albrecht, R.P. Wang, P.P. Ruden, M. Farahmand, K.F. Brennan: J. Appl. Phys. **83**, 4777 (1998)
32.28 E. Bellotti, B.K. Doshi, K.F. Brennan, J.D. Albrecht, P.P. Ruden: J. Appl. Phys. **85**, 916 (1999)
32.29 B.E. Foutz, S.K. O'Leary, M.S. Shur, L.F. Eastman: J. Appl. Phys. **85**, 7727 (1999)
32.30 S.K. O'Leary, B.E. Foutz, M.S. Shur, L.F. Eastman: Solid State Commun. **118**, 79 (2001)

32.31 T.F. de Vasconcelos, F.F. Maia Jr., E.W.S. Caetano, V.N. Freire, J.A.P. da Costa, E.F. da Silva Jr.: J. Cryst. Growth **246**, 320 (2002)
32.32 S.K. O'Leary, B.E. Foutz, M.S. Shur, L.F. Eastman: J. Electron. Mater. **32**, 327 (2003)
32.33 M. Wraback, H. Shen, J.C. Carrano, T. Li, J.C. Campbell, M.J. Schurman, I.T. Ferguson: Appl. Phys. Lett. **76**, 1155 (2000)
32.34 S. Krishnamurthy, M. van Schilfgaarde, A. Sher, A.-B. Chen: Appl. Phys. Lett. **71**, 1999 (1997)
32.35 D.K. Ferry, C. Jacoboni (Eds.): *Quantum Transport in Semiconductors* (Plenum, New York 1992)
32.36 N.W. Ashcroft, N.D. Mermin: *Solid State Physics* (Saunders College, Philadelphia 1976)
32.37 P.A. Sandborn, A. Rao, P.A. Blakey: IEEE Trans. Electron Devices **36**, 1244 (1989)
32.38 D.C. Look, J.R. Sizelove, S. Keller, Y.F. Wu, U.K. Mishra, S.P. DenBaars: Solid State Commun. **102**, 297 (1997)
32.39 N.A. Zakhleniuk, C.R. Bennett, B.K. Ridley, M. Babiker: Appl. Phys. Lett. **73**, 2485 (1998)
32.40 B.R. Nag: *Electron Transport in Compound Semiconductors* (Springer, Berlin, Heidelberg 1980)
32.41 M.S. Krishnan, N. Goldsman, A. Christou: J. Appl. Phys. **83**, 5896 (1998)
32.42 R. Oberhuber, G. Zandler, P. Vogl: Appl. Phys. Lett. **73**, 818 (1998)
32.43 W. Fawcett, A.D. Boardman, S. Swain: J. Phys. Chem. Solids **31**, 1963 (1970)
32.44 A. Bykhovski, B. Gelmont, M. Shur, A. Khan: J. Appl. Phys. **77**, 1616 (1995)
32.45 A.D. Bykhovski, V.V. Kaminski, M.S. Shur, Q.C. Chen, M.A. Khan: Appl. Phys. Lett. **68**, 818 (1996)
32.46 P. Lugli, D.K. Ferry: IEEE Trans. Electron Devices **32**, 2431 (1985)
32.47 K. Seeger: *Semiconductor Physics: An Introduction*, 9th edn. (Springer, Berlin, Heidelberg 2004)
32.48 S.K. O'Leary, B.E. Foutz, M.S. Shur, L.F. Eastman: J. Mater. Sci. Mater. Electron. **17**, 87 (2006)
32.49 W.A. Hadi, P.K. Guram, M.S. Shur, S.K. O'Leary: J. Appl. Phys. **113**, 113709 (2013)
32.50 V.M. Polyakov, F. Schwierz, F. Fuchs, J. Furthmüller, F. Bechstedt: Appl. Phys. Lett. **94**, 022102 (2009)
32.51 W.R.L. Lambrecht, B. Segall: Band structure of pure group III nitrides. In: *Properties of Group III Nitrides*, EMIS Datareviews, ed. by J.H. Edgar (Inspec, London 1994) p. 125
32.52 M.A. Littlejohn, J.R. Hauser, T.H. Glisson: J. Appl. Phys. **48**, 4587 (1977)
32.53 J.S. Blakemore: J. Appl. Phys. **53**, R123 (1982)
32.54 T.L. Tansley, C.P. Foley: J. Appl. Phys. **59**, 3241 (1986)
32.55 V.Y. Davydov, A.A. Klochikhin, V.V. Emtsev, S.V. Ivanov, V.V. Vekshin, F. Bechstedt, J. Furthmüller, H. Harima, A.V. Mudryi, A. Hashimoto, A. Yamamoto, J. Aderhold, J. Graul, E.E. Haller: Phys. Status Solidi B **230**, R4 (2002)
32.56 J. Wu, W. Walukiewicz, K.M. Yu, J.W. Ager III, E.E. Haller, H. Lu, W.J. Schaff, Y. Saito, Y. Nanishi: Appl. Phys. Lett. **80**, 3967 (2002)
32.57 T. Matsuoka, H. Okamoto, M. Nakao, H. Harima, E. Kurimoto: Appl. Phys. Lett. **81**, 1246 (2002)
32.58 E.G. Brazel, M.A. Chin, V. Narayanamurti, D. Kapolnek, E.J. Tarsa, S.P. DenBaars: Appl. Phys. Lett. **70**, 330 (1997)
32.59 J.G. Ruch: IEEE Trans. Electron Devices **19**, 652 (1972)
32.60 M.S. Shur, L.F. Eastman: IEEE Trans. Electron Devices **26**, 1677 (1979)
32.61 M. Heiblum, M.I. Nathan, D.C. Thomas, C.M. Knoedler: Phys. Rev. Lett. **55**, 2200 (1985)
32.62 A. Palevski, M. Heiblum, C.P. Umbach, C.M. Knoedler, A.N. Broers, R.H. Koch: Phys. Rev. Lett. **62**, 1776 (1989)
32.63 A. Palevski, C.P. Umbach, M. Heiblum: Appl. Phys. Lett. **55**, 1421 (1989)
32.64 A. Yacoby, U. Sivan, C.P. Umbach, J.M. Hong: Phys. Rev. Lett. **66**, 1938 (1991)
32.65 E. Tiede, M. Thimann, K. Sensse: Chem. Ber. **61**, 1568 (1928)
32.66 W.C. Johnson, J.B. Parsons, M.C. Crew: J. Phys. Chem. **36**, 2651 (1932)
32.67 R. Juza, H. Hahn: Z. Anorg. Allg. Chem. **239**, 282 (1938)
32.68 H.P. Maruska, J.J. Tietjen: Appl. Phys. Lett. **15**, 327 (1969)
32.69 Z.C. Huang, R. Goldberg, J.C. Chen, Y. Zheng, D.B. Mott, P. Shu: Appl. Phys. Lett. **67**, 2825 (1995)
32.70 L.F. Eastman, V. Tilak, J. Smart, B.M. Green, E.M. Chumbes, R. Dimitrov, H. Kim, O.S. Ambacher, N. Weimann, T. Prunty, M. Murphy, W.J. Schaff, J.R. Shealy: IEEE Trans. Electron Devices **48**, 479 (2001)
32.71 C.H. Oxley, M.J. Uren: IEEE Trans. Electron Devices **52**, 165 (2005)
32.72 M. Wraback, H. Shen, J.C. Carrano, C.J. Collins, J.C. Campbell, R.D. Dupuis, M.J. Schurman, I.T. Ferguson: Appl. Phys. Lett. **79**, 1303 (2001)
32.73 A. Matulionis, J. Liberis, L. Ardaravičius, M. Ramonas, I. Matulionienė, J. Smart: Semicond. Sci. Technol. **17**, L9 (2002)
32.74 C. Bulutay, B.K. Ridley, N.A. Zakhleniuk: Phys. Rev. B **68**, 115205 (2003)
32.75 R. Brazis, R. Raguotis: Appl. Phys. Lett. **85**, 609 (2004)
32.76 A.A.P. Silva, V.A. Nascimento: J. Lumin. **106**, 253 (2004)
32.77 S. Gökden: Physica E **23**, 198 (2004)
32.78 B.K. Ridley, W.J. Schaff, L.F. Eastman: J. Appl. Phys. **96**, 1499 (2004)
32.79 A. Matulionis, J. Liberis: IEE Proc. Circuits Devices Syst. **151**, 148 (2004)
32.80 A. Ilgaz, S. Gőkden, R. Tűlek, A. Teke, S. Őzçelik, E. Özbay: Eur. Phys. J. Appl. Phys. **55**, 30102 (2011)
32.81 D.R. Naylor, A. Dyson, B.K. Ridley: Solid State Commun. **152**, 549 (2012)
32.82 D.R. Naylor, A. Dyson, B.K. Ridley: J. Appl. Phys. **111**, 053703 (2012)
32.83 E. Bellotti, F. Bertazzi, S. Shishehchi, M. Matsubara, M. Goano: IEEE Trans. Electron Devices **60**, 3204 (2013)
32.84 S. Dasgupta, J. Lu, Nidhi, A. Raman, C. Hurni, G. Gupta, J.S. Speck, U.K. Mishra: Appl. Phys. Express **6**, 034002 (2013)

32.85 J.-Z. Zhang, A. Dyson, B.K. Ridley: Appl. Phys. Lett. **102**, 062104 (2013)
32.86 M. Farahmand, C. Garetto, E. Bellotti, K.F. Brennan, M. Goano, E. Ghillino, G. Ghione, J.D. Albrecht, P.P. Ruden: IEEE Trans. Electron Devices **48**, 535 (2001)
32.87 T. Li, R.P. Joshi, R.D. del Rosario: IEEE Trans. Electron Devices **49**, 1511 (2002)
32.88 C. Sevik, C. Bulutay: IEE Proc. Optoelectron. **150**, 86 (2003)
32.89 S.K. O'Leary, B.E. Foutz, M.S. Shur, L.F. Eastman: Appl. Phys. Lett. **87**, 222103 (2005)
32.90 V.M. Polyakov, F. Schwierz: Appl. Phys. Lett. **88**, 032101 (2006)
32.91 S.K. O'Leary, B.E. Foutz, M.S. Shur, L.F. Eastman: Appl. Phys. Lett. **88**, 152113 (2006)
32.92 J.S. Thakur, R. Naik, V.M. Naik, D. Haddad, G.W. Auner, H. Lu, W.J. Schaff: J. Appl. Phys. **99**, 023504 (2006)
32.93 Z. Yarar: Phys. Status Solidi B **244**, 3711 (2007)
32.94 V.M. Polyakov, F. Schwierz, I. Cimalla, M. Kittler, B. Lübbers, A. Schober: J. Appl. Phys. **106**, 023715 (2009)
32.95 P. Siddiqua, W.A. Hadi, A.K. Salhotra, M.S. Shur, S.K. O'Leary: J. Appl. Phys. **117**, 125705 (2015)
32.96 W.A. Hadi, M.S. Shur, S.K. O'Leary: J. Mater. Sci. Mater. Electron. **25**, 4675 (2014)
32.97 P. Siddiqua, W.A. Hadi, M.S. Shur, S.K. O'Leary: J. Mater. Sci. Mater. Electron. **26**, 4475 (2015)
32.98 T.V. Shubina, S.V. Ivanov, V.N. Jmerik, M.M. Glazov, A.P. Kalavarskii, M.G. Tkachman, A. Vasson, J. Leymarie, A. Kavokin, H. Amano, I. Akasaki, K.S.A. Butcher, Q. Guo, B. Monemar, P.S. Kop'ev: Phys. Status Solidi A **202**, 377 (2005)

33. II–IV Semiconductors for Optoelectronics: CdS, CdSe, CdTe

Minoru Isshiki, Jifeng Wang

Owing to their suitable band gaps and high absorption coefficients, Cd-based compounds such as CdTe and CdS are the most promising photovoltaic materials available for low-cost high-efficiency solar cells. Additionally, because of their large atomic number, Cd-based compounds such as CdTe and CdZnTe, have been applied to radiation detectors. For these reasons, preparation techniques for these materials in the polycrystalline films and bulk single crystals demanded by these devices have advanced significantly in recent decades, and practical applications have been realized in optoelectronic devices. This chapter mainly describes the application of these materials in solar cells and radiation detectors and introduces recent progress.

33.1 **Background** 853
33.2 **Solar Cells** 853
33.2.1 Basic Description of Solar Cells 853
33.2.2 Design of Cd-Based Solar Cells 854
33.2.3 Development of CdS/CdTe Solar Cells 855
33.2.4 CdZnTe Solar Cells 858
33.2.5 The Future of Cd-Based Solar Cells 858
33.3 **Radiation Detectors** 858
33.3.1 Basic Description of Semiconductor Radiation Detectors 858
33.3.2 CdTe and CdZnTe Radiation Detectors 859
33.3.3 Performance of CdTe and CdZnTe Detectors 860
33.3.4 Applications of CdTe and CdZnTe Detectors 862
33.4 **Conclusions** 863
References 863

33.1 Background

Cd-based compounds are very important semiconductor materials in the II–VI family. The attraction of Cd-based binary and ternary compounds arises from their promising applications as solar cells, γ- and x-ray detectors etc. These devices made from Cd-based materials are being widely applied in many fields.

33.2 Solar Cells

With the development of human society, energy sources in the earth are being slowly exhausted and we are faced with a serious problem. The solar cell is one substitute for fossil fuels and is being realized throughout the world. For this reason, solar-cell technologies have been developed since work was started by *Becquerel* in 1839 [33.1]. The solar cell has now been applied to daily life, industry, agriculture, space exploration, military affairs etc.

Solar cells have many advantages. Firstly, sunlight as an energy source for power generation is not only limitless but can be used freely. Secondly, since light is directly converted to electricity, the conversion process is clean, noise-free and not harmful to the environment, unlike a mechanical power generator. Thirdly, solar cells need little maintenance.

33.2.1 Basic Description of Solar Cells

A solar cell is a semiconductor device that directly converts light energy into electrical energy through the photovoltaic process. The basic structure of a solar cell is shown in Fig. 33.1. A typical solar cell consists of a junction formed between an n-type and a p-type semiconductor, of the same (homojunction) or different materials (heterojunction), an antireflection coating and Ohmic collecting electrodes. When the light irradiates the surface of a solar cell, photons with an energy

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_33

greater than the band gap of the semiconductor are absorbed by the semiconductor material. This absorption activates electron transitions from the valence to the conduction band, so that electron–hole pairs are generated. If these carriers can diffuse into the depletion region before they recombine, they can be separated by the applied electric field. At the p–n junction, the negative electrons diffuse into the n-type region and the positive holes diffuse into the p-type region. They are then collected by electrodes, resulting in a voltage difference between the two electrodes. When an external load is connected, electric current flows through the load. This is the origin of the solar cell's photocurrent.

Three parameters can be used to describe the performance of a solar cell: the short-circuit current density (J_{SC}), the open-circuit voltage (V_{OC}) and the fill factor (FF).

The short-circuit current density J_{SC} is the photocurrent output from a solar cell when the output terminals are short-circuited. In the ideal case, this is equal to the current density generated by the light J_L and is proportional to the incident photon flux. J_{SC} is determined by the spectral response of the device, the junction depth, and the series (internal) resistance, R_s. The open-circuit voltage V_{OC} is the voltage appearing across the output terminals of the cell when there is no load present, i. e., when $J = 0$. The relationship between V_{OC} and J_{SC} can be described by the diode equation, i. e.,

$$V_{OC} = \frac{nk_BT}{e} \ln\left(\frac{J_{SC}}{J_0} + 1\right), \tag{33.1}$$

where e, n, J_0, k_B and T are the charge on an electron, the diode ideality factor, the reverse saturation-current density, the Boltzmann constant and the absolute temperature, respectively. The fill factor (FF) is the ratio of the maximum electrical power available from the cell (i. e., at the operating point J_m, V_m) in Fig. 33.2) to the product ofV_{OC} and J_{SC}. It describes the rectangularity of the photovoltaic output characteristic.

$$FF = \frac{V_mJ_m}{V_{oc}J_{sc}} = \frac{P_m}{V_{oc}J_{sc}}. \tag{33.2}$$

The conversion efficiency (η) of a solar cell is determined as

$$\eta = \frac{P_m}{P_s} = \frac{V_{oc}J_{sc}FF}{P_s}. \tag{33.3}$$

Where P_s is the incident illumination power and P_m is the output electricity power, both being per unit area. The relationship between the short-circuit current density J_{SC} and the open-circuit voltage V_{OC} is shown in Fig. 33.2.

33.2.2 Design of Cd-Based Solar Cells

Solar cells can be made from silicon (Si), III–V compounds such as GaAs, or II–VI compounds such as Zn-based and Cd-based compound semiconductors. Of these materials, CdTe is the most attractive because of a number of advantages. CdTe is a direct-band-gap material with a band gap of 1.54 eV at room temperature. This is very close to the theoretically calculated optimum value for solar cells. CdTe has a high absorption coefficient (above 10^5 cm^{-1} at a wavelength of 700 nm), so that approximately 90% of the incident light is absorbed by a layer thickness of only 2 μm (compared with around 10 μm for Si), cutting down the quantity of semiconductor required. Therefore, CdTe is one of the most promising photovoltaic materials available for use in low-cost high-efficiency solar cells.

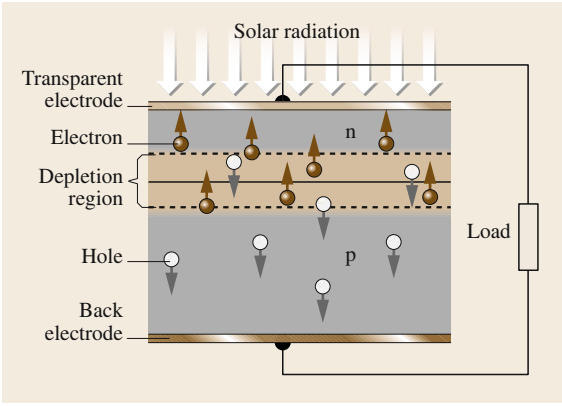


Fig. 33.1 The principle of photovoltaic devices

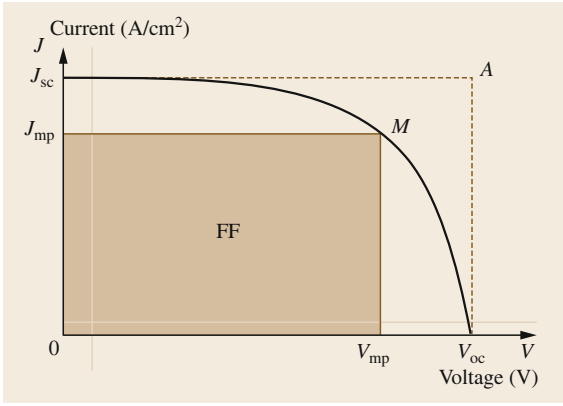


Fig. 33.2 Photovoltaic output characteristics of a solar cell

CdTe is the only material in which both n- and p-type conductivity can be easily controlled by doping acceptor or donor impurities in II–VI compound semiconductors. Furthermore, due to its high absorption coefficient and small carrier diffusion length, the junction must be formed closed to the surface, which reduces the carrier lifetime through surface recombination. In spite of this, the conversion efficiency has been gradually enhanced from 4% [33.2] to 16% [33.3] since the first CdTe-based solar cell was fabricated. CdTe solar cells can be made from polycrystalline or single-crystal material. However, it was found that this p–n homojuction structure is unstable due to the aging behavior of dopants in CdTe films and bulk single crystals [33.4]. Furthermore, it is also difficult to fabricate shallow p–n junctions with highly conducting surface layers, and high surface recombination velocities led to substantial losses that could not easily be avoided. For this reason, the CdS/CdTe heterojunction solar cell was proposed.

It has been proved that CdS/CdTe is a high-efficiency heterojunction solar cell. The theoretical calculation shows that this cell has a conversion efficiency of as high as 29% [33.5], and it is considered a promising alternative to the more widely used silicon devices. The typical structure of a CdS/CdTe solar cell is glass/indium tin oxide (ITO)/CdS/CdTe/back electrode, as seen in Fig. 33.3. The CdS/CdTe solar cell is based on the heterojunction formed between n-type CdS (only n-CdS is available) and p-type CdTe and is fabricated in a superstrate configuration where the incident light passes through the glass substrate, which has a thickness of 2−4 mm. This window glass is transparent, strong and cheap. This glass substrate also protects the active layers from the environment, and it provides the mechanical strength of the device. The outer face of the panel often has an antireflection coating to enhance absorption efficiency. A transparent conducting oxide (TCO) layer, usually tin oxide or indium tin oxide (ITO), acts as the front contact to the device. This is needed to reduce the series resistance of the device, which would otherwise arise from the thinness of the CdS layer. Usually, the polycrystalline n-type CdS layer is deposited onto the TCO. Due to its wide band gap ($E_g \approx 2.4$ eV at 300 K) it is transparent down to wavelengths of around 516 nm. The thickness of this layer is typically 100−300 nm. The CdTe layer is p-type-doped and its thickness is typically around 10 μm. Generally, the carrier concentration in CdS layer is 2–3 orders greater than that of the CdTe layer. As a result, the potential is applied mostly to the CdTe absorber layer, i. e., the depletion region is mostly within the CdTe layer. Therefore, the photogenerated carriers can be effectively separated in this active region.

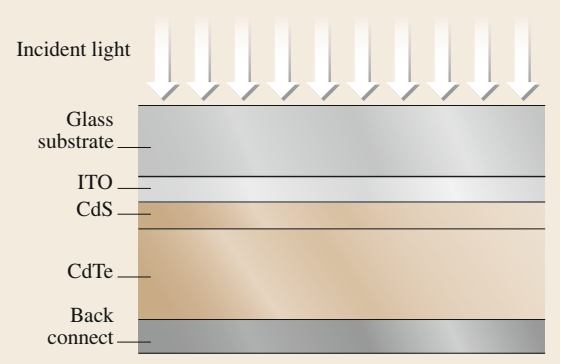


Fig. 33.3 A typical CdS/CdTe solar cell

The electrode preparation is very important in obtaining Ohmic contact or minimal series resistivity. Since there are no metals with a work function higher than the hole affinity of p-type CdTe (−5.78 eV) [33.6], it is difficult to produce a complete Ohmic contact on p-type CdTe surface. Many methods have been tried to fabricate Ohmic contact to p-type CdTe. Different metals or alloy and compounds such as Au, Al, and ZnTe/Cu etc. have also been used to serve as electrode materials. Although significant effort has been applied to this issue, the junction still inevitably displays some Schottky-diode characteristics. This is a problem that will be continuously studied.

33.2.3 Development of CdS/CdTe Solar Cells

The layers of CdS and CdTe for solar cells can be prepared using various techniques, which are summarized in Table 33.1. In these techniques, evaporation, close-space sublimation (CSS), screen printing and electrodeposition (ED) have been demonstrated to be very effective for fabricating high-efficiency large-area CdS/CdTe solar cells.

Study of solar cells based on CdTe single crystals and polycrystalline films were started as early as 1963 [33.14]. The cells were prepared with a p-type copper-telluride surface film as an integral part of the photovoltaic junction. The junction was thought to be a heterojunction between the p-type copper telluride and the n-type CdTe, with the transition region extending well into the CdTe side. A similar structure was reported for single-crystal cells. A conversion efficiency of up to 6% was obtained in the film cells, and 7.5% for single-crystal cells.

Dutton and *Muller* [33.15] fabricated CdS/CdTe solar cells by standard vapor-deposition procedures with a very thin CdTe layer. The existence of the

Table 33.1 Some fabrication techniques and conversion efficiency of CdS/CdTe solar cells

Fabrication techniques of CdS and CdTe	Conversion		Reference
	Efficiency (%)	Area (cm^2)	
Evaporation: physical vapor deposition (PVD), and chemical vapor deposition (CVD)	11.8	0.3	[33.7]
Close-spaced sublimation (CSS)	16	1.0	[33.3]
	8.4	7200	[33.8]
Electrodeposition (ED)	14.2	0.02	[33.9]
Screen printing (SP) and sintering	12.8	0.78	[33.10]
Metalorganic CVD (MOCVD)	11.9	0.08	[33.11]
PVD vacuum evaporation	11.8	0.3	[33.7]
Molecular beam epitaxy (MBE)	10.5	0.08	[33.12]
Sputtering deposition (SD)	14	0.09	[33.13]

interfacial layer was confirmed using x-ray diffraction studies, and substantiated by diode photocurrent measurements. The results showed a threshold photon energy that corresponds to the band gap of CdTe. The thin CdTe layer resulted in excellent rectification ratios and high reverse breakdown voltages. The observed photoresponse, *I–V* (current versus voltage), and *C–V* (capacitance versus voltage) characteristics were consistence with those predicted by a semiconductor heterojunction model of the CdS–CdTe interface. *Mitchell* et al. [33.16] prepared a variety of CdS/CdTe heterojunction solar cells with an ITO coating and a glycerol antireflection coating by vacuum evaporation of n-CdS films onto single-crystal p-CdTe substrates. Comparisons were made between cells prepared using different substrate resistivity, substrate surface preparations, and CdS film resistivity. The mechanisms of controlling the dark junction current, photocarrier collection, and photovoltaic properties were modeled, taking account of the interface states of the junction. A conversion efficiency of 7.9% was obtained under 85 mW/cm^2 of solar simulator illumination.

At the end of the 1970s, a new process, screen printing and sintering, was developed for fabricating CdS/CdTe polycrystal solar cells by *Matsushita* and coworkers [33.10, 17–20]. This technique proved to be a very effective method for fabricating large-area CdS/CdTe

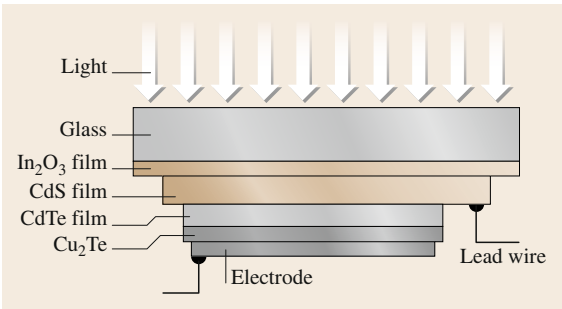


Fig. 33.4 Cross section of a ceramic thin-film CdTe solar cell

film solar cells. There are several important parameters in the screen-printing technique: the viscosity of the paste, the mesh number of the screen, the snap-off distance between the screen and the substrate and the pressure and speed of the squeegee. As cadmium chloride ($CdCl_2$) has a low melting point ($T_m = 568\,°C$), and forms a eutectic with both CdS and CdTe at low temperature, and the vapor pressure at 600 °C is high enough to allow complete volatilization of $CdCl_2$ after the sintering process, $CdCl_2$ is used as an ideal flux for sintering both CdS and CdTe.

Nakayama et al. [33.18] prepared the first thin-film CdS/CdTe solar cells using the screen-printing technique (Fig. 33.4). A glass plate, which was coated successively with In_2O_3 film and CdS ceramic thin film, was used as a transparent Ohmic contact substrate. The n-type CdS film with a thickness of about 20 μm and a resistivity of 0.2 Ω cm was prepared on this substrate by the screen-printing method. The CdTe paste was printed onto the CdS layer using the same technique and then heat-treated in an N_2 atmosphere at temperatures of 500–800 °C. As a result, the p-type CdTe layer with a thickness of 10 μm showed a resistivity of 0.1–1 Ω cm. A silver (Ag) paste electrode was applied to the p-type layer and an In–Ga alloy was applied to the CdS layer. The cell, with an active area of 0.36 cm^2, showed an intrinsic solar conversion efficiency of 8.1% under an illumination of 140 mW/cm^2 solar simulator (AM0). Subsequently, this group continued to work on improving CdS/CdTe solar cells [33.10, 19, 20]. From the practical point of view, a modular solar cell is necessary. For this purpose, thin-film CdS/CdTe solar cells with an efficiency of 6.3% were prepared on a borosilicate-glass substrate of $4 \times 4\,cm^2$ by successively screen printing and heating (sintering) of each paste of CdS, CdTe and carbon [33.19]. Although they fabricated some structures, including dividing the CdTe film on one substrate into five small cells, to prevent increases in the series resistivity, a 1 W module only showed an efficiency of 2.9% from 25 elemental cells with a $4 \times 4\,cm^2$ substrate. A main reason for this

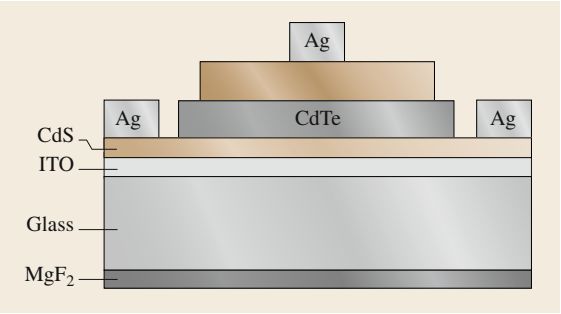


Fig. 33.5 The cross-sectional structure of a CdS/CdTe solar cell fabricating by the screen-printing method

low efficiency was thought to be the high series resistivity.

In order to decrease the resistivity of the carbon electrode further, this group changed the heating conditions of the carbon electrodes for CdTe and examined the effect of impurities in the carbon paste on the characteristics of solar cells [33.10]. They found that the series resistance (R_s) and the conversion efficiency (η) of solar cells exhibited strong dependence on the oxygen (O_2) partial pressure during the heating process. The results showed that the addition of Cu to the carbon paste causes R_s and the diode factor (n) to decrease, resulting in a remarkable improvement in η. Using a low-resistance contact electrode, which was made from 50 ppm added Cu carbon paste, a cell with an active area of 0.78 cm^2 displayed $V_{oc} = 0.754$ V, $I_{sc} = 0.022$ A, FF= 0.606 and $\eta = 12.8\%$.

Britt and *Ferekides* [33.21] reported the fabrication and characteristics of high-efficiency thin-film CdS/CdTe heterojunction solar cells. CdS films with 0.07–0.10 μm in thickness were deposited on a 0.5-μm SnO_2:F transparent conducting oxide layer by chemical-bath deposition and p-CdTe films with a thickness of 5 μm were deposited on the CdS layer by using the CSS method. Prior to the deposition of CdTe, CdS/SnO_2:F/glass was annealed in a H_2 atmosphere at temperatures of 350–425 °C for 5–20 min. The characteristics of the CdS/CdTe solar cell were $V_{oc} = 843$ mV, $J_{sc} = 25.1$ mA/cm^2, FF = 74.5%, corresponding to a total area greater than 1 cm^2 with an air mass 1.5 (AM1.5) conversion efficiency of 15.8%.

With modern growth techniques, high-efficiency CdS/CdTe solar cells have been developed using ultrathin CdS films having a thickness of 50 nm [33.3]. Figure 33.5 shows the cross-sectional structure of a CdS/CdTe solar cell. CdS films were deposited on an ITO glass substrates by the metal organic chemical vapor deposition (MOCVD) technique, and CdTe films were subsequently deposited by the CSS technique (see Fig. 33.6) for the fabrication of CdS/CdTe solar cells.

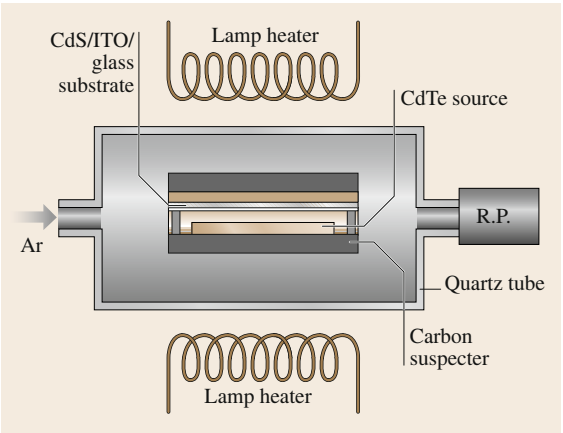


Fig. 33.6 Schematic of the close-spaced sublimation apparatus used for depositing CdTe films

A thin-film CdS/CdTe solar cell with an area of 1.0 cm^2 under AM1.5 conditions showed $V_{oc} = 840$ mV, $J_{sc} = 26.08$ mA/cm^2 and FF = 73%. As a result, a high conversion efficiency of 16.0% was achieved. Furthermore, they made developments in improving the uniformity of thickness and film qualities of CdS in order to obtain more efficient solar cells and to realize film deposition on large-area substrates of 30 × 60 cm or more [33.22]. They found clear differences in the electrical and optical properties between high- and low-quality CdS films. A photovoltaic conversion efficiency of 10.5% was achieved by a solar module with an area of 1376 cm^2 under AM1.5 measurement conditions.

In addition to dry processes such as screen printing, close-spaced sublimation etc., electrodeposition [33.23–26] has been also investigated for preparing polycrystalline CdTe layers for solar cells. *Awakura* and coworkers [33.24–26] reported the cathodic deposition behavior of CdTe thin layers under irradiation by visible light using ammoniacal basic aqueous solution as an electrolytic bath. Both deposition current density and current efficiency for the CdTe deposition were enhanced by irradiation. The deposited rate was over 10 times high than nonphotoassisted electrodeposition. This was believed to represent significant progress in reducing the cost of Cd-based solar cells.

To date, solar cells with an n-CdS/p-CdTe heterojunction have been reported with efficiencies as high as 16% [33.3]. Recently, it was reported that thin-layered n-CdS/p-CdTe heterojunction solar cells have already been manufactured industrially [33.27]. n-CdS film with a thickness range of 500–1000 Å was deposited from an aqueous solution directly onto a transparent conductive oxide (TCO) substrate. After the CdS film was annealed for densification and grain growing, p-CdTe layer was formed by electrochemical deposition.

The CdTe film was then annealed in air at 450 °C. Subsequently, monolithic TCO/CdS/CdTe was cut into discrete cells using an infrared laser. After other preparing process, a maximum efficiency of 10.6% for a CdTe 0.94 m^2 module with a power of 91.5 W was fabricated. The test results showed good stability. For practical applications, a 10-MW CdTe solar-cell manufacturing plant has been constructed [33.27].

33.2.4 CdZnTe Solar Cells

The ternary compound cadmium zinc telluride (CdZnTe, CZT) has potential for the preparation of high-efficiency tandem solar cell since its band gap can be tuned from 1.45 to 2.26 eV [33.28, 29]. *McCandless* et al. [33.28] deposited $Cd_{1-x}Zn_xTe$ films using the physical vapor deposition (PVD) and vapor transport deposition (VTD) techniques. The film composition was between 0.35 and 0.6, corresponding to a band gap from 1.7 to 1.9 eV. Post-deposition treatment of CdZnTe films in $ZnCl_2$ vapor at 400 °C resulted in no change to the alloy composition and caused recrystallization. Solar cells made from $Cd_{1-x}Zn_xTe$ films with $x \approx 0.35$ exhibited $V_{oc} = 0.78$ V and $J_{sc} < 10\,mA/cm^2$. These results were similar to those obtained from CdS window layers. Analysis of the spectral response indicated that $Cd_{1-x}Zn_xTe$ with $x \approx 0.35$ has a band gap of about 1.7 eV. *Gidiputti* et al. [33.29] used two deposition technologies, cosputtering from CdTe and ZnTe targets and coclose-spaced sublimation (CCSS) from CdTe and ZnTe powders, to prepare CdZnTe films. A structure similar to the CdTe superstrate configuration was initially utilized for cell fabrication: glass/ITO/CdS/CZT/graphite. Typical solar cell parameters obtained for CdZnTe/CdS (when $E_g(CdZnTe) = 1.72$ eV) solar cells were $V_{oc} = 720$ mV, and $J_{sc} = 2\,mA/cm^2$. However, the spectral response indicated increasing loss of photocurrent at longer wavelengths. In order to improve the collection efficiency, CdZnTe devices were annealed in a H_2 atmosphere. This postprocessing treatment showed that J_{sc} increased to over $10\,mA/cm^2$. Study of CdZnTe solar cells is being carried out in various directions.

33.2.5 The Future of Cd-Based Solar Cells

From the development process of Cd-based solar cells, the study on the laboratory scale is being transfered to large-scale deposition and cell fabrication. Modules are being developed throughout the world, using the screen printing, evaporative deposition and close-space sublimation [33.30–32]

In order to make a commercial Cd-based photovoltaic cell with its full potential, a large-scale high-throughput manufacturing process is required. The process must possess excellent yields and produce high-efficiency devices with good long-term stability. In order to progress towards these goals, a pilot system for continuous, inline processing of CdS/CdTe devices has been developed [33.33–35]. High-quality low-cost thin-film CdTe modules with an average total area efficiency of 8% and cascaded production-line yield of > 70% have been manufactured, and technology-development programs will further increase production-line module efficiency to 13% within five years [33.35].

33.3 Radiation Detectors

A radiation detector is a device that converts a radiation ray into electrical signal. They can be divided into gas-filled detectors, scintillation detectors and semiconductor detectors. Since semiconductor radiation detectors have a high spectrometric performance, and can be made portable, they have been applied in nuclear safeguards, medical physics and imaging diagnosis, industrial safety, nondestructive analysis, security and monitoring, nonproliferation and astrophysics [33.36, 37]. Meanwhile, they are also being gradually improved.

33.3.1 Basic Description of Semiconductor Radiation Detectors

The principle of operation of a semiconductor radiation detector is shown in Fig. 33.7. When high-energy photons (x- or γ-rays) radiate the detector, they lose the energy to induce electron–hole pairs in the semiconductor through photoelectric or Compton interactions, thereby increasing the conductivity of the material. In order to detect the change in conductivity, a bias voltage is required. An externally applied electric field separates the electron–hole pairs before they recombine, and electrons drift towards the anode, holes to the cathode. The charges are then collected by the electrodes. The collected charges produce a current pulse, whose integral equals the total charge generated by the incident photons, indicating the radiation intensity. The read-out goes through a charge-sensitive preamplifier, followed by a shaping amplifier. The multichannel analyzer is used for spectroscopic measurements of the radiation ray. This type of detector can produce both

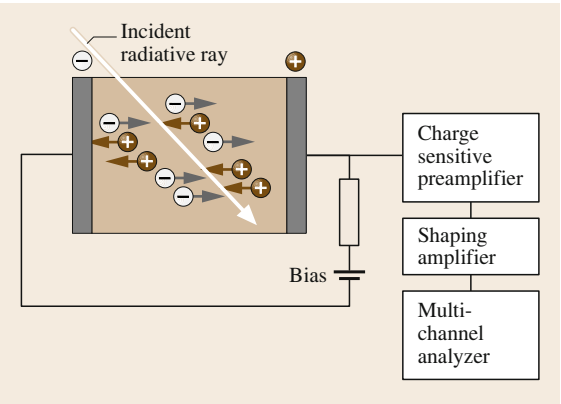


Fig. 33.7 The principle of operation of a semiconductor detector

count and energy information. Typically, detectors are manufactured in a square or rectangular configuration to maintain a uniform bias-current distribution throughout the active region. The wavelength of the peak response depends on the material's band-gap energy.

There are several important parameters to describe a detector performance. Detection efficiency is defined as the percentage of radiation incident on a detector system that is actually detected. Detector efficiency depends on the detector size and shape (larger areas and volumes are more sensitive), radiation type, material density (atomic number), and the depth of the detection medium. Energy resolution is one of the most important characteristics of semiconductor detectors. It is usually defined as the full-width at half-maximum (FWHM) of a single-energy peak at a specific energy. High resolution enables the system to more clearly resolve the peaks within a spectrum. So the reliability of the detector is also high. Detector sensitivity implies minimal detectable counts. It is defined as the number of counts that can be distinguished from the background. The relative detector response factor expresses the sensitivity of a detector relative to a standard substance. If the relative detector response factor is expressed on an equal-mass (-weight) basis, the determined sensitivity values can be substituted for the peak area.

33.3.2 CdTe and CdZnTe Radiation Detectors

Semiconductor radiation detectors are fabricated from a variety of materials including: germanium (Ge), silicon (Si), mercuric iodide (HgI_2), CdTe, CdZnTe etc. Typical detectors for a given application depend on several factors. Table 33.2 shows the features of some semiconductor detectors. From Table 33.2, Ge detectors have the best resolution, but require liquid-nitrogen cooling, which makes them impractical for portable applications. Si detectors also need cooling, and are inefficient in detecting photons with energies greater than a few tens of keV. CdTe and CdZnTe detectors possess their own advantages.

Single crystals of CdTe and CdZnTe are very important materials for the development of x- and γ-ray detectors. Currently Ge and Si detectors have to be used at liquid-nitrogen temperatures, and suffer from poor detection efficiency. Since the compound semiconductors CdTe and CdZnTe with high atomic number ($Z_{Cd} = 48, Z_{Zn} = 30, Z_{Te} = 52$) have a significantly higher photoelectric absorption efficiency in the $-100-500$ keV range, they can be fabricated into detectors that provide two advantages for use in portable instrumentation: (1) the large band-gap energy ($E_g > 1.54$ eV at room temperature) results in extremely small leakage current of a few nanoamps at room temperature. Therefore, these detectors have the potential to serve as useful detectors for radiation spectroscopy without the need for cryogenic cooling as required for more conventional silicon and hyper-pure germanium detectors; (2) the high density of the crystal provides excellent stopping power over a wide range of energies. The ability to operate at room temperature without the need for liquid-nitrogen cooling allows the construction of compact devices. CdTe and CdZnTe detectors can be fabricated into a variety of shapes and sizes, which makes it possible to produce detectors capable of meeting the requirements of a wide assortment of applications that are unsupported by other detector types. Their small size and relatively simple electronics allow them to open new areas of detector application. In many instances, CdTe and CdZnTe detectors can be substituted for other detector types in existing applications.

Table 33.2 The features of some semiconductor detectors

	NaI	Si	Ge	CdTe	$Cd_{0.9}Zn_{0.2}Te$
Bandgap (eV) at 300 K	–	1.14	0.67	1.54	1.58
Density (g/cm^3)	3.67	2.33	5.35	6.2	5.78
Energy per e–h pair (eV)	–	3.61	2.96	4.43	4.64
Detection/absorption	High	Low	Medium	High	High
Operation at room temp.	Yes	Yes	No	Yes	Yes
Spectrometic performance	Medium	High	Very high	High	High

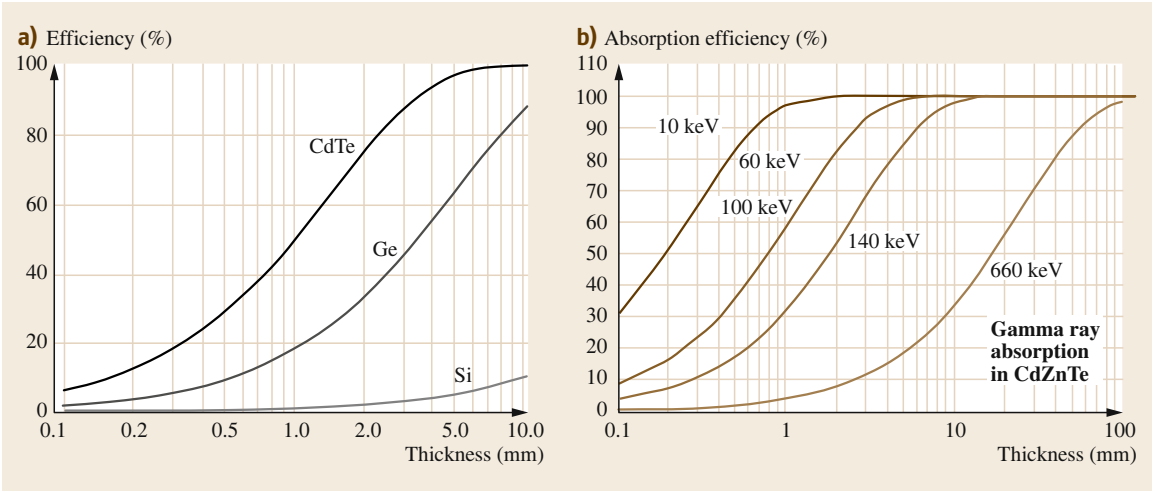


Fig. 33.8a,b γ-ray absorption in CdTe **(a)** and CdZnTe **(b)** detectors as a function of thickness

Figure 33.8 shows the γ-ray absorption efficiency in CdTe (Fig. 33.8a), and CdZnTe (Fig. 33.8b) detectors as a function of thickness. It was reported that, for a detector with a thickness of about 10 mm detecting 100 keV γ-rays, the detection efficiency for a CdTe detector is 100%, for a Ge detector 88%, for a Si detector only 12% [33.38]. In the case of a CdZnTe detector, CdZnTe with a thickness of only about 7 mm can absorb 100% of the γ-rays [33.39]. This efficiency assumes that the γ-ray is completely absorbed and the generated electron–hole pairs are completely collected. In fact, a large amount of charge loss occurs in CdTe and CdZnTe detector, depending on the quality of the materials.

33.3.3 Performance of CdTe and CdZnTe Detectors

Although CdTe and CdZnTe have many advantages over Si and Ge, including higher attenuation coefficients and lower leakage currents due to the wider bandgap, their disadvantages limit their application range [33.40]. The main disadvantage is that their crystals usually contain higher density defects. These defects are remaining impurities and structural defects such as mechanical cracks, twins, grain and tilt boundaries, and dislocations all lower the crystalline perfection of the materials.

In CdTe and CdZnTe detectors, the charge transport determines the collection efficiency of generated electron-hole pairs, and structural properties of materials control the uniformity of the charge transport. The problem of poor transport and collection of carriers, especially hole, can be overcome by applying high voltage to device and a range of electrode config. The former is related to high resistivity material, and later is related to the design of a detector. In addition, it was found that detected pulse signal intensity degraded with the time [33.41, 42]. This phenomenon was thought to be mainly related to the defects existing in bandgap. When the detector starts to work, these defects could act as slowly ionized acceptors. The low resistivity and poor quality material always exhibits such a polarization effect.

Summarizing above, improvement of a detector performance is equivalent to the improvement of the crystallinity of CdTe and CdZnTe. The key feature of all applications except substrate materials of CdTe and CdZnTe is the resistivity. This is because high resistivity can be obtained only by controlling native defects and impurity concentration. For many years, much effort was done in preparing high quality CdTe and CnZnTe single crystals.

Progresses in Crystal Growth of High Quality CdTe and CdZnTe

Various techniques have been applied to grow high-quality CdTe and CdZnTe single crystals. The Bridgman method [33.43, 44], the traveling-heater method [33.45, 46], growth from Te solvent [33.47], the gradient-freeze method [33.48, 49], and physical vapor transport [33.50, 51] are the most widely used. Crystals with applicable quality and size have become available, and have also fostered the rapid progress of research on CdTe.

As-grown CdTe single crystals commonly contain high concentration of both residual impurities and intrinsic defects. Preparation of high-purity high-resistivity CdTe and CdZnTe crystals has been widely attempted [33.52]. Early works was done by *Triboulet*

and *Marfaing* [33.53] in obtaining high-purity lightly compensated zone-melting growth following synthesis by the Bridgman method. The room-temperature carrier concentrations range from $1-5\times10^{13}\,\mathrm{cm}^{-3}$, and the resistivity from 100 to 400 Ω cm. The carrier mobility at 32 K reaches as high as $1.46\times10^{5}\,\mathrm{cm}^{2}/(\mathrm{V\,s})$. The total concentration of electrically active centers was estimated to be about $10^{14}\,\mathrm{cm}^{-3}$.

Usually, as-grown CdTe crystal shows p-type conductivity owing to the existence of remaining acceptor impurities or native defects. For this reason, many results have been reported in preparing high-resistivity CdTe. Chlorine (Cl) is thought to be a suitable donor for the compensation of these remaining acceptors because Cl can be doped quite uniformly due to its very small segregation coefficient [33.54].

The growth of high-resistivity CdTe:Cl single crystals with device quality was successfully performed by the traveling-heater method (THM) [33.54]. Solvent alloys for THM growth were synthesized in ampoules filled with Te, CdTe and $CdCl_2$ so that the molar ratio of Te/Cd was the same as in the solvent zone during the growth. The Cl concentration in the grown crystal was 2 weight-ppm. According to the dynamics of the crystal growth, it is important to control the shape of the solid–liquid to grow high-quality single crystals. Therefore, the solvent volume was optimized. The use of a slightly tilted seed from ⟨111⟩B was also effective in limiting the generation of twins with different directions. Single-crystal (111) wafers, larger than $30\times30\,\mathrm{mm}^2$ were successfully obtained from a grown crystal with a diameter of 50 mm. Pt/CdTe/In detectors with dimensions of $2\times2\times0.5\,\mathrm{mm}^3$ showed better energy resolution, because a higher electric field can be applied. The effective detector resistivity was estimated to be 10^{11} Ω cm.

CdTe doped with chlorine (Cl) or indium (In) with a resistivity of 3×10^{9} Ω cm and CdZnTe with a resistivity of 5×10^{10} Ω cm were grown by the high-pressure Bridgman (HPB) technique [33.55]. The material was polycrystalline with large grains and twins. Although the crystalline quality of HPB CdTe and CdZnTe is poor, the grains are large enough to obtain volume detectors of several cm^3. Photoluminescence (PL) spectra at 4 K showed that the free exciton could be observed and the FWHM of the bound excitons is very low. Some impurity emissions were identified. The detectors were fabricated from HPB CdTe and CdZnTe. These detectors showed excellent performance. Gamma-ray spectra were presented with high-energy resolution in an energy range from 60 to 600 keV. Using a $10\times10\times2\,\mathrm{mm}^3$ HPB detector, at a bias of 300 V, the peak at 122 keV from ^{57}Co had a FWHM of 5.2 KeV. Detectors with high-energy resolution were fabricated.

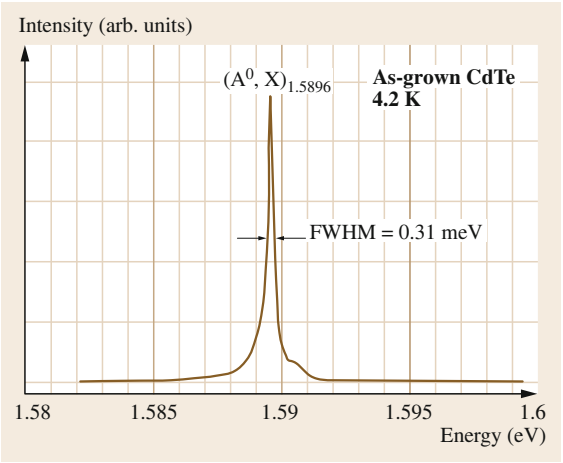


Fig. 33.9 A typical high-resolution PL spectrum of a high-purity CdTe sample

Because of the difference in vapor pressure between Cd and Te, grown crystals contain a large number of Cd vacancies (V_{Cd}). These Cd vacancies manifest acceptor behaviour. Therefore, high-resistivity CdTe can also be obtained by controlling the concentration of Cd vacancies. However, a prerequisite is that the purity of the CdTe crystal is high enough that the remaining impurities do not play a substantial role in determining the conductivity. Recently, the preparation of ultra-high-purity CdTe single crystals was reported [33.56, 57]. In order to obtain a high purity the starting materials of Cd and Te, Cd were first purified by vacuum distillation (VD) and the overlap zone-melting (OZM) method [33.56], while Te was purified using the normal freezing method [33.57]. The results of glow-discharge mass spectroscopy (GDMS) and the measurement of the residual resistivity ratio (RRR) showed that refined Cd and Te had the purity of 6N-up. Using the refined Cd and Te as starting materials, extremely high-purity twin-free CdTe single crystals were prepared by the traditional vertical Bridgman technique. Figure 33.9 shows a typical high-resolution PL spectrum of a high-purity CdTe sample. Only a sharp and strong emission line of (A^0, X) (an exciton bound to a neutral acceptor) at 1.5896 eV was observed. The FWHM of the (A^0, X) line is 0.31 meV. This is the narrowest value ever reported for CdTe single crystals grown by the Bridgman method. All of the above indicate that the sample is of high purity and quality. For a good performance detector, the leakage current must not exceed a few nanoamps, which requires a material with high resistivity. When growing CdTe crystals, the CdTe contain large number of Cd vacancies due to the extreme volatility of Cd. For these reason, CdTe single crystals were annealed in a Cd atmosphere in order to

obtain a high-resistivity crystal. Hall-measurement results showed that the conductive type changed from as-grown p-type to n-type at a Cd pressure around 1.5×10^{-2} atm, corresponding to a Cd source temperature of 500 °C. Samples annealed under these conditions showed a resistivity as high as $10^9\ \Omega$ cm.

Lachish [33.58] grew CdTe and CdZnTe crystals with excess cadmium in order to avoid the Te precipitates found in crystals grown by Bridgman method. During crystal growth one ampoule end was kept at a low temperature, which determines a constant, nearly atmospheric, vapor pressure in the system. The constant vapor pressure maintains a constant liquid composition and balanced amounts of cadmium and tellurium within the crystal. Although the highly donor-doped crystals are highly electrically conducting, successive annealing in Te vapor transforms the crystal wafers into a highly compensated state showing high electrical resistivity and high gamma sensitivity, depending on the doping level.

Improvements of Transport and Collection of Carriers in CdTe and CdZnTe Detectors

The transport and collection of carriers in CdTe and CdZnTe, especially holes, are the main problems that affect their performance. Poor carriers transport causes position-dependent charge collection and limits their capability as high-resolution spectrometers [33.59]. As a result, this limits the volume of detector. Therefore, the efficient detecting resolution and range are decreased. *Amman* et al. [33.60] studied the affect of nonuniform electron trapping on the performance of a γ-ray detector. An analysis of the induced charge signals indicated that regions of enhanced electron trapping are associated with inclusions, and that these regions extend beyond the physical size of the inclusions. Such regions introduce nonuniform electron trapping in the material that then degrades the spectroscopic performance of the material as a γ-ray detector. The measurements showed that the degree of nonuniformity that affects detector performance could be at the 1% level. Consequently, any useful characterization and analysis technique must be sensitive down to this level.

Detectors equipped with Ohmic contacts, and a grounded guard ring around the positive contact, have a fast charge-collection time. Conductivity adjustment, in detectors insensitive to hole trapping, optimizes the detector operation by a trade-off between electron lifetime and electrical resistance. Better hole collection has been achieved by designing detectors and electrodes of various geometries. *Nakazawa* et al. [33.61] improved the CdTe diode detectors by fabricating the guard-ring structure in the cathode face, and the leakage current of the detector decreased by more than an order of magnitude. The new CdTe detector was operated with a bias as high as 800 V at 20 °C and showed a good energy resolution (0.93 keV and 1 keV FWHM for 59.5 keV and 122 keV, respectively) and high stability for long-term operation at room temperature. Furthermore, they also developed a large-area ($20 \times 20\ \text{mm}^2$) detector with very high resolution. Owing to its high stopping power and high resolution at room temperature, pixel imagers have also been developed in the same group, and it is thought that this large diode detector could possibly be a substitute for scintillation detectors.

Since the carriers drift slow and have a short lifetime in CdTe detectors, the number of photons in the photo-peak is reduced and the spectrum is distorted by a tail towards lower energies. For these reasons, CdTe detectors are usually made into Schottky diodes, because this structure can withstand much higher bias voltage with a leakage current orders of magnitude lower than detectors with Ohmic contacts. *Takahashi* et al. [33.62] adopted the configuration of Schottky CdTe diode, which due to the low leaking current, makes it possible to apply a much higher bias voltage to ensure complete charge collection in relatively thin (< 1 mm) devices. Both the improved charge-collection efficiency and the low leakage current lead to an energy resolution of better than 600 eV FWHM at 60 keV for a $2 \times 2\ \text{mm}^2$ device without any electronics for charge-loss correction. Meanwhile, they also fabricated large-area detectors with dimensions of $21.5 \times 21.5\ \text{mm}^2$, with a thickness of 0.5 mm and an energy resolution of 2.8 keV. Stacked detectors can measure the energies as high as 300 keV. Furthermore, a large-array detector, consisting of 1024 individual CdTe diodes, was also made. Every detector had a dimension of $1.2 \times 5.0\ \text{mm}^2$. The total area, including the spaces between the detector elements, is $44 \times 44\ \text{mm}^2$. This array detector is expected to be used in next-generation Compton telescopes.

33.3.4 Applications of CdTe and CdZnTe Detectors

Owing to the convenience of the smaller collimator, better resolution and temperature stability, CdTe and CZT detectors have been used in safeguard applications by the International Atomic Energy Agency (IAEA) and some countries for over ten years [33.63, 64]. With the gradual improvement in performance, CdTe and CdZnTe detectors are replacing NaI detectors used in spent-fuel attribute tests [33.65], though their sensitivity is still low compared to NaI and Ge detectors, although already sufficient for many applications. CdTe detectors have a sensitive volume of

about 20–100 mm^3 and a probe diameter of 8–9 mm. CdZnTe detectors have a larger volume than CdTe detectors. The largest commercial CdZnTe detectors have a geometric volume of 1687 mm^3 ($15 \times 15 \times 7.5\ mm^3$). Detectors are mainly of hemispheric design to obtain high carrier-collection efficiency. These large-volume detectors have been made into portable and hand-held isotope-identification devices and are being used to detect radioactive sources.They will become commercially available in the near future.

At present, conventional x-ray film or scintillator mammograms are used in medical diagnostics such as screening for breast cancer. However, these show a nonlinear response to x-ray intensity and the detection quantum efficiency is low. Room-temperature semiconductor detectors such as CdTe and CdZnTe have favorable physical characteristics for medical applications. From the start of these investigations in the 1980s, rapid progress has been achieved [33.66–71]. *Barber* [33.67] presented results concerning, first, a CdTe two-dimensional (2-D) imaging system ($20 \times 30\ mm^2$ with 400×600 pixels) for dental radiology and, second, a CdZnTe fast pulse-correction method applied to a $5 \times 5 \times 5\ mm^3$ CdZnTe detector (energy resolution of 5% for a detection efficiency of 85% at 122 keV) for medical imaging. After that, a 2 mm-thick CdZnTe detector was fabricated for application to digital mammography [33.69]. The preliminary images showed high spatial resolution and efficiency. Furthermore, CdTe and CdZnTe detectors with a thickness of 0.15–0.2 mm were fabricated [33.71]. The detectors are indium-bump-bonded onto a small version of a chip. Their detection quantum efficiency was measured as 65%. This result showed that CdTe and CdZnTe detectors are superior to scintillator-based digital systems, whose quantum efficiency is typically around 30–40%. This showed that CdTe and CdZnTe detectors have potential applications in medical imaging, as well as industrially for nondestructive evaluation inspection.

In universe exploration, Cd-based detectors can be used in an advanced Compton telescope (ACT) planned as the next-generation space-based instrument devoted to observations of low/medium-energy γ-rays ($\approx$ 0.2–30 MeV) and to the nonthermal energy exploration telescope (NeXT) [33.72]. In the universe, radiation rays have a wide energy range of 0.5–80 keV. In order to detect this wide range of radiation, *Takahashi* et al. [33.73–75] proposed a new focal-plane detector based on the idea of combining an x-ray charge-coupled device (CCD) and a CdTe pixel detector as the wide-band x-ray imager (WXI). The WXI consists of a soft-x-ray imager and a hard-x-ray imager. For the detection of soft x-rays (10–20 keV) with high positional resolution, a CCD with a very thin dead layer will be used. For hard x-rays, CdTe pixel detectors serve as absorbers. This study is now under way.

33.4 Conclusions

Devices fabricated from Cd-based compounds, such as solar cells and radiation detectors, are being applied in our daily life. In the case of solar cells, although recent success have improved their conversion efficiency and reduced the cost, many problems remain. Fundamental understanding of the CdTe-based solar-cell properties is limited, particularly as a result of their polycrystalline nature. Therefore, the fundamental electronic properties of polycrystalline Cd-based thin films should be studied deeply. Other challenges are to reduce the cost and to lengthen the operating life span. These problems are being studied and solved [33.76].

In the case of radiation detectors, CdTe and CdZnTe detectors have many advantages, such as room-temperature operation, high count rates, small size, and direct conversion of photons to charge, which make them attractive candidates for a wide variety of applications in industrial gauging and analysis, as well as medical instrumentation and other areas. These detectors are being made available commercially at present. However, they have severe problems such as polarization effects, long-term stability and their high price. Further efforts should still be focused on the preparation of high-quality materials and improvement of the stability and reliability of detectors. We are confident that radiation detectors made from Cd-based compounds will achieve more widespread application.

References

33.1 E. Becquerel: Compt. Rend. Acad. Sci. (Paris) **9**, 561 (1839)

33.2 Y.A. Vodakov, G.A. Lomakina, G.P. Naumov, Y.P. Maslakovets: Sov. Phys. Solid State **2**, 1 (1960)

33.3 T. Aramoto, S. Kumazawa, H. Higuchi, T. Arita, S. Shibutani, T. Nishio, J. Nakajima, M. Tsuji, A. Hanafusa, T. Hibino, K. Omura, H. Ohyama, M. Murozono: Jpn. J. Appl. Phys. **36**, 6304 (1997)

33.4 B. Yang, Y. Ishikikawa, T. Miki, Y. Doumae, M. Ishiki: J. Cryst. Growth **179**, 410 (1997)
33.5 A.W. Brinkman: Properties of narrow gap cadmium-based compounds. In: *Electronic Materials Information Services*, Vol. 10, ed. by P. Capper (IEE, London 1994) p. 591
33.6 R.W. Swank: Phys. Rev. **156**, 844 (1967)
33.7 R.G. Little, M.J. Nowlan: Progr. Photovolt. **5**, 309 (1997)
33.8 Y.-S. Tyan, E.A. Perez-Albuerne: Proc. 16th IEEE Photovolt. Specialists Conf. (IEEE, New York 1982) p. 794
33.9 J.M. Woodcock, A.K. Turner, M.E. Özsan, J.G. Summers: Proc. 22nd IEEE Photovolt. Specialists Conf., Las Vegas (IEEE, New York 1991) p. 842
33.10 K. Kuribayashi, H. Matsumoto, H. Uda, Y. Komatsu, A. Nakano, S. Ikegami: Jpn. J. Appl. Phys. **22**, 1828 (1993)
33.11 J. Britt, C. Ferekides: Appl. Phys. Lett. **62**, 2851 (1993)
33.12 H.W. Schock, A. Shah: Proc. 14th Eur. Photovolt. Sol. Energy Conf., ed. by H.A. Ossenbrink, P. Helm, H. Ehmann (H.S. Stephens, Bedford 1997) p. 2000
33.13 A.D. Compaan, A. Gupta, J. Drayton, S.-H. Lee, S. Wang: Phys. Stat. Solid B **241**, 779 (2004)
33.14 D.A. Cusano: Solid State Electron. **6**, 217 (1963)
33.15 R.W. Dutton, R.S. Muller: Solid State Electron. **11**, 749 (1968)
33.16 K.W. Mitchell, A.L. Fahrenbruch, R.W. Bube: J. Appl. Phys. **48**, 4365 (1977)
33.17 H. Uda, A. Nakano, K. Kuribayashi, Y. Komatsu, H. Matsumoto, S. Ikegami: Jpn. J. Appl. Phys. **22**, 1822 (1983)
33.18 N. Nakayama, H. Matsumoto, K. Yamaguchi, S. Ikegami, Y. Hioki: Jpn. J. Appl. Phys. **15**, 2281 (1976)
33.19 S. Ikegami, T. Yamashita: J. Electron. Mater. **8**, 705 (1979)
33.20 N. Nakayama, H. Matsumoto, A. Nakano, S. Ikegami, H. Uda, T. Yamashita: Jpn. J. Appl. Phys. **19**, 703 (1980)
33.21 J. Britt, C. Ferikides: Appl. Phys. Lett. **62**, 2851 (1993)
33.22 M. Tsuji, T. Aramoto, H. Ohyama, T. Hibino, K. Omura: Jpn. J. Appl. Phys. **39**, 3902 (2000)
33.23 M.P.R. Panicker, M. Knaster, F.A. Kröger: J. Electrochem. Soc. **125**, 566 (1978)
33.24 K. Murase, H. Uchida, T. Hirato, Y. Awakura: J. Electrochem. Soc. **146**, 531 (1999)
33.25 K. Murase, M. Matsui, M. Miyake, T. Hirato, Y. Awakura: J. Electrochem. Soc. **150**, 44 (2003)
33.26 M. Miyake, K. Murase, H. Inui, T. Hirato, Y. Awakura: J. Electrochem. Soc. **151**, 168 (2004)
33.27 D.W. Cunningham, M. Rubcich, D. Skinner: Prog. Photovolt. **10**, 59 (2002)
33.28 B. McCandless, K. Dobson, S. Hegedus, P. Paulson: Proc. NCPV Sol. Program Rev. Meet., Denver (NREL, Golden 2003) p. 401, NREL/CD-520-33586
33.29 G. Gidiputti, P. Mahawela, M. Ramalingan, G. Sivaraman, S. Subramanian, C.S. Ferekides, D.L. Morel: Proc. NCPV Sol. Program Rev. Meet., Denver (NREL, Golden 2003) p. 896, NREL/CD-520-33586
33.30 P.D. Maycock: PV News **17**, 3 (1998)
33.31 R.C. Powell, U. Jayamaha, G.L. Dorer, H. McMaster: Proc. NCPV Photovolt. Program, Rev., ed. by M. Al-Jassim, J.P. Thornton, J.M. Gee (American Institute of Physics, New York 1995) p. 1456
33.32 D. Bonnet, H. Richter, K.-H. Jager: Proc. 13th Eur. Photovolt. Sol. Energy Conf., ed. by W. Freiesleben, W. Palz, H.A. Ossenbrink, P. Helm (Stephens, Bedford 1996) p. 1456
33.33 K. Zweibel, H. Ullal: Proc. 25th IEEE Photovolt. Specialists Conf., Washington DC 1996 (IEEE, New York 1996) p. 745
33.34 K.L. Barth, R.A. Enzenroth, W.S. Sampath: Proc. NCPV Sol. Program Rev. Meet., Denver (NREL, Golden 2003) p. 904, NREL/CD-520-33586
33.35 A. Abken, C. Hambro, P. Meyers, R. Powell, S. Zafar: Proc NCPV Sol. Program Rev. Meet., Denver (NREL, Golden 2003) p. 393, NREL/CD-520-33586
33.36 K. Zanio: In: *Semiconductors and Semimetals*, Vol. 13, ed. by R.K. Willardson, A.C. Beer (Academic, New York 1978) p. 164
33.37 R. Triboulet, Y. Marfaing, A. Cornet, P. Siffert: J. Appl. Phys. **45**, 2759 (1974)
33.38 G. Sato, T. Takahashi, M. Sugiho, M. Kouda, T. Mitani, K. Nakazawa, Y. Okada, S. Watanabe: IEEE Trans. Nucl. Sci. **48**, 950 (2001)
33.39 C. Szeles: Phys. Stat. Solid B **241**, 783 (2004)
33.40 H. Yoon, J.M. Van Scyoc, T.S. Gilbert, M.S. Goorsky, B.A. Brunett, J.C. Lund, H. Hermon, M. Schieber, R.B. James: Proc. Infrared Appl. Semicond. II. Symp. Boston, ed. by D.L. McDaniel Jr., M.O. Manasreh, R.H. Miles, S. Sivananthan, P.A. Warrendale (Materials Research Society, Pittsburgh 1998) p. 241
33.41 R.O. Bell, G. Entine, H.B. Serreze: Nucl. Instrum. Methods **117**, 267 (1974)
33.42 P. Siffert, J. Berger, C. Scharager, A. Cornet, R. Stuck, R.O. Bell, H.B. Serreze, F.V. Wald: IEEE Trans. Nucl. Sci. **23**, 159 (1976)
33.43 R.K. Route, M. Woff, R.S. Feigelson: J. Cryst. Growth **70**, 379 (1984)
33.44 K.Y. Lay, D. Nichols, S. McDevitt, B.E. Dean, C.J. Johnson: J. Cryst. Growth **86**, 118 (1989)
33.45 R.O. Bell, N. Hemmat, F. Wald: Phys. Stat. Solid A **1**, 375 (1970)
33.46 R. Triboulet, Y. Mafaing, A. Cornet, P. Siffert: J. Appl. Phys. **45**, 375 (1970)
33.47 K. Zanio: J. Electron. Mat. **3**, 327 (1974)
33.48 M. Azoulay, A. Raizman, G. Gafni, M. Roth: J. Cryst. Growth **101**, 256 (1990)
33.49 A. Tanaka, Y. Masa, S. Seto, T. Kawasaki: Mater. Res. Soc. Symp. Proc. **90**, 111 (1987)
33.50 W. Akutagawa, K. Zanio: J. Cryst. Growth **11**, 191 (1971)
33.51 C. Ceibel, H. Maier, R. Schmitt: J. Cryst. Growth **86**, 386 (1988)
33.52 M. Isshiki: *Wide-Gap II–VI Compounds for Opto-Electronic Applications* (Chapman Hall, London 1992) p. 3
33.53 R. Triboulet, Y. Mafaing: J. Electrochem. Soc. **120**, 1260 (1973)
33.54 M. Funaki, T. Ozaki, K. Satoh, R. Ohno: Nucl. Instrum. Methods A **322**, 120 (1999)

33.55 M. Fiederle, T. Feltgen, J. Meinhardt, M. Rogalla, K.W. Benz: J. Cryst. Growth **197**, 635 (1999)
33.56 B. Yang, Y. Ishikawa, Y. Doumae, T. Miki, T. Ohyama, M. Isshiki: J. Cryst. Growth **172**, 370 (1997)
33.57 S.H. Song, J. Wang, M. Isshiki: J. Cryst. Growth **236**, 165 (2002)
33.58 U. Lachish: CdTe *and* CdZnTe *Crystal Growth and Production of Gamma Radiation Detectors*, http://urila.tripod.com/crystal.htm
33.59 T. Takahashi, S. Watanabe: IEEE Trans. Nucl. Sci. **48**, 950 (2001)
33.60 M. Amman, J.S. Lee, P.N. Luke: J. Appl. Phys. **92**, 3198 (2002)
33.61 K. Nakazawa, K. Oonuki, T. Tanaka, Y. Kobayashi, K. Tamura, T. Mitani, G. Sato, S. Watanabe, T. Takahashi, R. Ohno, A. Kitajima, Y. Kuroda, M. Onishi: IEEE Trans. Nucl. Sci. **51**, 1881 (2004)
33.62 T. Takahashi, T. Mitani, Y. Kobayashi, M. Kouda, G. Sato, S. Watanabe, K. Nakazawa, Y. Okada, M. Funaki, R. Ohno, K. Mori: IEEE Trans. Nucl. Sci. **49**, 1297 (2002)
33.63 R. Arlt, D.E. Rundquist: Nucl. Instrum. Methods Phys. Res. A **380**, 455 (1996)
33.64 T. Prettyman: Proc. 2nd Workshop Sci. Mod. Technol. Safeguards, Albuquerque, ed. by C. Foggi, E. Petraglia (European Commission, Albuquerque 1998)
33.65 W.K. Yoon, Y.G. Lee, H.R. Cha, W.W. Na, S.S. Park: INMM J. Nucl. Mat. Manag. **27**, 19 (1999)
33.66 C. Scheiber, J. Chambron: Nucl. Instrum. Methods A **322**, 604 (1992)
33.67 H.B. Barber: J. Electron. Mater. **25**, 1232 (1996)
33.68 L. Verger, J.P. Bonnefoy, F. Glasser, P. Ouvrier-Buffet: J. Electron. Mater. **26**, 738 (1997)
33.69 S. Yin, T.O. Tümay, D. Maeding, J. Mainprize, G. Mawdsley, M.J. Yaffe, W.J. Hamilton: IEEE Trans. Nucl. Sci. **46**, 2093 (1999)
33.70 C. Scheiber: Nucl. Instrum. Methods A **448**, 513 (2000)
33.71 S. Yin, T.O. Tümay, D. Maeding, J. Mainprize, G. Mawdsley, M.J. Yaffe, E.E. Gordon, W.J. Hamilton: IEEE Trans. Nucl. Sci. **49**, 176 (2002)
33.72 T. Tanaka, T. Kobayashi, T. Mitani, K. Nakazawa, K. Oonuki, G. Sato, T. Takahashi, S. Watanabe: New Astron. Rev. **48**, 269 (2004)
33.73 T. Takahashi, B. Paul, K. Hirose, C. Matsumoto, R. Ohno, T. Ozaki, K. Mori, Y. Tomita: Nucl. Instr. Meth. A **436**, 111 (2000)
33.74 T. Takahashi, K. Nakazawa, T. Kamae, H. Tajima, Y. Fukazawa, M. Nomachi, M. Kokubun: SPI **4851**, 1228 (2002)
33.75 T. Takahashi, K. Makishima, Y. Fukazawa, M. Kokubun, K. Nakazawa, M. Nomachi, H. Tajima, M. Tashiro, Y. Terada: New Astron. Rev. **48**, 309 (2004)
33.76 V.K. Krishna, V. Dutta: J. Appl. Phys. **96**, 3962 (2004)

34. II–VI Narrow Bandgap Semiconductors: Optoelectronics

Ian M. Baker

The main purpose of this chapter is to describe the applications and technology of II–VI narrow bandgap semiconductors, focussing on HgCdTe for infrared sensors. It provides a historical record of the development HgCdTe through photoconductors and photodiode arrays. The solid-state physics of HgCdTe is described to provide a foundation to explain the main crystal growth processes and device technologies. It concludes with a review of the research and development programs in centers around the world on third-generation infrared detector technology (so-called GEN III detectors). These include small pixel technology, higher operating temperature (HOT) device structures, two-color array technology, electron avalanche photodiodes (e-APDs), multifunctional HgCdTe detectors, retina level processing, and future trends for HgCdTe infrared detector arrays.

34.1 **Overview** 867
34.1.1 Introduction to Solar and Infrared Imaging 868
34.1.2 Historical Perspective and Early Detectors 868
34.1.3 HgCdTe Infrared Detector Market in the Modern Era 869
34.2 **Applications and Sensor Design** 870
34.3 **Photoconductive Detectors in HgCdTe** 871
34.4 **Sprite Detectors** 872
34.5 **Introduction to Photovoltaic Devices in HgCdTe** 873
34.5.1 Ideal Photovoltaic Devices 874
34.5.2 Nonideal Behavior in HgCdTe Diodes 875
34.5.3 Influence of Silicon on Photodiode Array Performance 875
34.5.4 Theoretical Foundations for HgCdTe Array Technology 876
34.6 **Manufacturing Technology for Photodiode Arrays** 879
34.6.1 Liquid Phase Epitaxial (LPE)-Based Processes 879
34.6.2 Vapour Phase Epitaxial (VPE)-Based Processes 882
34.7 **Advanced HgCdTe Technologies** 886
34.7.1 Small Pixel Technology 886
34.7.2 Higher Operating Temperature (HOT) Device Structures 886
34.7.3 Two-Color Array Technology 888
34.7.4 HgCdTe Electron Avalanche Photodiodes (e-APDs) 889
34.7.5 Multifunctional HgCdTe Detectors 891
34.7.6 Retina-Level Processing 892
34.7.7 Conclusions and Future Trends for Photovoltaic HgCdTe Arrays 892
34.8 **Emission Devices in II–VI Semiconductors** 892
34.9 **Potential for Reduced Dimensionality in HgTe–CdTe** 893
References 893

34.1 Overview

The main commercial application for narrow bandgap semiconductors is in infrared radiation detection. There are relatively few simple, elemental, or compound semiconductors with the correct forbidden energy gap to sense photons within the infrared spectrum, particularly at longer wavelengths. It is also beneficial to accurately match the spectrum to certain atmospheric windows, so the ability to tailor the wavelength is very important. A great deal of work has been carried out to synthesize crystals using an alloy of two different compounds with widely separate forbidden energy gaps (or bandgaps). II–VI compounds have crystal properties that make them very suitable for mixing, and they have a range of bandgaps from near zero for the semimetal Hg compounds to more than 1 eV.

However, among the wider bandgap II–VI semiconductors, only Cd, Zn, Mn, and Mg have been shown to open up the bandgap of the Hg-based semimetals HgTe and HgSe. By far the most developed alloy system is $Hg_{1-x}Cd_xTe$ (HgCdTe), a semiconductor formed from the semimetal HgTe and the wide bandgap semiconductor CdTe. By adjusting the alloy composition x, the

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_34

properties, including bandgap, can be varied smoothly between HgTe and CdTe. The advantages of this system arise from a number of features, notably: close lattice matching, high optical absorption coefficient, low-carrier generation rate, high electron mobility, and readily available doping techniques.

There has been some interest in developing alternative ternary alloys to replace HgCdTe since from theoretical considerations the already weak Hg–Te lattice bond is further destabilized by alloying with CdTe, and there is a potential for obtaining materials with increased hardness and detectors with better temperature stability. Most other compound combinations have crystal growth problems or doping limitations that make them unsuitable for device fabrication with the possible exception of $Hg_{1-x}Mn_xTe$, $Hg_{1-x}Zn_xTe$, and $Hg_{1-x}Cd_xSe$. Photoconductive detectors have been reported using $Hg_{1-x}Mn_xTe$, and photovoltaic detectors using both materials, but devices have not reached the maturity of $Hg_{1-x}Cd_xTe$. Also, the technology reported for these materials is very similar to that of HgCdTe, and so many of the detector design, technology, and performance factors described in this chapter for HgCdTe are also relevant to other II–VI alloys.

The main purpose of this chapter is to describe the applications, device physics, and technology of II–VI narrow bandgap devices, focussing on HgCdTe. It concludes with a review of the research and development programs in centers around the world on third-generation infrared detector technology (so-called GEN III detectors).

34.1.1 Introduction to Solar and Infrared Imaging

The Earth's atmosphere is conveniently transparent in the visible region, but it severely filters much of the radiation at longer wavelengths. Figure 34.1 shows the quantum spectra for solar and thermal radiation at the Earths' surface for specific conditions. The solar spectrum is scaled to a lux equivalent of 17 500 which is typical of a bright sunlit day. The thermal spectrum is the natural photon radiation emitted by all objects around us and depends on the temperature and the spectral emissivity of the materials in the scene. Passive thermal imaging is the term used to describe imaging of this radiation. Figure 34.1 shows the photon emittance of a body at a temperature of 20 °C peaking around 10 μm. After atmospheric absorption, the solar and infrared spectrum is divided into many windows which have names or letters as shown. The thermal infrared is particularly compromised by water vapor absorption, and there are only two windows used for imaging: the so-called medium wave infrared band (MWIR) between

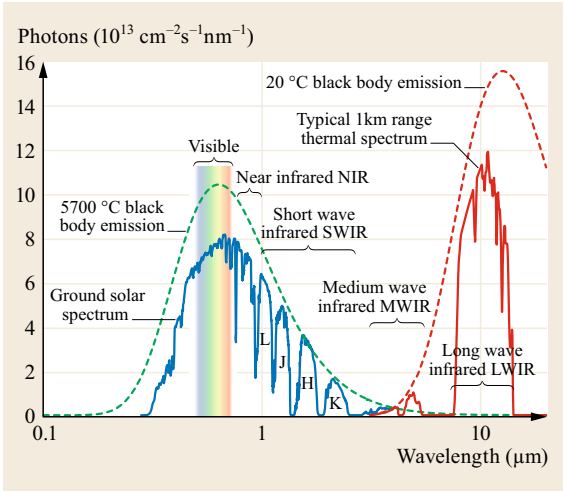


Fig. 34.1 Solar and thermal quantum spectra at the Earths' surface with associated band definitions. The solar data corresponds to a bright sunlit day of 17 500 lux

3.2 and 5.0 μm with a carbon dioxide notch around 4.3 μm and the long wave infrared band (LWIR) from 7.7 to 14 μm. The change in photon flux with temperature is small requiring sensitive detectors in contrast to visible sensors that need a wide dynamic range. HgCdTe is uniquely sensitive to the whole solar and thermal infrared range from 0.4 to 14 μm.

Figure 34.1 shows that the flux levels in the visible and thermal infrared are quite similar. In fact, the flux in the MWIR band is similar to the flux in the visible band at a scene illuminance of 250 lux corresponding to average room lighting. Infrared sensors can, therefore, exploit many of the focal plane technologies and signal processing developed for digital cameras.

34.1.2 Historical Perspective and Early Detectors

The first infrared photon detectors, based on thallous sulfide, were developed in the United States during World War I and were sensitive to about 1.4 μm. The next important developments occurred before and during World War II in Germany, with work on thin-film polycrystalline PbS devices with a response up to 2.5 μm, later extended further to the 3–5 μm region using PbSe and PbTe. During the 1950s and 1960s, indium antimonide (InSb) detectors emerged with a wavelength to 5.5 μm, and during the same period, impurity photoconductivity was studied in doped germanium and doped silicon. Germanium was favored at this time because it was available with fewer compensating impurities and gave better detector performance. Ge:Hg, with an impurity activation energy of 0.09 eV,

provided the first practical detector to be used in real-time thermal imaging, employing linear arrays for the 8–13 μm region. Because it required cooling to below 30 K extrinsic silicon devices have not found favor for terrestrial applications.

The next important phase of semiconductor infrared detector research took place during the late 1960s and 1970s, when research efforts were directed toward an intrinsic detector for the 8–13 μm band that would operate more conveniently at around 80 K. Two alloy semiconductors were developed, $Hg_{1-x}Cd_xTe$ (HgCdTe) and $Pb_{1-x}Sn_xTe$ (LTT). The first report of the synthesis of the semimetal HgTe and the wide-bandgap semiconductor CdTe to form the semiconductor alloy HgCdTe was published by the Royal Radar Establishment in Malvern, UK, by *Lawson* et al. [34.1]. This landmark paper reported both photoconductive and photovoltaic response at wavelengths extending out to 12 μm, and pointed out that this new alloy semiconductor showed promise for intrinsic infrared detectors. Soon thereafter many centers around the world switched detector development for major thermal-imaging programmes to the HgCdTe system. High-performance LW and MW linear arrays were produced for the first generation of thermal imaging equipment developed and manufactured in the 1970s through to the present. Many hundreds of thousands of detectors have been delivered.

34.1.3 HgCdTe Infrared Detector Market in the Modern Era

LWIR clearly has the most energy and was favored in older optically scanned systems where the dwell time was short and high flux was essential. Since the late 1990s, the LWIR market has been dominated by infrared detectors based on heat absorption. These include ferroelectrics, such as barium strontium titanate, or microbolometers, such as vanadium oxide (VOx) or amorphous silicon (α-Si). Uncooled cameras have an established volume market for relatively short-range imaging. They tend to need fast optics to achieve useful sensitivity and a typical focal ratio is F/1.2. For longer ranges, the diameter and weight of the germanium objective lens become an important issue. High-performance, cooled cameras based on HgCdTe have so much sensitivity in reserve that they can use focal ratios around F/4. It is straightforward and affordable to use long telephoto optical designs. They can also have fast frame rates with some cameras able to run at more than 1000 frames per second.

At first sight, the MW window looks compromised by the photon flux, some two orders lower than the LW window. However, it is much easier to make two-dimensional (2-D) detectors for this wavelength, and the photon flux disadvantage can be offset by staring architectures that enable a longer integration (or stare) time. The pixel size can be proportionately smaller with wavelength and typical sizes range from 8 to 15 μm. Array formats can be much larger than uncooled arrays, and HD 1024×768 and 1280×768 formats are used in the top-end cameras. High-performance thermal imaging market is currently shared between two material systems: indium antimonide (InSb), which has sensitivity to 5.5 μm, and HgCdTe. InSb has most of the cooled thermal imaging market because it was the earliest to be developed. HgCdTe has some key discriminators that may make it the material of choice for future *third-generation* developments. HgCdTe can be operated at much higher temperature than InSb so cameras can be smaller and have lower power consumption. HgCdTe can be grown in complex heterostructures with layers of different bandgap and doping for advanced sensors. The device technology is amenable to small pixels with small optical crosstalk. The dark current is relatively low in HgCdTe, providing opportunities in scientific applications and long-range imaging, especially in association with avalanche photodiode arrays.

The SW window is rarely used for passive thermal imaging because of low flux, but active imaging using a source, such as infrared illuminators, lasers, or even nightglow, is of continuing interest and it is the main band used in astronomy. Other materials can be used for SW detectors but products such as InGaAs tend to have high noise beyond 1.7 μm due to defects arising from the lattice mismatch with the InP substrate. HgCdTe has almost ideal solid-state properties for electron avalanche photodiode (e-APD) devices). The high mobility ratio of HgCdTe results in single-carrier multiplication with low noise figures, and such devices are exploited in photon-starved applications, such as long-range imaging and astronomy.

There are emerging technologies for the higher performance market based on quantum effects in III–V materials, such as type II superlattices, barrier photoconductors (so-called nBn devices), and multiple quantum well (MQW) detectors (usually using AlGaAs/GaAs technology). These systems are claimed to have higher ultimate performance than HgCdTe, better reliability, and lower manufacturing cost. Recent developments in wafer scale growth of HgCdTe, however, have set new standards. Also, the two drawbacks of lower quantum efficiency (QE) and the absence of a good surface passivation process (HgCdTe can use interdiffused CdTe) may limit how close these systems can get to the ultimate performance.

34.2 Applications and Sensor Design

Often the detector is the performance-limiting component in the camera, and it is necessary to use detectors with a sensitivity limited only by the random rate of arrival of photons from the scene (so-called background limited detectors, BLIP detectors). In a narrow gap semiconductor using both photoconductive and photovoltaic detectors, it is necessary to cool so that the thermal generation and associated excess noise are suppressed and the sensitivity becomes *BLIP limited*. The choice of cooler is, therefore, a key technology in an infrared detector, and the common options are described here. One of the points to emphasize is that the focal plane array is only one of a number of components that need to be optimized to preserve the BLIP performance. The optical design, cryogenics, and signal processing are also important to maintain the sensitivity of the focal plane array.

The means of cooling depends on the detector type and application. For instance photoconductive arrays often need a lot of cooling power, because of the Joule heating on the focal plane and the need for many bond wires that add to the thermal load. Two-dimensional arrays, on the other hand, often consume little power and have few connections, allowing alternative cooling methods. Other considerations are cooldown time, power, weight, space volume, robustness, and life.

Historically, the Joule–Thompson cooler was the first available method of cooling, relying on compressed gas bottles for power. The cooling power can be very high. Most first generation detectors use Joule–Thompson cooling. In more recent times, the high cooling power is exploited routinely in fast cooldown applications. Figure 34.2 illustrates a typical fast cooldown product, which can be cooled in a few seconds to liquid argon temperature (92 K) where BLIP performance is easily achieved.

Most modern thermal imaging applications use a cryocooler based on the Stirling cycle, which can achieve temperatures as low as 55 K but is normally regulated to a temperature in the range of 80–140 K. Figure 34.3 shows a typical integrated detector–cooler assembly (IDCA). The cut-away in Fig. 34.3 shows the front end of the detector in more detail. The infrared array is bonded directly on the cold finger of the cooler. To ensure only photons from the scene reach the de-

Fig. 34.2 Typical fast cooldown detector using Joule–Thompson cooling. (Courtesy of Leonardo MW Ltd)

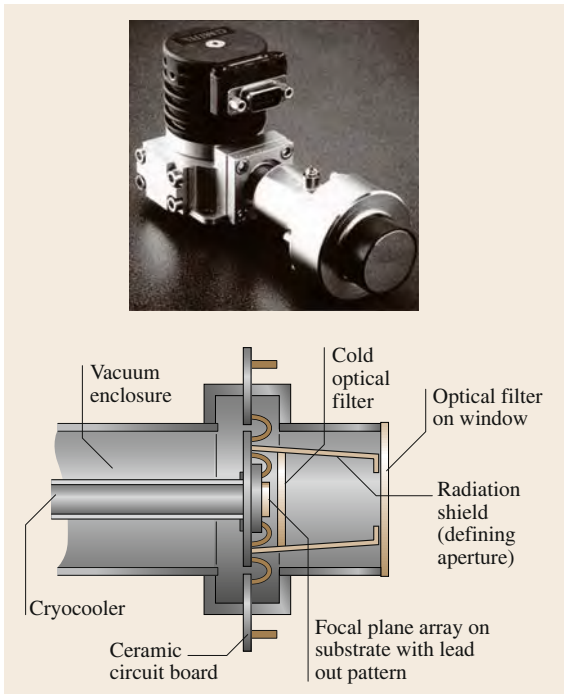


Fig. 34.3 Typical integrated detector–cooler assembly (IDCA). (Courtesy of Leonardo MW Ltd)

Fig. 34.4 Typical thermoelectrically cooled detector. (Courtesy of Leonardo MW Ltd)

tector, a cold stop or a radiation shield is bonded to the cold finger. The cold stop will have internal baffles and special coatings to stop internal reflections and is a vital component to achieve the BLIP operation of the array. The detector will have an optical pass-band filter that controls the upper and lower wavelengths. This will vary with the application, but typical bands are 3.2–4.2 μm, 3.7–5.0 μm, or 7.7–10.6 μm. Two methods are commonly used. The filter can be incorporated in the cold stop, and is, therefore, cooled, producing a very efficient block to stray radiation from other parts of the dewar. Almost as good performance can be achieved with filters on the front window, and a well-designed cold stop.

The third means of cooling is Peltier cooling (usually based on bismuth telluride devices). A typical detector is illustrated in Fig. 34.4. State-of-the-art coolers can achieve temperatures as low as 180 K, but in general they are not as efficient as Stirling cryocoolers, and the cooling performance can be poor in high ambient temperatures. Nevertheless, the main strengths of long-term reliability, greater shock resistance, and smaller weight and space volume have created a large volume market, especially for small photoconductor arrays. The temperature of 180 K is not quite cold enough for thermal imaging, and the applications tend to be those that can tolerate lower performance infrared detectors.

34.3 Photoconductive Detectors in HgCdTe

A photoconductor usually comprises a small slab or element of material with two contacts. The aim is to detect the change in resistance of the element when the photon flux is changed. The first HgCdTe devices were photoconductive because of the simplicity of the technology, and the relative ease of achieving near-ideal infrared performance and excellent reliability. Photoconductors were developed in the early 1970s when only bulk grown material (from a Bridgeman process or a solid-state recrystallization process) was available. HgCdTe quickly proved to be the best option and other systems including $Hg_{1-x}Cd_xSe$, $Hg_{1-x}Zn_xTe$, and $Hg_{1-x}Mn_xTe$ could not be developed to the same level. HgCdTe photoconductive detectors have been in routine production since the early 1980s and are often called first-generation detectors. They are the key component of the US Common Module Thermal Imager and, in the form of the SPRITE detector they feature in the UK Class II Common Module imager. Detectors for thermal imaging are most commonly fabricated with a peak response in the 8−12 μm atmospheric window region and are cooled to 80 K by means of a Joule–Thompson expansion cooler or, more recently, a Stirling engine cryocooler. They are also quite commonly made for operation in the 3−5 μm atmospheric window, and either use cooling to 80 K or employ Peltier (or thermoelectric) coolers to cool to around 200 K. Peltier-cooled MW detectors are used in many small, hand held cameras, and a whole host of heat sensing applications. The reliable performance, low levels of defects and easily understood physics has led to a long product life for photoconductive arrays. However, the array size is limited and first-generation thermal imaging systems need to employ complex optics to scan the infrared image over the array to build up a scene.

Early HgCdTe used bulk crystal growth techniques and the simple device geometry allows arrays to be manufactured by chemomechanical polishing of thick slices of material. Because the electron mobility is many orders higher than the hole mobility, it is advantageous to work with n-type material where the photoconductive gain is very high. N-type material with a low doping level can be manufactured directly, using the Bridgeman growth process, or can be produced by two-temperature annealing. Low doping is essential for long-carrier lifetime. N-type material can utilize anodic oxidation to passivate the surfaces and provide an accumulated surface that repels minority carriers, preventing them from interacting with surface recombination sites. Surface recombination velocity as low as 50 cm s^{-1} is routinely achieved. Also, crystalline defects in n-type material appear to have a weak effect on the carrier lifetime.

Ion milling is used to prepare the contact windows. The resulting highly doped n^+ layer provides an excellent ohmic contact for electron flow, and a blocking contact to minority carrier holes (so-called excluding contact). This leads to nonequilibrium effects. When the bias is applied, no holes can be injected into the material to sustain the flow of holes in the bulk away from the contact. As a consequence, both hole and electron densities drop to maintain space–charge neutrality. The dominant thermal generation process in n-type material is Auger 1, and the generation rate can drop by an order and can substantially increase the responsivity and detectivity [34.2].

Figure 34.5 shows a photomicrograph and scanning electron microscope view of a typical photoconductive element. It comprises a small slab of material, typically 50 μm square and 8 μm thick with two ohmic contacts.

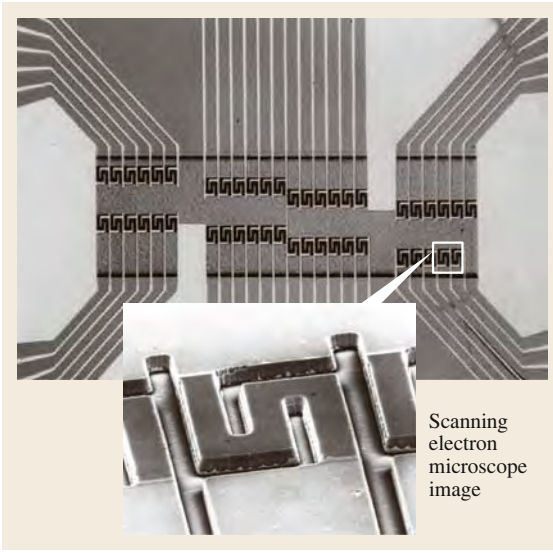


Fig. 34.5 Photomicrograph of 48-element photoconductive array using labyrinthed structures for enhanced responsivity. (Courtesy of Leonardo MW Ltd) ◄

The element illustrated has been *labyrinthed* to improve its performance. The device is operated using a constant current bias, supplied by means of a voltage source, and a series resistance, large compared to the detector resistance. The signal is measured as a voltage change across the detector, so that the device is essentially a radiation-sensitive resistor.

Despite its apparent simplicity, a rigorous analysis of photoconductor operation is quite complex [34.2, 3]. The analysis here is restricted to a one-dimensional treatment. *Kolodny* and *Kidron* [34.4] and *Oda* [34.5] have performed a 2-D and 3-D treatment.

Photoconductive HgCdTe detectors have been very successful in producing arrays of up to a few hundred elements for use in first-generation thermal imaging systems. The limitations of photoconductive detectors are apparent when very large focal plane arrays are required. The low impedance of the photoconductor makes it unsuitable for injecting into silicon charge transfer devices or field effect transistors; therefore, each element requires a lead-out through the vacuum encapsulation to an off-focal plane amplifier. The complexity of the dewar, therefore, limits the size of the array and the technology is not suitable for 2-D arrays.

34.4 Sprite Detectors

The SPRITE detector was invented in the early 1970s [34.6, 7] and it forms the basis of the UK Common Module detector, produced in large numbers. A thorough description of the device is reported by *Elliott* and *Jones* [34.8]. It is a very elegant method for amplifying the sensitivity of a photoconductive device. It is also a very compact device which enables highly efficient radiation shielding and economical use of material. SPRITE detectors are used in serial–parallel scan-imaging systems, because they perform a time-delay and signal integration function, in addition to a detector function. The operating principle of the device is illustrated in Fig 34.6. It consists of a strip of n-type HgCdTe, typically 700 μm long, 60 μm wide, and 10 μm thick, with three ohmic contacts. A constant-current circuit provides a bias through the two end contacts. The third contact is a potential probe for signal read-out.

There are two important points concerning the operation of the device. First, the bias field must be sufficiently high to ensure that excess minority carriers, generated along the filament length, can reach the negative end contact before recombination occurs. Second, the bias field must be chosen so that the ambipolar drift velocity, V_a, is equal to the image scan velocity. Consider now an image feature scanned along the filament, as illustrated in Fig. 34.6. The density of excess carriers increases during the scan as $[1-\exp(-x/V_a\tau)]$, where x is the distance from the positive bias contact along the scan direction, and τ is the excess-carrier life-

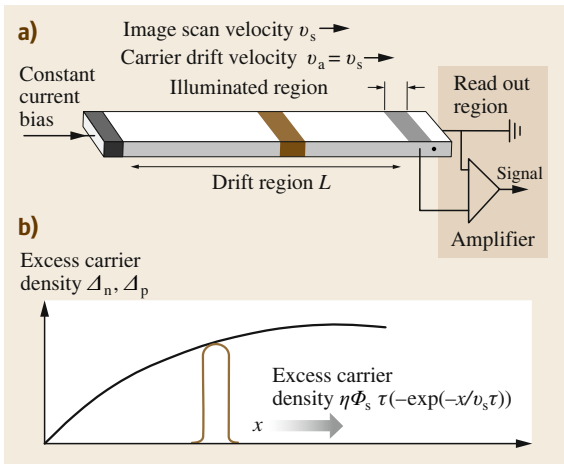


Fig. 34.6a,b Operating principle of a SPRITE detector

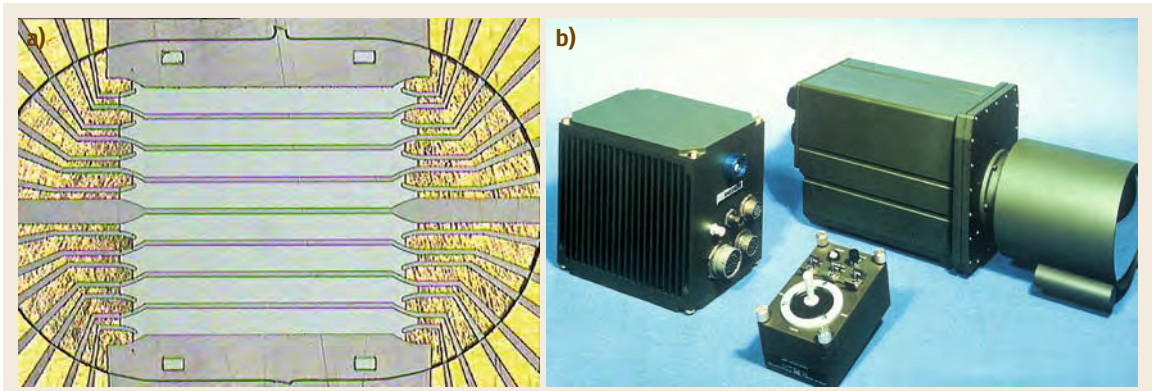


Fig. 34.7a,b UK Class II Common Modules Imager, **(a)** Horned OctoSPRITE **(b)** Camera modules. (Courtesy of Leonardo MW Ltd)

time. When the illuminated region enters the readout zone, the increased conductivity modulates the voltage on the contact and provides an output signal. Since the integration time of the signal flux, which is τ in a long element (or L/V_a in a shorter element), is greater than the dwell time on a conventional discrete element in a fast-scanned, serial system, a larger conductivity modulation and, hence, a larger output signal is observed. Therefore, the SPRITE detector continuously performs a time-delay and integration TDI function, and the signal-to-noise ratio is improved as a result of coherent integration of the signal and incoherent integration of the noise. The net gain in the signal-to-noise ratio is proportional to the square root of the integration time. For an amplifier-limited or a Johnson noise-limited detector, the gain is proportional to the integration time. Sprites can achieve a signal-to-noise improvement of a factor of 4 over a single photoconductor element.

The device depends on three aspects. The minority-carrier lifetime must be maximized to give the longest possible integration length, the speed of the carriers (holes) must match a practical image-scanning speed, and the thermal spread of the carriers must be minimized as this will blur the image. In practice, this implies the need for large minority-carrier lifetime, and low minority-carrier diffusion lengths. Lifetime values up to $\approx 10\,\mu s$ are obtained in alloy compositions suitable for 8–14 μm, 80 K operation, with N-type extrinsic carrier density $\approx 10^{14}\,cm^{-3}$. Recombination is principally by the Auger 1 process in these compositions. Lifetime values of 15–30 μs are obtained in compositions used for 3–5 μm operation at thermoelectric temperatures. The hole mobility is low in HgCdTe and results in short diffusion lengths even with long lifetimes, yielding a spatial resolution in the 8–14 μm band SPRITE devices of $\approx 50\,\mu m$ and $\approx 100\,\mu m$ in 3–5 μm devices if the full integration is used. Frequency response into the MHz region is achieved.

SPRITE detectors have been commonly used in 8-, 16-, and 24-row arrays and are commercially available in these forms. Larger Sprite arrays are limited by Joule heating in the elements which increases the heat load of the cooler to impractical levels. The Horned OctoSPRITE was the final development used in the UK Class II Common Module as shown in Fig. 34.7.

For comprehensive reviews, see [34.2, 9–11].

34.5 Introduction to Photovoltaic Devices in HgCdTe

Photovoltaic arrays have inherently low power consumption and can be easily connected to a silicon integrated circuit to produce a retina-like focal plane array (FPA). Photovoltaic devices emerged in the early 1980s and tended to fall into two categories: small matrix arrays mostly in the MW (3–5 μm) and long linear arrays mostly in the LW band.

Long linear arrays were first developed for systems with just one dimension of scanning. Because the image is scanned across the array very quickly, the dwell time is as short as a few tens of microseconds, and it is essential to use LW detectors with high QE and near BLIP-limited performance. HgCdTe became quickly established as the sensor of choice. Development work in the 1980s and 1990s in the United States and Europe has led to a family of detectors for LW thermal imaging, such as the US SADA II system based on a 480×6 format or the UK STAIRS *C* system based on a 768×8

array format (producing a high definition 1250×768 image). In the detector examples quoted, there are six or eight rows of diodes, and the signal from each row is delayed in time and added to provide an enhancement. This is called TDI. An important advantage of this scheme is that defective pixels can be deselected, so typically the STAIRS system will only use the best six of the eight diodes in one channel. Consequently, the detector performance can be relatively insensitive to defects in the HgCdTe and this is the reason why LW long linear arrays have reached maturity much earlier than LW matrix arrays.

Long linear arrays can also use the MW band, but the flux levels are over an order less than the LW band and the performance is compromised by the short integration time in a scanned system. MW detectors then tend to use matrix arrays where the integration time can approach the frame time. MW detectors have now reached maturity in large 2-D arrays with near-ideal performance and few pixel defects, and they produce remarkable sensitivities down to 10 mK. The MW band thermal imaging market ranges from small arrays using half-TV (320×256 or 384×288) or full-TV formats (640×480 or 640×512) to high-definition formats. The market is divided between indium antimonide (InSb) and HgCdTe. Broadly, manufacturing companies from the United States and Israel tend to prefer InSb and European, and some US companies have specialized in HgCdTe. Figure 34.8 shows a representative image from a state-of-the-art HgCdTe 2-D array. Larger arrays are produced for science applications, such as astronomy, and up to $4\,\mathrm{K} \times 4\,\mathrm{K}$ arrays are under development [34.12].

Unlike photoconductive arrays, manufacturers use a variety of different technologies for photovoltaic arrays. Sections 34.5.1–34.5.4 describe the theory of photovoltaic detectors and the fundamental principles behind the HgCdTe device technology. These principles help to explain the various approaches used by manufacturers to produce detectors in both the LW and MW bands described in Sect. 34.6.

34.5.1 Ideal Photovoltaic Devices

The current–voltage characteristic for an ideal diffusion-limited diode exposed to a photon flux φ_λ is given by

$$I_\mathrm{d} = I_\mathrm{s}\left[\exp\left(\frac{qV_\mathrm{d}}{k_\mathrm{B}T}\right) - 1\right] - qA\eta\phi_\lambda \,, \tag{34.1}$$

where I_d is the diode current, I_s is the diffusion current, V_d is the diode bias voltage (taking negative values in reverse-bias), η is the QE, and A is the detector area. The second term in (34.1) is the photocurrent, I_ph. The QE is given by

$$\eta = (1-R)\left[1 - \exp\left(-\alpha_\lambda t\right)\right] F \,, \tag{34.2}$$

where R is the reflectivity of the front surface, normally minimized by an antireflection coating, α_λ is the absorption coefficient, t is the optical path length through the absorber, typically around $5-10\,\mu\mathrm{m}$ to give adequate absorption in HgCdTe detectors, and F is a geometry factor that describes the number of photogenerated carriers within the pixel which reach the junction before recombining.

The condition that the detector be background limited is crucial for high-performance systems and to meet this condition the internal thermal generation must be much less than the photon generation in the lowest flux case, that is, $I_\mathrm{s} \ll I_\mathrm{ph}$. By differentiating (34.1), I_s is given by $k_\mathrm{B}T/qR_0$, where R_0 is the zero bias resistance.

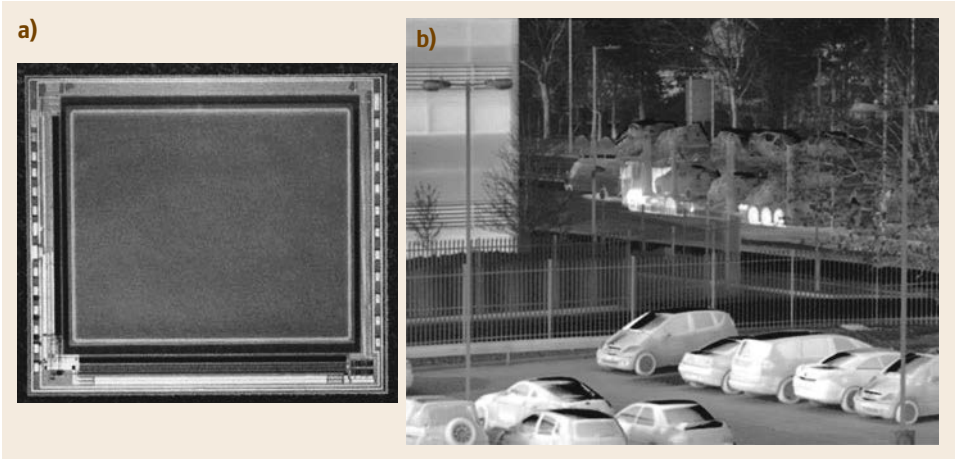


Fig. 34.8a,b Typical focal plane array **(a)** and thermal image **(b)**. (Courtesy of Leonardo MW Ltd)

The condition for background-limited performance is, therefore, often written as $R_0 \gg k_B T/qI_{ph}$, and this is the origin of the commonly used figure of merit, R_0A. As an example, in the case of an LW band system, say 10 μm cut-off, with an F2 optic and a lowest background scene temperature of −40 °C, an R_0A of greater than 30 Ω cm^2 is required. Detectors need to be cooled sufficiently to suppress thermally generated currents, but there are always pressures to avoid very low temperatures to give savings in power and cooldown time and better engine reliability. This is particularly the case for cut-off wavelengths longer than 10 μm, and detector technologies have emerged that provide high R_0A as a first priority.

The white noise current at low frequencies is given by

$$i_n^2 = 2qI_s\left[1+\exp\left(\frac{qV_d}{k_BT}\right)\right]B + 2q\,(qA\eta\phi_\lambda)\,B\,, \tag{34.3}$$

where B is the bandwidth in hertz. In the absence of photocurrent, the noise at the zero bias is equal to the Johnson noise, $4k_BTB/R_0$. In reverse bias, it tends to the normal expression for shot noise, $2q(I_s + I_{ph})B$. Note that the mean-square shot noise in the reverse bias is half that of the Johnson noise at the zero bias.

For photoconductor arrays, the established figure of merit is detectivity, D^*. For photodiode arrays, the photocurrent is integrated on a capacitor, so a better figure of merit is noise-equivalent temperature difference (NETD) [34.13]. For an integrated charge Q_{int} (comprising photocurrent, dark current, and stray photons), the minimum variance is the square root of Q_{int}. To derive the equation for NETD, it is useful to introduce the *thermal contrast ratio*, c. This is the proportional increase in photon flux per degree change in blackbody temperature. Typical values for c are 4 %/K in the MWIR and 2 %/K in the LWIR.

In terms of integrated charge,

$$cQ_{ph} = \frac{dQ_{ph}}{dT} \quad \text{or} \quad dT = \frac{dQ_{ph}}{cQ_{ph}} = \text{NETD}$$

if dQ_{ph} is replaced by the noise rtQ_{int}. The expression for NETD is shown in the following equation

$$\text{NETD} = \frac{1}{cQ_{ph}}\sqrt{Q_{int}}\,. \tag{34.4}$$

Note that in the ideal BLIP case, $Q_{int} = Q_{ph}$, and the NETD is given by the simple expression

$$\text{NETD} = \frac{1}{c\sqrt{Q_{ph}}}\,. \tag{34.5}$$

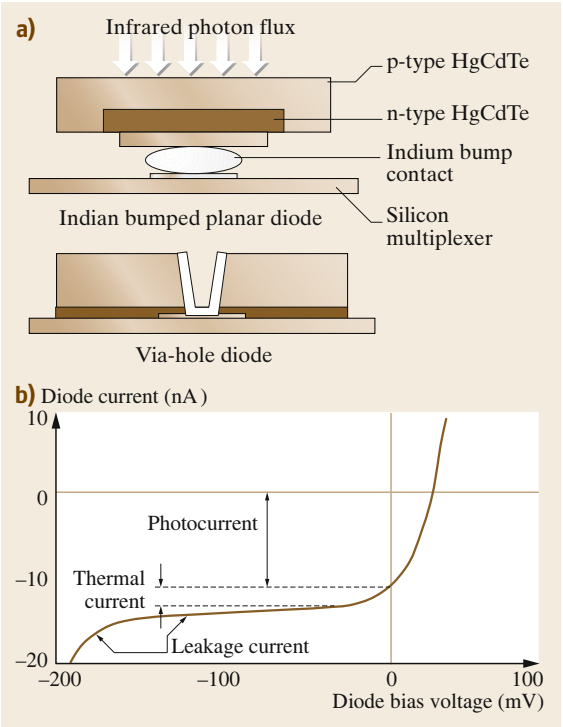


Fig. 34.9a,b Types of HgCdTe diodes **(a)** and a typical current–voltage characteristic for an LW diode showing main nonideal features **(b)**

34.5.2 Nonideal Behavior in HgCdTe Diodes

At longer wavelengths or higher operating temperatures, a number of leakage currents can impact the I–V characteristics and degrade the performance of the detector. Figure 34.9 illustrates a typical I–V characteristic for an LW diode, along with two types of HgCdTe photodiodes, a planar diode and a via-hole diode. The diagram illustrates how factors such as thermally generated current, shunt resistance, and breakdown currents at a high bias can affect the diode behavior. The performance can be degraded directly, for example, extra currents produce white noise which may be significant if the leakage currents are of the same order as the photocurrent. In addition, some leakage current mechanisms are associated with trapping or tunneling and often produce excess, low frequency noise (often called $1/f$ noise for the typical trend with frequency).

34.5.3 Influence of Silicon on Photodiode Array Performance

Additional constraints arise from the need to inject photocurrent into a silicon-integrated circuit often called an ROIC for readout integrated circuit. The simplest

method is called direct injection and literally means injection of the photocurrent into the source of a common gate MOSFET. A capacitor on the drain performs the integration. Good injection efficiency is only achieved if the diode has a dynamic resistance higher than the input impedance of the MOSFET. Unless the dynamic impedance of the diode is high enough, the signal can be attenuated, additional noise added, and nonlinearity generated in the response. Achieving a high dynamic resistance has been a driver in the development of most infrared technologies. In MW arrays with cut-offs up to 6 μm, high injection efficiencies are seen routinely because the photodiodes are near-ideal. In the LW band, there will be a limit to the upper cut-off wavelength and/or flux levels that routinely allow high injection efficiencies using direct injection.

The input impedance of a MOSFET is a function of the source–drain current (in this case, the total diode current) and is usually expressed in terms of the transconductance, g_m, given by $qI_d/(nk_BT)$ for low injected currents (n is an ideality factor that can vary with technology usually in the range of 1−2). The expression for injection efficiency, ε, is approximately given by

$$\varepsilon = \frac{R_v I_d}{R_v I_d + \dfrac{nk_BT}{q}} , \tag{34.6}$$

where R_v is the dynamic impedance of the diode and I_d is the total injected diode current equal to I_{ph} in the background-limited case.

In the background-limited case, (34.6) becomes a simple function of $I_{ph}R_v$. For a system requiring an efficiency of 99.9%, the value for $I_{ph}R_v$ approaches 10 V and in systems with slow optics and low background temperatures, the requirement for R_v can exceed $10^9\,\Omega$. There are more complex injection circuits that effectively reduce the input impedance and allow lower HgCdTe resistances to be used. Good comprehensive summaries are presented by *Kozlowski* [34.14] and *Kozlowski* et al. [34.15]. These buffer amplifier circuits may not be compatible with small pixels and consume more power but for some low photon flux applications offer a practical solution.

The ROIC can have a controlling effect on the sensitivity of 2-D arrays. Both charge coupled devices (CCDs) and standard foundry complementary metal-oxide-semiconductor (CMOS) devices have been used for multiplexing infrared 2-D arrays, but CMOS is now the preferred choice and CMOS circuits will operate well at low temperatures. The function of the multiplexer is to integrate the photocurrent from the photodiode array, perform a limited amount of signal processing, and sequentially readout the signals. As shown in Sect. 34.5.1, the sensitivity depends on the number of electrons integrated and this requires the integration capacitance in each pixel to be fairly high. Normally, the capacitor is created using a MOSFET structure with a thin gate oxide dielectric and capacitance values approaching 1 pF are possible producing NETD values as low as 10 mK. It is important to note that this NETD limitation has nothing to do with the HgCdTe, but it only depends on the pixel design. Furthermore, as the pixel size is reduced, the capacitance is decreased disproportionately, so the NETD cannot always be maintained in small pitch arrays.

34.5.4 Theoretical Foundations for HgCdTe Array Technology

Good detailed summaries of photodiode fundamentals have been provided by *Reine* et al. [34.16, 17] and the reference lists contained therein. The purpose of this section is to highlight the key issues that control the thermal generation, leakage currents, and QE in detectors. From this analysis, it is easier to understand the measures taken by detector manufacturers to make practical high-performance detectors.

Thermal Diffusion Currents in HgCdTe

When minority-carrier generation occurs due to a thermal process within a diffusion length from the p-n junction, it gives rise to a current that cannot be distinguished from the photocurrent. This is called thermal diffusion current and has two commonly reported generation mechanisms: Auger or Shockley–Read–Hall (S-R-H).

Auger recombination in HgCdTe is a phenomenon involving the interaction of three carriers. For instance, Auger 1 generation in n-type material is due to the interaction of two electrons and a hole and is dependent on the square of the carrier concentration. Fortunately, it is relatively easy to produce low-carrier concentrations in n-type HgCdTe ($< 5\times10^{14}\,\mathrm{cm}^{-3}$), and where there is a choice, this will be selected as the absorber type. In p-type material, Auger recombination involves two holes and an electron, and is referred to as Auger 7. It is well reported that Auger 7 lifetimes are an order greater than Auger 1, so in p-type material, the lifetime will transition from Auger 7 to Auger 1 at higher temperature. An empirical relationship for the minimum Auger generated dark current has been proposed called *Rule 07* by *Tennant* et al. [34.18, 19] and it is applied across a wide parameter space. This is often used as a baseline for HgCdTe devices when judging newer infrared technologies.

In p-type material, dark current generation through trap levels often dominates Auger 7. The reason may be

that crystal defects such as dislocations tend to decorate n-type and create local stress fields exposing trap levels. The lifetime in p-type HgCdTe is reported by *Lacklison* and *Capper* [34.20], *Polla* et al. [34.21], and *Radford* et al. [34.22] to show an inverse, linear dependence on doping, and this is attributed to the S-R-H proces. S-R-H recombination has been associated with the Hg vacancy and can be modeled by a strong donor level located $\approx 30\,\mathrm{mV}$ from the conduction band, which appears to be independent of composition in the $x = 0.2$–0.3 range. The density of these donor S-R centers is said to be proportional to the Hg vacancy concentration, but lower by a factor of ≈ 20. If Hg vacancies are replaced by acceptor atoms, such as Cu, Ag, Na, or Au, the lifetime can be increased by over an order. The use of extrinsic doping is an effective way to reduce thermal currents from the p-side in homojunction devices. A summary of the lifetime limiting mechanisms in "n"-type and "p"-type material as a function of composition, temperature, and doping level is given by *Kinch* [34.23].

Expressions for the diffusion current in photovoltaic devices have been derived by *Reine* et al. [34.16]. The fundamental R_0A expression is given by

$$R_0A\,(n-\mathrm{side}) = \frac{N_\mathrm{d}}{qn_\mathrm{i}^2}\sqrt{\frac{k_\mathrm{B}T\tau_\mathrm{h}}{q\mu_\mathrm{h}}}\,, \tag{34.7}$$

where N_d is the net donor concentration on the n-side, n_i is the intrinsic carrier concentration, which dominates the temperature dependence, and τ_h and μ_h are the minority carrier lifetime and mobility, respectively.

A similar expression describes the contribution of the p-side. In practice, the diffusion length, L, in normally doped HgCdTe is often larger than the typical absorber thickness of 3–5 μm. In this case, the volume available for the generation of diffusion current is restricted, and a suitable modified expression is

$$R_0A\,(n-\mathrm{side}) = \frac{k_\mathrm{B}TN_\mathrm{d}\tau_\mathrm{h}}{q^2n_\mathrm{i}^2t_\mathrm{n}}\,, \tag{34.8}$$

where t_n is the thickness of the n-type material.

In the case of L being greater than t_n, the surfaces and contacts can act as sources of extra diffusion current if the surface recombination velocity is greater than the diffusion velocity, D/L. It is essential to ensure properly passivated surfaces and for contacts employ a minority carrier barrier, such as a higher doped or wider bandgap layer under the metal contact.

In order to engineer a detector with low thermal diffusion current (or high R_0A), it is important to recognize that it is only necessary to use one side of the junction for collecting photocurrent; the other side, in principle, can be made wider bandgap, thereby minimising the thermal current contribution. Devices with layers of different bandgaps are called heterostructures. The most common design is to use a wider bandgap on the p-side to reduce the thermal diffusion current and use the n-side as an absorber taking advantage of long minority carrier diffusion lengths to maximise the QE.

Thermal Generation Through Traps in the Depletion Region

The thermal generation rate, g_dep, within the depletion region via traps, is given by the usual Shockley–Read–Hall expression

$$g_\mathrm{dep} = n_\mathrm{i}^2\left(n_1\tau_\mathrm{n0} + p_1\tau_\mathrm{p0}\right)^{-1}\,, \tag{34.9}$$

where n_1 and p_1 are the electron and hole concentrations, respectively, which would be obtained if the Fermi energy was at the trap energy, and τ_n0 and τ_p0 are the lifetimes in the strongly n- and p-type regions.

The leakage current is often known as the g–r current. Normally, one of the terms in the denominator of (34.9) will dominate, and for the case of a trap at the intrinsic level, n_1 and $p_1 = n_\mathrm{i}$, giving $g_\mathrm{dep} = n_\mathrm{i}/\tau$. It is the weaker dependence on n_i, and, therefore, on temperature, which distinguishes generation within the depletion layer from thermal diffusion current.

Where the depletion region intercepts the surface, there is often enhanced generation due to the presence of a high density of interface states. This can be exacerbated if the surface passivation is not properly optimized to give a flat band potential at the junction. In an extreme case, the surface on one side of the junction may become inverted, creating an extension of the depletion layer along the surface, and leading to high-generation currents. A practical solution to this mechanism is to widen the bandgap in the material where the junction intercepts the surface, the so-called heteropassivation. This can be achieved by using a thin film of CdTe together with a low-temperature anneal [34.24], and this is the commonest passivation technique used by manufacturers.

Interband Tunneling

Due to the very low effective mass of the electron in HgCdTe, direct band-to-band tunneling can occur from filled states in the valence band to empty states at the same energy in the conduction band. An expression for the current due to this process has been developed by *Anderson* [34.25]. The tunneling current increases very rapidly as the applied voltage or doping is increased, or the bandgap or temperature decreased. It is the normal

modern practice to use low doping on one side of the junction to minimize interband tunneling under normal operating conditions; nevertheless, at low temperature or for very long wavelength devices, interband tunneling can become dominant.

Trap-Assisted Tunneling

Trap-assisted tunneling is generally accepted to be one of the main causes of leakage current and excess noise in LW diode arrays, but it is not a fundamental limitation, often being associated with impurities or structural defects within the depletion region. The definition of a trap is a center with a capture coefficient for minority carriers many times larger than that for majority carriers. (Otherwise, it is a generation center.) The role played by traps in the depletion region is very complicated allowing for three possible two-step processes: thermal–tunnel, tunnel–thermal, and tunnel–tunnel. The nonthermal step can include tunneling of electrons from the valence band to traps, and tunneling from the traps to the conduction band.

The formulation for trap-assisted tunneling (TAT) for HgCdTe was developed by *Wong* [34.26] and by *Anderson* and *Hoffman* [34.27], based upon original work by *Sah* [34.28] in silicon. *DeWames* et al. [34.29], *Nemirovsky* et al. [34.30], and *Rosenfeld* and *Bahir* [34.31] have shown that models based on a thermal-tunnel process can explain the bias-dependent behavior of the reverse-bias current in long wavelength diodes.

Nemirovsky et al. [34.32] have shown that many of the observed properties of their n^+-p diodes can be explained by a TAT process via a thermally excited, bulk S-R-H center, modeled for the special case of the trap residing at the Fermi level. The physical picture for this assumption is that there is a uniform distribution of trapping centers throughout the bandgap, and the barrier for tunneling is lowest at the uppermost center that is still occupied. Hence, the occupied trapping center that coincides with the Fermi level has the highest transition probability and plays the dominant role in the thermal TAT process. This model helps to explain several unusual properties, such as the observed weak dependence of the leakage current on temperature and the doping level. Note that this is in contrast to direct band-to-band tunneling. Also, there is a much weaker dependence on diode bias voltage. The model also predicts the commonly observed *ohmic* region illustrated in Fig. 34.9 at a bias around −100 mV.

The physical origin of the trap has not been established yet but it appears to be an acceptor-type impurity or defect within the depletion region. The population of such traps is likely to be dependent on the HgCdTe material and the junction-forming technology, and so the quality of LW diodes is highly process specific.

Impact Ionization

Underlying TAT a more fundamental source of leakage current has been proposed which is called impact ionization. The effect arises because in HgCdTe, the electron scattering mechanisms tend to be weak and hot electrons can penetrate deep into the conduction band where they readily avalanche. Leakage current arises because extra electron–hole pairs are created within the depletion region due to impact ionization by minority-carrier electrons from the p-side. The leakage mechanism has been confirmed to have a linear relationship with optically injected minority carriers over a wide temperature range [34.33, 34]. To suppress this source of leakage current, the n-side doping must be very low [34.35]. Routinely achieving low-carrier concentrations ($< 5 \times 10^{14}\,\mathrm{cm}^{-3}$) is an important aim for homojunction technology. Most manufacturers introduce a donor, such as indium, to the crystal to control the n-type level to around $3-5 \times 10^{14}\,\mathrm{cm}^{-3}$. The carrier concentration in the p-region has a second-order effect compared with that of the n-region, for the range of concentrations normally used.

Photocurrent and Quantum Efficiency

HgCdTe has a strong optical absorption coefficient, and only thin absorber layers are needed to produce high QE. Most device structures have a back-reflector which enables two optical passes, so typically in MW detectors, the absorber need only be $2-3\,\mu\mathrm{m}$ thick and about twice this in LW detectors. Ideally, the absorption should occur well within a diffusion length of the p-n junction to avoid signal loss due to recombination. Potential recombination mechanisms are radiative, Auger, and S-R-H. Photons generated in radiative recombination may be reabsorbed in a process called photon recycling. In many device structures, this effectively extends the lifetime and radiative recombination is negligible, but it also poses a risk for crosstalk, especially in two-color devices. A long-carrier lifetime is nearly always observed in n-type material with low-carrier concentration. Device engineers tend to favor using n-type absorbers for the best QE and try to minimize the volume of the p-region for lower thermal leakage currents.

An interesting idea for improving QE and dark current is to perforate the absorber with tube-like holes to create a so-called photon-trapping structure (PTS). *Schuster* and *Bellotti* [34.36] summarise the status.

Excess Noise Sources in HgCdTe Diodes

For most applications, the detector performance depends on operating the diode in a small reverse bias where the dynamic resistance is at a maximum. It is then necessary to minimize extraneous leakage currents. The control of these leakage currents and the

associated low frequency noise is, therefore, of crucial interest to the device engineer.

In the field of thermal imaging, HgCdTe FPAs can now be manufactured with very few non-BLIP pixels over a wide spectral range from 1.0 to 12 μm. Most manufacturers use interdiffused CdTe as a passivant and have eliminated surface effects which currently is the weakness of many emerging III–V technologies. At higher operating temperature or very long wavelength, it becomes more difficult to suppress the effect of crystal defects and so it is important to summarise the research.

Firstly, there is a clear distinction between early research on LPE technologies grown on (111) CdZnTe and later work on molecular beam epitaxy (MBE) or metal organic vapor phase epitaxy (MOVPE). The reverse-bias characteristics and $1/f$ noise of HgCdTe diodes made in LPE are well known to depend strongly on dislocations intercepting the junction as revealed by defect etches. Threading dislocations, emanating from the substrate and possibly decorated during growth, are the most potent forms (for instance [34.34] and the references therein) giving excess dark current and noise, often of a bimodal nature. Some authors [34.37, 38] report that the dark current and noise mechanism through dislocations are TAT. Earlier many authors reported an empirical relationship between TAT leakage current and $1/f$ noise with a trend roughly given as $\alpha i_{\mathrm{L}}^{\beta}$, where α and β are device-dependant variables with β taking a value close to 0.5 (for instance [34.39]).

Some device structures have lateral collection architectures such as via-hole and planar devices. Here there is a possibility for n-type dislocation pipes to extend the depletion layer into the P-region. This has been discussed by *Kinch* et al. [34.40] and *Strong* et al. [34.41]. The n-type pipe can be viewed as a one dimensional MOSFET whose conductivity can be affected by nearby trapping events. In this way, a single trap can modulate the absorber volume and produce fluctuations in dark current, photocurrent and junction capacitance. A slow trap will result in a random telegraph signal spectrum (RTS) and a fast trap a $1/f$ noise spectrum. This type of noise source is reduced in VPE processes because the absorber volume is usually less than the diffusion length.

The current trend is to move toward lower temperature growth processes and growth on substrates with poor lattice match and this has refocused attention on grown-in dislocations. Also, the modern trend to operate infrared detectors at higher temperature (to reduce camera power) exaggerates the effect. MOVPE material on (100) orientation shows a faint pattern of etch pits suggesting no threading-type dislocations but a background of weak mosaic crystal boundaries [34.42]. It is apparent that the type of crystal defect affects the diode performance in different ways. Dislocations in the junction would be expected to contribute significant TAT current, but in VPE devices, the junction is routinely placed up the composition grade into a wider bandgap area and the thermal step in the TAT process is switched off. It is apparent then that VPE processes are much more tolerant of crystal defects due to bandgap engineering and the ability to control the absorber thickness.

34.6 Manufacturing Technology for Photodiode Arrays

There are four basic photodiode processes used commercially at present, two based on liquid phase epitaxy (LPE) growth and two on vapor phase epitaxy (VPE) growth. This section describes the processes but does not attempt to compare them as they are all under intense development toward smaller pixel sizes and higher operating temperature. All the processes are developed at least to 10 μm pixel size and manufacturers have different strategies to minimize crosstalk.

34.6.1 Liquid Phase Epitaxial (LPE)-Based Processes

LPE of HgCdTe at present provides the lowest crystal defect levels, and very good short and long-range uniformity. *Capper* et al. [34.43] has produced a detailed summary of the field. LPE layers are grown using an isothermal super saturation or programmed cooling technique or some combination. A detailed knowledge of the solid–liquid–vapor phase relation is essential to control the growth particularly in view of the high Hg pressure. Challenges include: the compositional uniformity through the layer, the surface morphology, the incorporation of dopants, and the specifications for thickness, wavelength, etc. A common component leading to high structural quality is the use of lattice-matched substrates of CdZnTe. These are grown by bulk processes, such as horizontal Bridgman, and can supply layers as large as 60 cm^2. The CdZnTe substrates must be of the highest quality and often this is a significant cost driver for the process.

Two different technical approaches are used: growth from Hg-rich solutions and growth from Te-rich solutions. Advantages of the Hg-rich route include: excel-

lent surface morphology, a low liquidus temperature, which makes cap layer growth more feasible and the ease of incorporation of dopants. Also, large melts can provide for very good compositional and thickness uniformity in large layers and give consistent growth characteristics over a long period of time. Growth from Te-rich solutions can use three techniques to wipe the melt onto the substrate: dipping, tipping, and sliding. A sliding boat uses small melt volumes and is very flexible for changing composition, thickness, and doping. Tipping and dipping can be scaled up easily and can provide thick, uniform layers, but the large melts limit flexibility. Double layers are also more difficult to grow.

Most manufacturers have taken their chosen growth system and tailored it to provide optimum material for their device technology. In particular, the use of dopants and the deliberate introduction of compositional grades are very specific to the device structure. A crucial figure of merit, however, is the dislocation count that controls the number of defects in 2-D arrays. Etch pit densities of $3-7\times10^4\,\mathrm{cm}^{-2}$ are typically seen in the Te-rich sliding boat process reproducing the substructure of the CdZnTe substrate [34.44]. The etch pits are associated with threading dislocations which appear to be normal or near normal to the layer surface. The substrate defect level can be as low as mid-$10^3\,\mathrm{cm}^{-2}$ in some horizontal Bridgman CdZnTe, but this is not easy to reproduce. Similar defect densities are found in CdTeSe material, but the impurity levels have proved difficult to control in the past. Device processing, therefore, must expect to cope with defect levels in the mid-$10^4\,\mathrm{cm}^{-2}$ range for routine CdHgTe epilayers and this will set the ultimate limit on the number of defects in HgCdTe 2-D arrays. Several groups have used the LPE process on low cost substrates including: CdZnTe or CdTe on GaAs/Si wafers [34.45] or the PACE technology on sapphire described in detail in Sect. 34.6.2, *Wafer Scale Processes Using Low Cost Substrates*.

Junction-Forming Techniques

For n–p devices, crystal growers can obtain the desired p-type level by controlling the density of acceptor-like mercury vacancies within a carrier concentration range of, say, 10^{16}–$10^{17}\,\mathrm{cm}^{-3}$. Neutralizing the Hg vacancies and relying on a background level of donors to give the n-type conversion create the photodiode junctions. Mercury can be introduced by thermal diffusion from a variety of sources but high temperature processes are not very compatible with HgCdTe at the device level. However, type conversion can be readily achieved by processes such as: ion beam milling, as described by *Blackman* et al. [34.46], and ion implantation, as described by *Margalit* et al. [34.47], *Kolodny* and *Kidron* [34.48], *Bubulac* et al. [34.49] and *Bubulac* and *Viswanathan* [34.50], and *Syz* et al. [34.51]. *Kinch* [34.52] also reports type conversion using plasma-enhanced milling in the VIAP process and *White* et al. [34.53] has similar observations for H_2/CH_4 plasmas. The common feature is that the conversion depth is much deeper than would be expected from the implantation range alone.

The current knowledge on type conversion using ion beam milling is described by *Baker* and *Maxey* [34.34], and the current knowledge for ion implantation is summarized by *Bubulac* and *Viswanathan* [34.50]. The explanation for the behavior of HgCdTe under ion beam bombardment involves a number of physical mechanisms. Firstly, the low binding energies, ionic bond nature, and open lattice of HgCdTe facilitate the liberation of free mercury at the surface and subsequent injection by the ion beam. The injection mechanism probably involves a recoil implantation process. Once the Hg interstitial is injected, the mobility is apparently extremely high, and there is some evidence that this is stimulated by the ion beam in a process related to the anomalously high diffusion rates of impurities often observed in SIMS analysis. Another factor is the movement of dislocations under the influence of the ion beam and possibly stimulated by strain fields. A number of workers report that the n-type carrier concentration in the converted region is very low and in fact, fast-diffusing impurities, such as Group IB elements Cu, Ag, and Au, and Group IA elements Na, K, and Li, which reside on the metal sublattice, are swept out of the n^- region of the diode by the flux of Hg interstitials. This impurity sweep-out effect is serendipitous because it means that in ion-beam-generated junctions, the n-regions are very pure, free of S-R centers, and have weak doping, creating an ideal structure for high-performance detectors. The detailed atomic level processes taking place, including the role of mercury interstitials, dislocations, and ion bombardment in the junction-forming process, are complex and not well understood in detail. Despite the complex physics involved, manufacturers have achieved good phenomenological control of the junction depth and n-dopant profiles with a variety of processes.

Via-Hole Technologies Using LPE

Via-hole devices share many of the process stages of photoconductive arrays and have been commercially produced since 1980. Arrays generated using this process are known as loophole devices [34.54] and HDVIP devices [34.52]. By using a thin monolith of HgCdTe bonded rigidly to the silicon, the thermal expansion mismatch problem is overcome because the strain is

taken up elastically. This makes the devices mechanically and electrically robust and suitable for very long linear arrays up to 16 mm in length [34.44].

The process flow of the via-hole device using LPE material is illustrated in Fig. 34.10. The process has two simple masking stages. The first defines a photoresist film with a matrix of holes of, say, 5 µm diameter. Using ion beam milling, the HgCdTe is eroded away in the holes until the contact pads are exposed. The holes are then backfilled with a conductor, to form the bridge between the walls of the hole and the underlying metal pad. The junction is formed around the hole during the ion beam milling process as described in Sect. 34.6.1, *Junction-Forming Techniques*. The second masking stage enables the p-side contact to be applied around the array. Figure 34.10 shows a photomicrograph of one corner of the hybrid illustrating the membrane-like nature of the HgCdTe on the silicon.

The HDVIP process [34.52] results in a similar structure but uses a plasma-etching stage to cut the via-hole and an ion implantation stage to create a stable junction and damage region near the contact. In order to achieve higher lifetimes and lower thermal currents, Cu is introduced at the LPE growth stage. This is swept out during the diode formation and resides selectively in the p-region, partially neutralising the S-R centers associated with Hg vacancies and achieving dark currents approaching that of fully doped heterostructures. In via-hole processes, the cylindrical shape of the junction minimizes the intersection with threading dislocations (see Sect. 34.5.4, *Excess Noise Sources in HgCdTe Diodes*). This has been demonstrated by the difference in yields when the HDVIP device is extended by a planar junction [34.52].

LW HgCdTe detectors were originally developed for long linear arrays and the processes can be extended to 2-D arrays with one additional consideration. The earth plane of the array must carry the sum of the photocurrent from all the pixels in the array, and if there is insufficient conductance, a voltage will appear between the center and edge of the array. This is called *substrate debiasing* and in the extreme case will inhibit injection in the center diodes. In any event, it provides a potential long-range crosstalk mechanism and detector designers will try to minimize the resistance of the earth plane layer. In homojunction arrays, the p-type common layer cannot be doped high enough and a special metal grid is needed, as shown in the array in Fig. 34.11. In other technologies, the common layer must be doped high enough to prevent long-range crosstalk and use junctions with high dynamic resistance.

Planar Device Technology Using LPE

The planar device structure illustrated in Fig. 34.12 is the simplest device structure currently used. It is consistent with a number of junction-forming processes, e.g., ion implantation, diffusion, and ion milling. The matrix of junctions is mass connected to a silicon ROIC using indium bumps. The strength of the process is the simplicity and the compatibility with epitaxially grown materials. *Tribulet* et al. [34.55] describe a process based on high-quality LPE material and ion-implanted junctions, which is conducive to good quality detectors and volume production. In the simplest form,

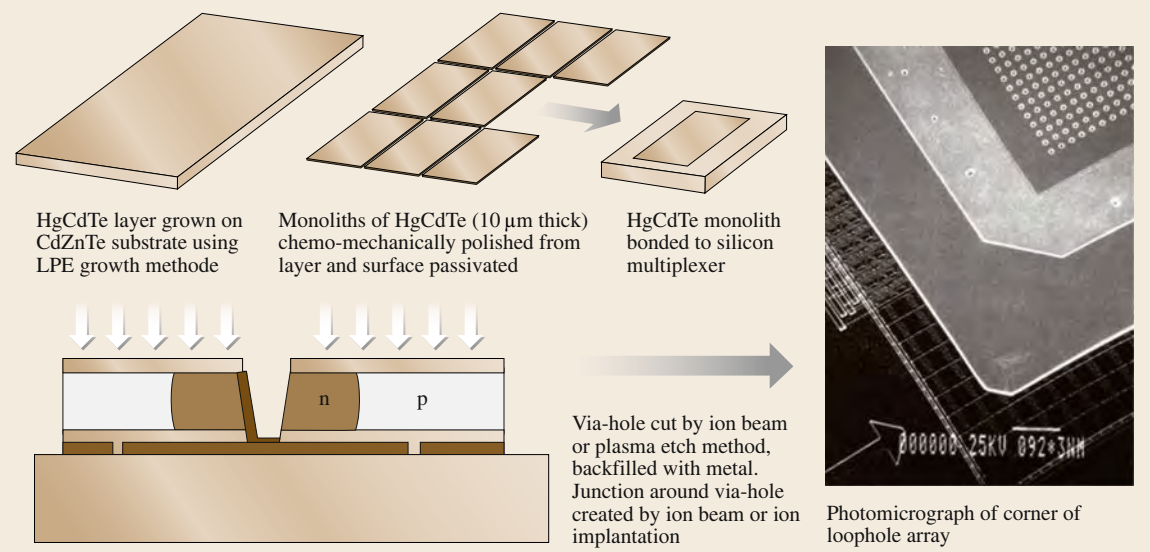


Fig. 34.10 Schematic of via-hole process

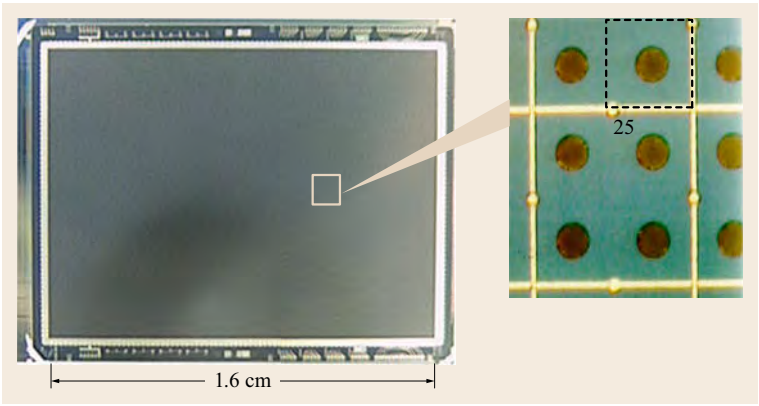


Fig. 34.11 LW HDVIP array with metal grid for substrate debiasing control. (Courtesy of DRS Technologies)

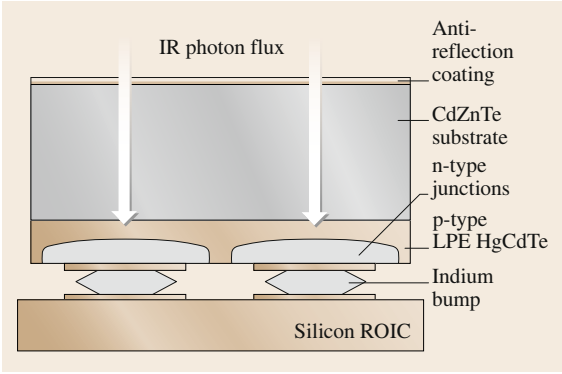


Fig. 34.12 Schematic of planar indium-bumped hybrid

the process needs three masking stages, for the junction and pixel contact, and one for the contact to the p-side. More recent publications have described the status at two manufacturing centers for instance: [34.56, 57] with the listed references.

The device is backside illuminated, that is, it is illuminated through the substrate, and because of interdiffusion, there is a graded bandgap region which aids diffusion of carriers to the junction. Careful control of the junction geometry is needed to avoid crosstalk due to the diffusion of minority carriers into adjacent pixels, especially in the case of small pixel sizes. The thermal expansion mismatch between the HgCdTe/CdZnTe substrate combination and the silicon multiplexer requires that in larger arrays, the CdZnTe substrate is thinned. In this case, the spectral sensitivity can be extended into the visible.

34.6.2 Vapour Phase Epitaxial (VPE)-Based Processes

Two growth processes are widely used: MBE and MOVPE. A good detailed summary of MBE technology has been prepared by *Wu* et al. [34.58]. *Maxey* et al. [34.59] describe in detail the MOVPE process. A brief comparison is included here.

MBE offers the lowest temperature growth under an ultra-high vacuum environment, and in common with MOVPE in situ doping and control of the composition and interfacial profiles. These are essential for the growth of advanced and novel device structures. Typically, growth is carried out at 180–190 °C on (211) CdZnTe substrates. Effusion cells of CdTe, Te, and Hg are commonly used. Hg is incorporated in the film only by reacting with free Te; thus, the composition depends on the Te-to-CdTe flux ratio. The structural perfection depends strongly on the Hg-to-Te flux ratio, and growth is usually restricted to a tight temperature range. Indium is the most widely used n-dopant and is well activated. P-type dopants are less conveniently incorporated in situ, but manufacturers have devised a number of processes to force As onto the proper Te site. Again the Hg-to-Te ratio and growth temperature are crucial to achieve good activation. MBE structural problems center mainly on pinholes or voids.

MOVPE offers higher growth rates which enable multiple runs per day to be achieved with a single growth reactor. MOVPE growth depends on transporting the elements cadmium and tellurium (and dopants iodine and arsenic) at room temperature as volatile organometallics. They react along with Hg vapor in the hot gas stream above the substrate or catalytically on the substrate surface. The drive to lower temperatures and, hence, Hg equilibrium pressures has resulted in the adoption of the Te precursor di-isopropyl telluride, which is used for growth in the 350–400 °C range. A key step in the success of this process is to separate the CdTe and HgTe growth, so that they can be independently optimized. This is called the IMP process for interdiffused multilayer process. The resulting stack of alternating CdTe and HgTe layers relies on the fast interdiffusion coefficients in the pseudobinary to homogenize the structure at the growth temperature. The

wavelength of $Hg_{1-x}Cd_xTe$ is determined by the composition x which is, in turn, dependent on the growth rates during the CdTe and HgTe cycles. The key difference to MBE is that control of HgCdTe composition can be simply accomplished by the length of time the Cd source flows into the reactor for the CdTe part of the IMP cycle. Also, since the dopants are incorporated on the CdTe cycle, the doping density is relatively independent of composition. Arsenic is used for p-doping (*Tris*-di-methyl amino arsine) and iodine for n-doping (*Iso*-butyl iodide), and are actively incorporated during growth without a requirement for an ex-situ anneal. The main morphological risk for MOVPE is the creation of hillocks that can cause a cluster of pixel defects. Particles or other surface roughening features can lead to the creation of these hillock features due to the fast ⟨111⟩ microcrystal growth rates that enhance any surface perturbation. Growth orientations 3–4° off (100) are used primarily to reduce both the size and density of hillocks.

Double-Layer Planar Heterojunction Devices (DLPH)

When a device uses different compositions or doping levels, a nomenclature is adopted for describing the structure. The definition used in this chapter is as follows. A diode comprising an n-type layer on a p-type layer is described as an n on p device, so the first letter is the layer nearest the contact. A superscript '−' or '+' denotes particularly low or high doping. Wider bandgap materials such as those used in barriers or contact areas are described as N or P.

Mesa heterostructures have a technologically difficult passivation stage on the sidewalls of the mesa, which can lead to reverse-bias leakage and uniformity problems if not optimum. *Arias* et al. [34.60] have demonstrated a clever diffused heterostructure, which overcomes the sidewall problem, called double-layer planar heterostructure (DLPH). This is a back-illuminated arsenic-implanted p–n–N planar buried junction structure [34.61] illustrated in Fig. 34.13. The junction is formed by arsenic implantation into a three layer N–n–N film grown in situ by MBE onto a CdZnTe substrate. The unique feature of this structure is that the junction is buried beneath the top wide bandgap n-type layer. The junction intercepts the surface in wider gap CMT, thereby reducing the generation rates for any surface defects that may be present. The arsenic-doped region must extend deep enough into the narrow gap absorber layer in order to collect photocarriers. This necessitates that the wide bandgap layer be thin (on the order of 0.5 μm) and that the compositional interdiffusion between the wide gap and absorber be minimal. Growth by MBE at low temperature (175 °C) satisfies these requirements. Another feature of this structure is

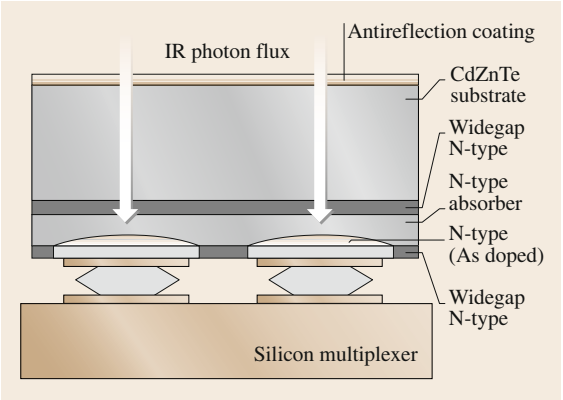


Fig. 34.13 Schematic of double-layer planar heterostructure (DLPH)

the wide bandgap N-type buffer layer between the absorber and the substrate. This buffer layer keeps carriers that are photogenerated in the absorber away from the substrate interface where there can be interface recombination. An empirical relationship for the dark current across a wide parameter space has been proposed called Rule 07 by *Tennant* et al. [34.18, 19]. This is often used as a baseline for HgCdTe devices when judging newer infrared technologies.

Recent progress on FPAs manufactured by the DLPH process is described by *Beletic* et al. [34.12]. The arrays are aimed at the civil space and astronomy markets, and they must meet very demanding specifications. For astronomy, the zodiacal background sets the specification for dark current as 0.01 e/s/pixel, a figure routinely met in specialized HgCdTe devices. In order to maximize the science return, the arrays need to be physically large and are often mosaicked into larger focal planes. The practical limit for DLPH arrays is around 6 cm^2 limited by the availability of single-crystal substrates of CdZnTe, and the largest array is 4 k × 4 k with 15 μm pixel size. Manufacturing such an array represents a triumph of electronic and thermo-mechanical engineering, and a photograph is shown in Fig. 34.14.

Mesa Heterojunction Devices (MHJ)

For heterostructures that are grown in situ, it is necessary to completely isolate the junction forming a so-called mesa device. The device structure is illustrated in Fig. 34.15, and this is sometimes called a double-layer heterojunction device (DLHJ). Each pixel is electrically isolated by a mesa slot that extends through the absorber to eliminate lateral collection which can often degrade the MTF of LPE-based processes. Manufacturers will create the mesa slot by the use of wet etches or dry processes such as ion beam milling, or a combina-

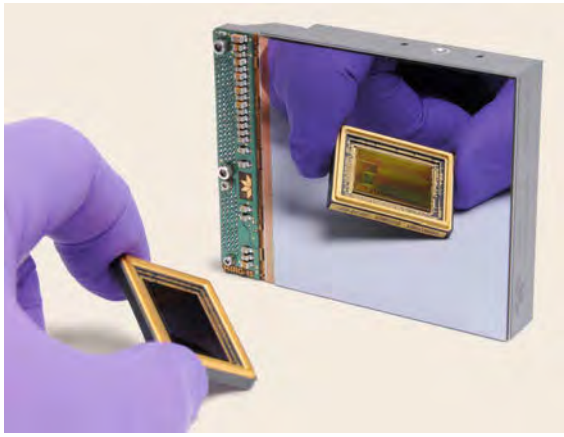


Fig. 34.14 4 k × 4 k SW array with 15 µm pixel size for astronomy. (Courtesy of Teledyne Imaging Sensors)

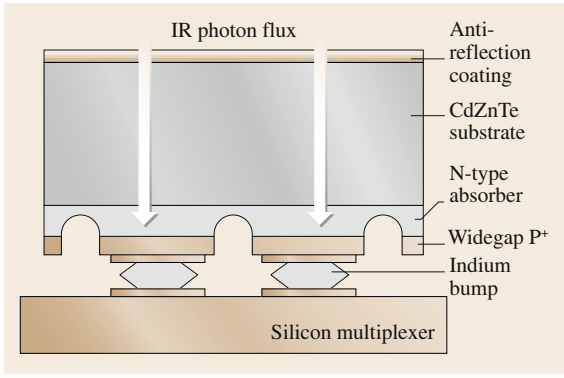


Fig. 34.15 Schematic of mesa heterojunction (MHJ)

tion of both. The sidewalls are then coated with a CdTe layer that is interdiffused at high temperature. The widening of the bandgap around the edges of the absorber effectively separates carriers from surface states and minimizes junction currents where the junction intercepts the sidewall. The high temperature anneal also promotes long-term stability and immunity from life at high operational temperatures. HgCdTe is the only infrared material that permits so-called heteropassivation, and it has a strong impact on the manufacturing yields and long-term process stability.

The main advantage of this process is that the electro-optic properties of the heterojunctions can be fully exploited. Layer thicknesses and doping levels are set at the growth stage. Normally, n-type material is chosen as the absorber because it is easier to get low doping levels and this provides long diffusion lengths and is relatively free of S-R-H centers. The more problematic P-side can then have a wider bandgap which tends to deactivate dark current generation. Because the absorber can have minimal volume in this process the dark currents are usually an order less than LPE homojunction technologies.

The first double-layer heterojunction devices, DLHJs, were reported from the United States in the 1980s with work continuing into the 1990s with LPE material. For summaries of this work, see *Tung* et al. [34.62], *Wang* [34.63], *Pulz* et al. [34.64], and *Norton* et al. [34.65]. *Tung* [34.66] reported the first n^-–P^+ structure grown from MBE which has the advantage of very low crosstalk because of the complete electrical isolation of the absorber.

In the modern era, mesa heterojunction technology is a commonly used technology for MBE and MOVPE wafer scale processes. A review of MOVPE is given by *Hipwood* et al. [34.37, 38] and *Baker* et al. [34.42]. With mesa sidewall shaping, the cone can become an optical concentrator with photons trapped within the mesa cone and completely absorbed. This ensures that stray photons do not scatter and contribute to the background current of other pixels. Stray photons reduce the integration time and have a direct impact on thermal contrast. Figure 34.16 shows a mesa heterojunction array with a 16 µm pitch and the ray-tracing analysis. The sidewall reflections have skewed paths through the absorber effectively lengthening the absorber by about a factor of 3. MHJs can, therefore, have the low optical and electrical crosstalk and this is especially important in small pixels.

Wafer Scale Processes Using Low-Cost Substrates

There is a dual purpose in developing vapor phase epitaxy technologies. Firstly, it opens up the possibility of growing complex multilayer materials for a whole range of new device structures and secondly it allows growth on lower cost substrates than CdZnTe. CdZnTe is expensive and only available in sizes up to about 60 cm^2, limited mainly by Zn segregation in the boule. It is brittle and soft making wafers hard to handle and can have a recurring problem of impurity outdiffusion. Also, cadmium is the subject of a number of international protocols concerning its use and disposal. CdZnTe substrates are mined from sawn up slices taken from large bulk crystal ingots grown by either vertical gradient freezing (VGF) or Bridgman methods. Conservative estimates predict that growth on alternative substrates reduces the cadmium utilization by two orders of magnitude.

Silicon, gallium arsenide, germanium, and sapphire are the most common substrates used because they are available commercially in large area wafers at relatively low cost. Figure 34.17 illustrates a typical process flow.

Of the four candidate substrate materials, silicon has the worst lattice mismatch and the worst ther-

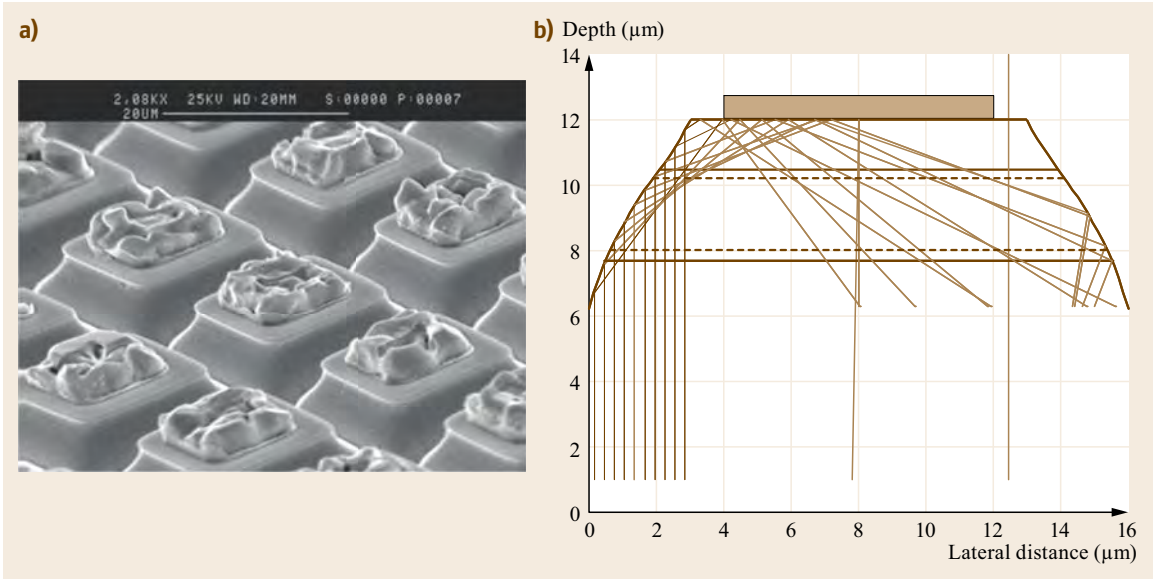


Fig. 34.16a,b 16 μm pitch mesa MOVPE device **(a)** and optical modeling **(b)**

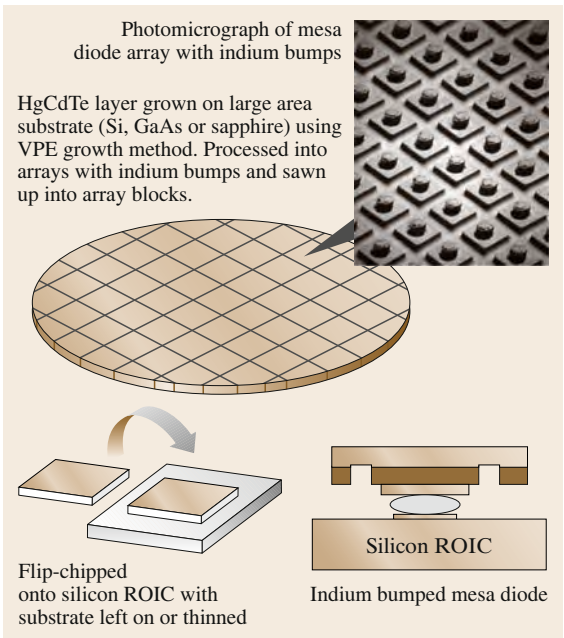


Fig. 34.17 Wafer scale processing using VPE HgCdTe material, mesa diodes, and indium bump interconnect technology

mal expansion difference. Nevertheless, the elegance of matching the ROIC thermal expansion is a powerful advantage when considering very large arrays and some early work explored growth on silicon. *Maxey* et al. [34.59] report a variety of device structures made using MOVPE layers grown on 75 mm diameter GaAs on silicon layers; however, the arrays had poor uniformity compared to those on GaAs alone. Similarly, *Carmody* et al. [34.67] found that buffer layers of CdTe grown on GaAs had better crystal quality, fewer dislocations, higher mobility, and longer minority carrier lifetimes than grown-on silicon. In general, silicon has fallen out of favor because of the device quality.

Commercial GaAs substrates have a number of advantages for use as substrates: they are readily available at low cost, have good crystal quality and an oxide that is easy to remove prior to growth. GaAs substrates of 76.2 mm diameter cost some five times less than GaSb (used in III–V detectors) and about 50 times less than CdZnTe. Since the substrate is at the front end of the process, the cost impact per array is divided by the yield and the number of arrays per wafer. It, therefore, has a superlinear relationship on the size of the array. There are also selective etches to remove the substrate at the end of the process and this is essential for two reasons: First, the thermal expansion difference between the ROIC and GaAs is high enough to cause cracking and, second, the thick substrate can cause optical crosstalk. *Arias* et al. [34.68] pioneered MBE growth on GaAs and demonstrated that good quality material could be grown. MBE growth on silicon substrates of 100 mm diameter has been used to grow MW HgCdTe layers as described by *Varesi* et al. [34.69]. The MBE used buffer layers of CdZnTe and CdTe to minimize the density of misfit dislocations.

Another approach, known as PACE, has been reviewed by *Bostrup* et al. [34.70]. The process uses a 75 mm diameter substrate of highly polished sapphire,

which remains with the device throughout. The thermal expansion problem is overcome by using a carrier substrate of sapphire, so that the silicon multiplexer is sandwiched in the middle. The process starts with a buffer layer of CdTe deposited using an organometallic vapor phase epitaxy (OMVPE) process to 8.5 μm thickness followed by a 10 μm thick p-type HgCdTe layer grown by LPE. The n–p diodes are formed by ion implantation. This process has been used to fabricate arrays up to 2048×2048 ($37\,mm^2$) in the SW band [34.14, 71]. The arrays are backside illuminated through the sapphire substrate.

A number of device types including SW, MW, LW, e-APD, and two-color arrays have been fabricated using MOVPE growth on low-cost 75 mm GaAs [34.42]. The process consists of growing a 2 μm thick layer of CdTe which begins the process of turning over misfit dislocations as revealed by transmission electron microscopy. Then, a thick layer of HgCdTe is grown which continues the process of improving the crystal quality. Provided the p–n junction is far from the substrate, the dark current performance of LW diodes can exceed that of the best LPE. The GaAs substrate is removed after bump bonding.

34.7 Advanced HgCdTe Technologies

Third Generation or GEN III is a commonly used term to describe advanced infrared detectors. The definition of a *GEN III* detector can differ between different nations, but the general guideline is any detector that offers an imaging advantage over conventional first- and second-generation systems. Common agreed examples are, for instance, megapixel arrays with high density, dual color or even multispectral arrays, higher operating temperature, fast readout rates, very low NETD due to pixel-level signal compression, and retina-level signal processing. Some of the research progress is described here in more detail.

34.7.1 Small Pixel Technology

Electro-optic systems strive to reduce the pixel size because it has a direct effect on the cost, size, weight, and power consumption of the overall camera. Alternatively, in conjunction with faster optics, the resolution can be improved and this is a useful retrofit path for existing cameras. Faster optics also increases the ratio of photocurrent to dark current, allowing higher operating temperatures. Clearly, the limitation of small pixel sizes is set by the diffraction limit of the optics (for instance [34.72]). *McEwen* et al. [34.73] have proposed that the pixel size should satisfy *Sparrow*'s *Undulation Condition* [34.74], setting a minimum value of $0.5\lambda f_{no}$ where λ is the mean photon wavelength that depends on the scene temperature and atmospheric conditions and f_{no} is the focal ratio of the optic. Even this is not exact because the human eye can resolve beyond this. However, many authors argue that oversampling the image provides benefits in the IR notably reduction of aliasing, suppression of point sources, and improved image processing. For cameras searching for small subpixel targets, there is the potential to suppress false alarms. On the negative side, small pixels can be difficult to manufacture compromising yields and performance. Also, maintaining a good resolution (MTF) in the presence of optical and electrical crosstalk limits the choice of technology. In very small pixels, the silicon ROIC becomes challenging as the integration capacitance is reduced and silicon MOSFETs tend to become noisy at small geometries. The degree that manufacturers can reduce the scale of their technologies depends on a detailed analysis of all these arguments.

Pixel sizes in the 10–12 μm range for MWIR systems are available in the market place with good maturity from planar devices [34.75], via-hole devices [34.52], DLPH [34.76], mesa heterojunctions [34.77].

Mesa heterojunctions based on MOVPE have been described by *McEwen* et al. [34.73] with a pixel size of 8 μm and implemented in a 1280×1024 pixel array called SuperHawk. The paper stresses the design of the mesa cone to act as an optical cavity and minimize crosstalk mechanisms limiting the spatial resolution of the device. Near-perfect MTF was measured using an F/2.8 optic and 3.5–5.0 μm spectral range.

The fabrication of 5 μm pixel dimension arrays on the DARPA Lambda Scale Program has been described by *Kinch* et al. [34.78] and *Robinson* et al. [34.79] using the HDVIP $n^+/n^-/p$, homo-junction, via-hole technology, and *Tennant* et al. [34.80] using a modified DLPH technology. Very large arrays can be made economically, and the small pixel size suits the fast optics often required in wide-area surveillance applications.

34.7.2 Higher Operating Temperature (HOT) Device Structures

Operating FPAs at higher temperatures are a key modern requirement and have far-reaching consequences on the size, weight, power, and cost of future thermal

imaging systems. The power consumption of high-performance thermal imaging cameras is often dominated by the Stirling cooling engine and operating a small cooler at 150 K instead of 80 K approximately halves the power as well as reducing the cooldown time. Higher temperatures also significantly improve the mean time before failure of a Stirling engine cooler. A satisfactory operating temperature clearly depends on the optics, scene temperature, waveband, and signal processing. Making comparisons from the literature is, therefore, difficult but rationalizing the data to a typical practical camera specification of F4, 290 K background, 3.5–5.0 μm waveband, and 50 mK NETD provides a rough guide to the status quo. Epitaxial InSb can be operated up to 100 K with a stretch potential to 120 K. LPE processes, nBn barrier devices, and Type II superlattices currently operate around 140–150 K. VPE processes using mesa cones can operate at 160 K with the stretch potential to 200 K. The main reason why VPE processes can be better is due to geometry. The absorber thickness can be minimized and reflections within the mesa cone effectively create long path lengths to maintain QE.

It is important to recognize that the bandgap of HgCdTe varies with temperature. For instance, a 5.7 μm cutoff array at 80 K becomes 5.0 μm at 200 K and so for HOT operation a lower x value should be selected. In contrast, InSb has a narrowing bandgap at higher temperature which further degrades its potential.

For each process, there are some common denominators for improving the performance at higher temperature. Manufacturers try to reduce the absorber volume for instance by using thinner structures and employ weak doping levels to limit Auger generation. However, the limit of acceptable image quality is not usually controlled by dark current but rather by the generation of excess noise and the increase in defective pixels. To that extent, it is not fundamental but should improve with a reduction in crystal defects. Ultimately, thermal emission from the radiation shield will start to dominate over about 200 K.

The best reported HOT imaging is described by *Pillans* et al. [34.81] using mesa heterodiodes in MOVPE with lightly doped p-type absorbers. Figure 34.18 shows an image at 210 K from a scene on a warm day using a 16 μm pixel array in the full MWIR band. The higher temperature has more temporal noise (not evident in the image) but completely acceptable spatial resolution. One of the reasons this is possible is that arrays at 80 K can have NETD values around 15 mK and the NETD can be degraded by an order of magnitude before there is recognisable degradation in an image. First generation cameras had equivalent NETD values

Fig. 34.18 Imaging at 210 K operating temperature. (Courtesy of Leonardo MW Ltd)

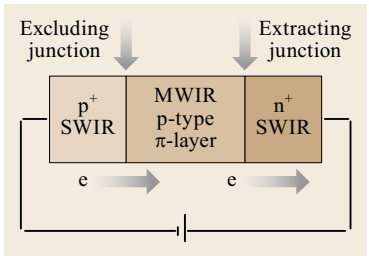


Fig. 34.19 Schematic of non-equilibrium HOT device structure

of 200 mK and were completely acceptable. Also, the application of image-processing software can mitigate many of the noise artifacts.

Operation at close to room temperature has been proposed primarily as a result of research in the UK in the 1980s and 1990s. Good summaries have been reported by *Elliott* [34.82] and *Elliott* et al. [34.83]. The possibility of operating detectors at temperatures near room temperature was proposed by *Elliott* and *Ashley* [34.84]. The basic structure is a $Pp^{-}N$ illustrated in Fig. 34.19.

The active volume of the absorber is small relative to a minority carrier diffusion length and it is operated in strong non-equilibrium by reverse biasing the minority carrier contact to completely extract all of the intrinsically generated minority carriers. To preserve space–charge neutrality the majority carrier concentration drops to the background dopant concentration and Auger generation is effectively suppressed. Remaining components are associated with S-R centers and possibly injection from the contact and surface regions. The realization of HOT LW arrays will depend on the suppression of high $1/f$ noise associated with the reverse bias operation. In MW arrays it is also important to suppress competing recombination mecha-

nisms. Nevertheless the device concept is a route toward high-performance infrared detectors with minimal cooling.

34.7.3 Two-Color Array Technology

Thermal emission from scenes with a mixture of materials can have spectral variations analogous to color in the visible. For instance the Si–O bond in many rocks can suppress LW radiation in comparison to man-made materials. Realtime multicolor sensors are still in development but a good stepping stone is the dual-waveband detector (DWB) with simultaneous readout of flux levels in 2 separate infrared bands. MW/LW arrays are well developed because there are visibility benefits to both bands depending on the conditions. MW has longer range in high humidity and better spatial resolution giving increased identification ranges in good conditions. In contrast, LW is better in more turbulent conditions, more useful for fast-moving targets and is also better for penetrating most smokes and fogs. Another interest is to separate hot objects from solar reflections and this can performed by comparing the thermal waveband with the SW band or two thermal bands. DWBs incorporate both wavebands into a single detector, giving the user the ability to switch between the two as circumstances require. This clearly offers a cost advantage over the need for two separate detectors. Image fusion can be used to highlight areas of interest by exposing spectral differences. For a given system aperture, MW offers twice the usable resolution of LW, which can be exploited in various ways in system design.

Systems designers would prefer to have co-spatial and co-temporal pixels but this is very difficult to achieve in practice. Such arrays need two indium bumps per pixel and a deep etch to make contact to the intermediate layer. Manufacturers will try to make the colors side by side within the blur circle of the optics for quazi-co-spatial or butt the integration periods of the MW and LW for quazi-co-temporal. In this case, it may be acceptable because the MW integration period usually dominates.

A common approach is to use a so-called bias-selectable detector [34.85–87] which needs only one bump and is compatible with small pixel sizes. Figure 34.20 shows a schematic structure. The simplest device is an n^-–P–n^- structure. The two n-type regions are the MW and LW absorbers, with appropriately chosen bandgaps. The intermediate wide-gap p-type region prevents transistor action between the two absorbers. The design uses one indium bump per pixel. The two junctions are stacked vertically so that the radiation travels through the MW absorber before the LW absorber, and the band to be detected is selected by the polarity of the applied bias.

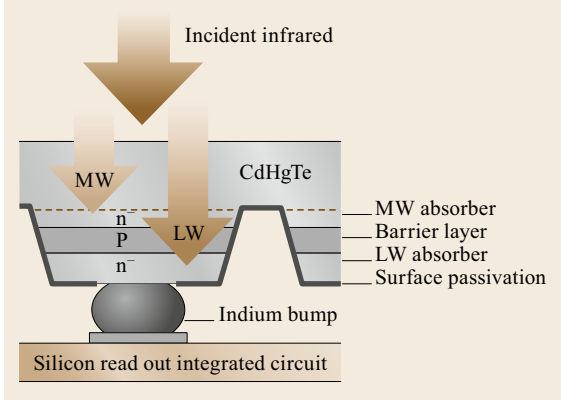


Fig. 34.20 MW/LW bias-selectable architecture for VPE processes

Several successful arrays have been reported. *Reine* et al. [34.88], report a 64×64 simultaneous MW/LW dual-band HgCdTe array on 75 μm pitch. The array is fabricated from a four-layer P–n–N–P film grown in situ by MOVPE on CdZnTe and mounted on a custom ROIC. This approach has been applied to the double layer planar heterostructure, DLPH, in a process called SUMIT for Simultaneous Unipolar Multispectral Integrated Technology [34.89]. The DLPH structure can be turned into a two-color device by having two absorbers of different bandgap separated by a wide bandgap barrier. An etch step is used to expose the lower, shorter wavelength absorber, and form an arsenic doped junction. A pixel size of 40 μm for each color is used in 128×128 demonstrators for MW/MW arrays. A variant of this basic process has also been presented by AIM [34.90], and they have achieved two colors in a concentric arrangement in 192×192 and 256×256 arrays with 56 μm pitch, again using a full custom ROIC.

Kinch [34.52] reports a quazi-co-spatial via-hole type structure with a 5.2 μm monolith on top of 4.2 μm material. Each cell has two via holes connecting the two layers. One is for connecting the longer wavelength array to the silicon in the normal way and the other is isolated from the longer wavelength material and connects the shorter wavelength to the silicon. The 2-color arrays utilize standard 640×480, 25 μm pitch ROICs (as used in LW FPAs). The unit cell design in these ROICs incorporates one buffer amplifier circuit for every 2 HgCdTe pixels. The 2-color array actually uses 4 mono-color pixels to form a 2-color unit cell, thus resulting in $50\,\mu m \times 50\,\mu m$ geometry, with 2 buffer amplifiers per unit cell, one for each color. Figure 34.21 shows a typical spectral separation.

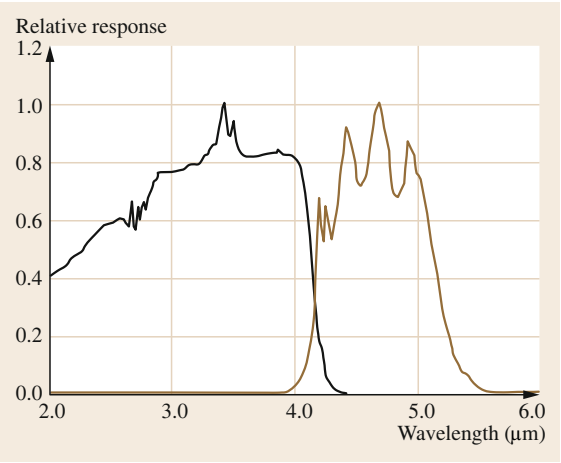


Fig. 34.21 Spectral response of MW/MW 2-color HDVIP HgCdTe FPA. (Courtesy of DRS Technologies)

DWB detectors in the LW and MW bands have been realized through multilayer VPE growth structures ([34.91–93] and references therein). These papers detail the array sizes that are commercially available and the quality of the arrays in each band.

Spectral crosstalk is an important performance parameter for dual-band detectors. There are two components of spectral crosstalk. One is unabsorbed MW radiation which appears as LW signal. The second is from photon recycling, where carriers in the MW absorber recombine radiatively with the generated photon passing into the LW absorber. Optical filters are employed in DWB detectors and when weighted for radiation from a 293 K blackbody, the spectral crosstalk of dual-band detectors can be as low as 0.2% [34.77].

34.7.4 HgCdTe Electron Avalanche Photodiodes (e-APDs)

Avalanche photodiode arrays (e-APDs) are proving to be vital components in many future infrared systems especially those with low photon counts. Although the concept dates back to the 1980s [34.33] the first e-APD arrays were produced in the early 2000s (for instance [34.94]). A number of centers developed e-APDs for laser-gated imaging (LGI) or burst-illumination LIDAR (BIL), now well established for long-range target identification [34.95]. Without avalanche gain the best detectors can sense a return signal of around 200 photons, but some requirements call for single photon detection. In this case avalanche gains of up to 100× are needed. It is remarkable that HgCdTe e-APDs have been able achieve this in arrays as large as 320×256 and at pixel sizes ≈ 20–$30\,\mu$m. In the commercial area there is growing interest in avalanche photodiode arrays for the 1.0–2.5 μm region. Ground-based astronomical telescopes now rely on adaptive optics to correct for atmospheric distortion. The trend is to sense the wavefront using guide stars in the 1.0–2.5 μm region. In current detectors the best sensitivity is typically around 30 photons, but this can be reduced to a few photons with avalanche gain providing much faster and more accurate wavefront correction. Atmospheric science, particularly remote gas analysis, is calling for LIDAR systems with near-single photon sensitivity. Typically the pulse laser would be tuned to a resonant frequency of the target gas molecule and the return signal would be tracked in time to give a depth profile.

The solid-state mechanisms for avalanche gain in HgCdTe have been well described and three recommended papers are [34.96–98] and the references therein. Figure 34.22 shows a potential energy schematic for a typical photodiode and it illustrates the single-carrier, cascade-like gain mechanism, together with the history-dependent behavior that leads to the observed low noise figures. The heavy hole mass is $\approx 0.55\,m_o$ across the entire HgCdTe composition range, resulting in low hole mobility values, and a significant degree of optical phonon scattering. Thus the heavy hole acquires energy from applied fields very inefficiently, and readily loses what it does gain to optical phonons. The effective mass of the electron, on the other hand, is very small, $m_e^*/m_o \approx 7 \times 10^{-2}\,E_g$, where E_g is the bandgap in eV. Its mobility is high, particularly at low temperatures, and scattering by optical phonons is weak, which results in significant energy gain at even modest applied fields. The conduction band of HgCdTe is also devoid of any low-lying secondary minima, which allows for large electron energy excursions deep into the band, and hence the high probability of impact ionization, with the generation of electron–hole pairs. In HgCdTe, this is an electron gain mechanism so the absorber must be p-type, which dictates an n-on-p structure. Theoretical studies indicate that there is an optimum width for the depletion region (gain zone) of 1.5–2.5 μm; at the lower end is a risk of gain saturation and tunneling currents; and at the upper end, alloy or phonon scattering starts to impact the ionization threshold voltage. One of the reasons APD arrays have matured so quickly is that these conditions are easily met using near-standard manufacturing processes.

In fact, the e-APD offers an extraordinary amplifier specification: voltage controlled gain at the point of absorption, electron gain values up to 1000×, virtually zero power consumption, bandwidths to GHz, high stability, high uniformity, no impact on the silicon pixel design and negligible added noise. On the cautionary side, the noise figure of HgCdTe e-APDs is often

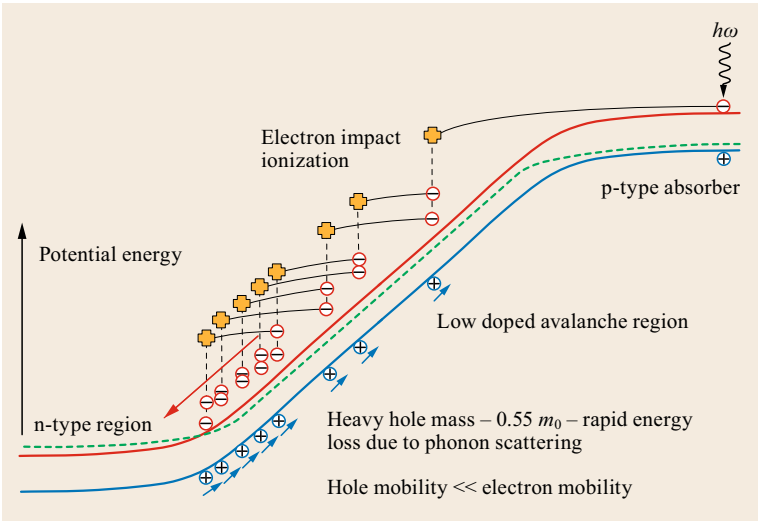


Fig. 34.22 Potential energy schematic – illustrating history-dependent avalanche gain in HgCdTe

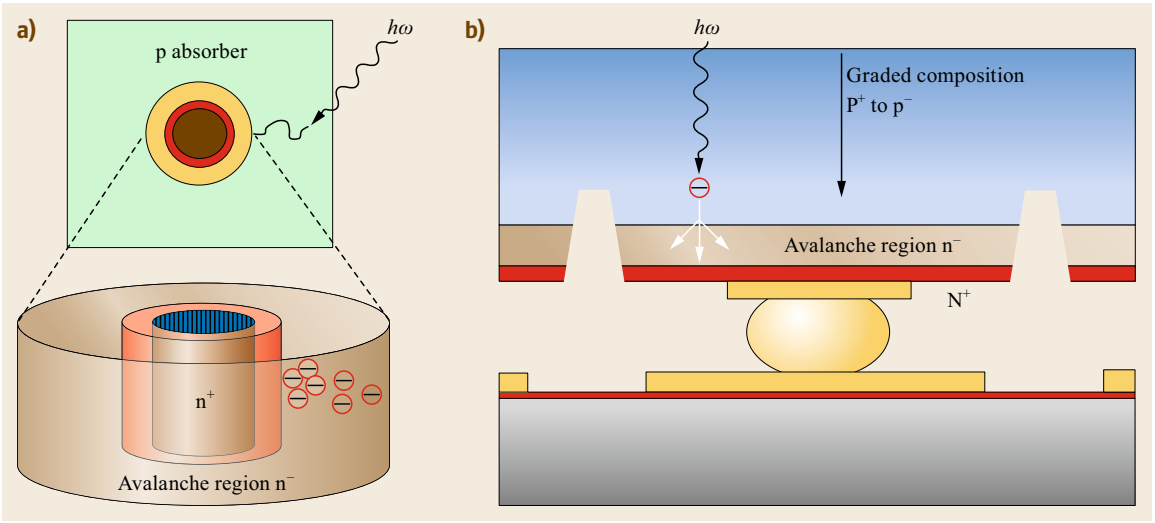


Fig. 34.23a,b Schematics of the device structures used for APDs: **(a)** LPE/via-hole technology; **(b)** MOVPE/mesa technology

quoted as being negligible but APD noise figures only describe the gain fluctuation. Noise due to dark current is not included and can be significant particularly as it is gain-amplified in many devices. Most photon starved applications can exploit avalanche gain but some care needs to be taken with dark current.

Three technologies are currently used for e-APDs – planar diffused processes as described in Sect. 34.6.1, *Planar Device Technology Using LPE*, via-hole processes as described in Sect. 34.6.1, *Via-Hole Technologies Using LPE* and mesa heterojunctions as described in Sect. 34.6.2, *Mesa Heterojunction Devices (MHJ)*. The via-hole and mesa heterojunction e-APD devices are illustrated in Fig. 34.23. In via-hole processes [34.99] the open-ended n^+n^-/p^+ cylindrical diode geometry is attractive because of the stronger dependence of diode potential with distance away from the n^+ region. This allows for large bias voltages to be used before the diode depletion width attains a value equal to a phonon mean free path, resulting in efficient, phonon-free, electron avalanche gain. Heterojunction processes allow the absorber, junction region, and multiplication region to be independently optimized. This is important for controlling the peak fields around the junction to allow high voltages without incurring excessive dark current.

Avalanche gain normally has an exponential dependence on applied voltage but the exponent and ioniza-

tion threshold voltage are dependent on device level parameters such as the electric field distribution and the composition through the depletion region. *Rothman* et al. [34.100] have developed a local field impact ionization and gain model for SW e-APDs that takes into account gain saturation and phenomenological expressions for the local impact ionization. Equation (34.10) is a good baseline for illustrating the behavior of avalanche gain with voltage, composition, and depletion layer width.

$$M = \exp\left[aE_g^{-1}V^c W^{1-c}\exp\left(-\frac{bW}{V}\right)\right], \quad (34.10)$$

where M is the avalanche gain, a, b, and c are fitted constants taken here as 22.6, 32 500 and 0.6, respectively, V is the applied voltage and W the depletion layer width in cm. Figure 34.24 shows an illustrative example of how the gain varies.

At first sight, it would appear to be beneficial to always use narrow bandgap compositions but in practical systems there is a risk of increased dark current and stray light. In this case, the noise equivalent photons (NEPh) can be calculated as a function of the e-APD parameters and extra noise. The derivation involves the solution of a quadratic and a useful form is given by

$$\mathrm{NEPh} = \frac{F}{2Q}\left(1+\left[1+\left(\frac{2\mathrm{FN}}{FTM}\right)^2\right]^{0.5}\right), \quad (34.11)$$

where F is the noise figure, Q is the QE, FN is the fixed noise, and T is the transfer function (gain from pixel to output – typically in the region of 0.7).

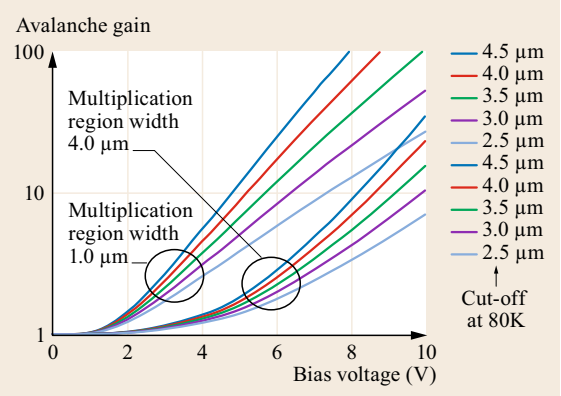


Fig. 34.24 Variation of avalanche gain with bias voltage and cut-off wavelength for two different multiplication region widths. (After [34.100])

e-APDs are being increasingly considered for astronomy, initially for wavefront sensors but increasingly for imaging and spectroscopic instruments. *Finger* et al. [34.101] and *Atkinson* et al. [34.102] summarise the progress with the Saphira device which is designed for the fast frame rate and non-destructive readout. They show that the avalanche gain can increase the photon sensitivity in proportion to the gain. Low noise measurements can be achieved by multiple sampling and Fowler sampling techniques for read noise as low as 0.27 electrons rms – an order better than with conventional detectors. This class of detector will be used for single-photon sensing for a range of scientific and defence applications.

34.7.5 Multifunctional HgCdTe Detectors

An important consequence of the avalanche gain is that it allows much more complex FPAs to be designed. This is because the photon noise is amplified and it relaxes the noise requirement on the silicon. In effect, the size of the silicon components can be decreased allowing for more complex circuits in the pixel and more functionality. There are often no disadvantages to using more avalanche gain, so it can be used as an enabling step for multifunctional FPAs.

An essential modern requirement is the ability to perform detection, recognition, identification, and intent over long ranges, day and night. Burst-illumination LIDAR (BIL) is an emerging field [34.103]. Collimated pulse lasers can give instantaneous brightness far in excess of passive infrared emission levels even at extreme long range. The brightness allows the use of long focal length telescopes made from glass (because of the 1.55 μm wavelength). Also at 1.55 μm there is better optical resolution compared with the longer wavelengths of passive IR systems. It is therefore fairly straightforward to achieve resolution improvements of over an order of magnitude compared with standard thermal imaging. Another advantage stems from the short laser pulse (typically 20 ns wide) which has the effect of freezing the scene, including any own-platform jitter, and atmospheric effects. There will be distortion from frame to frame but images can be de-rotated and re-registered to restore stable displayed scenes. Such systems also need thermal imaging for scene context and a detector which combines BIL and conventional thermal imaging reduces the number of apertures and system complexity.

It is very straightforward to manufacture e-APD arrays from HgCdTe with a wavelength in the MW band so they can be used for both SW active and MW passive applications. Multifunctional arrays have been

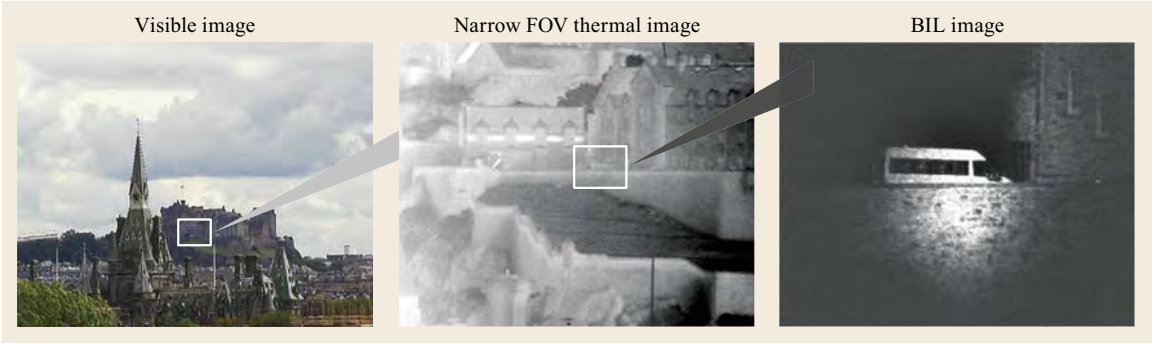


Fig. 34.25 Example of dual-mode passive and active imaging. The active BIL image is for a focal ratio of F10 and avalanche gain of ×30. (After [34.103])

manufactured with a 3.2–4.2 μm thermal channel and three-dimensional (3-D) BIL [34.104]. Figure 34.25 shows nonmilitary sample images demonstrating the dual-mode behavior.

34.7.6 Retina-Level Processing

In very large arrays operating at high frame rates, the downloading of signal data and the subsequent signal processing can be daunting, and it is called the data processing bottleneck problem. The human retina presents an example of how evolution has dealt with the problem. The eye performs a number of image-processing operations in the *z-plane* including, in order logarithmic photon sensing, spatial filtering, temporal filtering, motion sensitivity, and data decomposition. An essential function is the difference of Gaussian or DoG filter which is used for both edge and contrast enhancement [34.105].

FPAs with neuromorphic processing have been explored at a number of centers [34.106, 107]. A good example is the MIRIADS program (Miniature InfraRed Imaging Applications Development System). MIRIADS uses a neuromorphic FPA, with temporal high pass filtering, frame co-adding, and a DoG operation to detect motion, enhance edges, and reject ambient light levels. The current reported array is a 64×64. More information is presented by *Baxter* et al. [34.108].

34.7.7 Conclusions and Future Trends for Photovoltaic HgCdTe Arrays

Photovoltaic 2-D HgCdTe detectors are serving many applications worldwide and are supplanting photoconductive detectors in second-generation systems for military, commercial, and scientific applications, offering improved temperature sensitivity, lower power consumption, weight, and volume. The continuing development work within manufacturing centers, supported by the wider scientific community, is aimed at increasing the operating temperature and reducing manufacturing costs. Medium-term research is being directed at the best advanced materials and device structures for the next generation of HgCdTe arrays, which will combine the state-of-the-art background-limited performance with low-cost wafer-scale processing, including growth on the silicon readout for wafer-scale devices. Band-gap engineering will produce heterostructure detectors with much higher operating temperatures. This will enable background-limited operation in MW arrays above 200 K to exploit novel cryogenic engines. Device structures will be extended to produce bi-spectral and multispectral capability. Arrays will have larger physical size, smaller pixels, higher sensitivity, and on-focal plane signal processing to achieve a faster frame rate. The ultimate performance potential of HgCdTe will ensure that it is the material of choice for all high-performance infrared systems.

34.8 Emission Devices in II–VI Semiconductors

Infrared emission devices, in the form of light-emitting diode (LEDs) and laser diodes, are used in a broad range of applications, such as gas sensors, infrared scene simulators, molecular spectroscopy, free-space communications, fiber-optic communications, and LIDAR (light detector and ranging). A useful summary of emission device theory and current progress has been presented by *Gordon* [34.109]. There is competition from lead salt semiconductors (in the 3–4 μm range) and narrow gap III–V compounds. HgCdTe LEDs use

a forward-biased junction to inject minority carriers into an n-type region typically with low doping to encourage radiative recombination. The external photon efficiency is limited by factors, such as competing recombination mechanisms, re-absorption of the photons, and internal reflection. Nevertheless, efficiencies as high as 5% have been achieved at 4 μm using an immersion lens to overcome internal reflection [34.110].

One of the disadvantages of the LED is that the spectrum is often too broad for many applications and lasers are needed. Lasers also offer greater output power and higher spatial coherence, but they are more difficult to make and have additional noise mechanisms. Infrared lasers covering the 3–5 μm band have been demonstrated with pulsed operation at 80 K. It is only recently that the complex double heterojunction structures needed for lasing have been available from low-temperature VPE processes, so the field is relatively young. Nevertheless, peak powers up to several milliwatts have been recorded in diode-pumped lasers and over 1 W in optically pumped lasers.

Using a structure similar to that reported in Sect. 34.7.2 and illustrated in Fig. 34.19 the carrier densities, in reverse bias, can be lowered to well below the thermal equilibrium values. Under this condition, the absorption of photons can exceed the emission rate and the surface of the device can look cold. This is called negative luminescence [34.111], and devices have been developed for use as temperature references in thermal imaging cameras.

Infrared LEDs have also been reported [34.112] in the dilute magnetic semiconductors $Hg_{1-x}Mn_xTe$ and $Hg_{1-x-y}Cd_xMn_yTe$, where, in principle, the bandgap could be tuned by the applied magnetic field.

34.9 Potential for Reduced Dimensionality in HgTe–CdTe

The emergence of low-temperature growth processes will enable HgTe–CdTe structures to be grown with dimensions small enough to show quantum confinement effects. At present, this research is in its early stages, and there are no active programs to make devices but this could change in future. HgTe–CdTe superlattices offer many benefits for detectors. The minority carrier lifetime could be enhanced beyond that of bulk material providing better QE. The optical absorption coefficient can be much larger than bulk material, allowing LW detectors to use absorbers as thin as a few microns. Furthermore, tunneling currents can be reduced by orders of magnitude, permitting operation at much longer wavelengths. The state of the art of reduced dimensionality material is summarized by *Meyer* and *Vurgaftman* [34.113].

Quantum well lasers are also known to offer significant advantages over double heterojunction devices with bulk active regions. For instance, the maximum operating temperatures for pulsed and CW lasers could be increased by over 100 K, and this would have important consequences for the attractiveness of narrow gap II–VI emitters.

Acknowledgments. The author would like to acknowledge the advice and support of my management and colleagues at Selex Southampton, in particular, Drs Peter Capper and Chris Maxey for their scientific contributions over many years. It is appreciated that many scientists and associates in other companies around the world have maintained dialog to advance scientific knowledge in this field. Thanks also to my wife Lesley for patience and help with the English.

References

34.1 W.D. Lawson, S. Nielsen, E.H. Putley, Y.S. Young: J. Phys. Chem. Solids **9**, 325 (1959)
34.2 C.T. Elliott, N.T. Gordon: Infrared detectors. In: *Handbook on Semiconductors*, ed. by C. Hilsum (North-Holland, Amsterdam 1993) p. 841
34.3 C.T. Elliott: Infrared detectors. In: *Handbook on Semiconductors*, Vol. 4, ed. by C. Hilsum (North-Holland, Amsterdam 1981) p. 727
34.4 A. Kolodny, I. Kidron: Infrared Phys. **22**, 9 (1992)
34.5 N. Oda: Proc. SPIE **915**, 20 (1988)
34.6 C.T. Elliott: Electron. Lett. **17**, 312 (1981)
34.7 C.T. Elliott, D. Day, D.J. Wilson: Infrared Phys. **22**, 31 (1982)
34.8 C.T. Elliott, C.L. Jones: *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications* (Chapman Hall, New York 1997), Chap. 16
34.9 R.M. Broudy, V.J. Mazurczyk: (HgCd)Te photoconductivity detectors. In: *Semiconductors and Semimetals*, Vol. 18, ed. by R.K. Williardson, A. Beer (Academic, Cambridge 1981) p. 157
34.10 M.B. Reine: Proc. SPIE **443**, 2 (1983)

34.11 M.B. Reine, K.R. Maschoff, S.B. Tobin, P.W. Norton, J.A. Mroczkowski, E.E. Krueger: Semicond. Sci. Technol. **8**, 788 (1993)
34.12 J.W. Beletic, R. Blank, D. Gulbransen, D. Lee, M. Loose, E.C. Piquette, T. Sprafke, W.E. Tennant, M. Zandian, J. Zino: Proc. SPIE **7021**, 70210H (2008) doi:10.1117/12.790382
34.13 N.T. Gordon, I.M. Baker: In: *Infrared Detectors and Emitters: Materials and Devices*, ed. by P. Capper, C.T. Elliott (Kluwer Academic, Dordrecht 2001) p. 23
34.14 L.J. Kozlowski: Proc. SPIE **2745**, 2 (1996)
34.15 L.J. Kozlowski, J. Montroy, K. Vural, W.E. Kleinhans: Proc. SPIE **3436**, 162 (1998)
34.16 M.B. Reine, A.K. Sood, T.J. Tredwell: Photovoltaic infrared detectros. In: *Semiconductors and Semimetals*, Vol. 18, ed. by R.K. Willardson, A.C. Beer (Academic, New York 1981) p. 201
34.17 M.B. Reine: Photovoltaic detectors. In: *Infrared Detectors and Emitters: Materials and Devices*, *Electronic Materials Ser.*, *Vol. 8*, ed. by P. Capper, C.T. Elliott (Springer, New York 2001)
34.18 W.E. Tennant, D. Lee, M. Zandian, E. Piquette, M. Carmody: J. Electron. Mater. **37**, 1406 (2008)
34.19 W.E. Tennant: J. Electron. Mater. **39**, 1030 (2010)
34.20 D.E. Lacklison, P. Capper: Semicond. Sci. Technol. **2**, 33 (1987)
34.21 P.L. Polla, R.L. Aggarwal, D.A. Nelson, J.F. Shanley, M.B. Reine: Appl. Phys. Lett. **43**, 941 (1983)
34.22 W.A. Radford, R.E. Kvaas, S.M. Johnson: Proc. IRIS Special Group Infrared Mater., Menlo Park (1986)
34.23 M.A. Kinch: *Fundamentals of Infrared Detector Materials*, Vol. TT76 (SPIE, Bellingham 2007)
34.24 O.P. Agnihotri, C.A. Musca, L. Faraone: Semicond. Sci. Technol. **13**, 839 (1998)
34.25 W.W. Anderson: Infrared Phys. **20**, 353 (1980)
34.26 J.Y. Wong: IEEE Trans. Electron Devices **27**, 48 (1980)
34.27 W.W. Anderson, K.J. Hoffman: J. Appl. Phys. **53**, 9130 (1982)
34.28 C.T. Sah: Phys. Rev. **123**, 1594 (1961)
34.29 R.E. DeWames, J.G. Pasko, E.S. Yao, A.H.B. Vanderwyck, G.M. Williams: J. Vac. Sci. Technol. **A6**, 2655 (1988)
34.30 Y. Nemirovski, D. Rosenfeld, R. Adar, A. Kornfeld: J. Vac. Sci. Technol. **A7**, 528 (1989)
34.31 D. Rosenfeld, G. Bahir: IEEE Trans. Electron Devices **39**, 1638 (1992)
34.32 Y. Nemirovsky, R. Fastow, M. Meyassed, A. Unikovsky: J. Vac. Sci. Technol. **B9**(3), 1829 (1991)
34.33 C.T. Elliott, N.T. Gordon, R.S. Hall, G.J. Crimes: J. Vac. Sci. Technol. A **8**, 1251 (1990)
34.34 I.M. Baker, C.D. Maxey: J. Electron. Mater. **30**(6), 682 (2003)
34.35 I.M. Baker, G.J. Crimes, C.K. Ard, M.D. Jenner, J.E. Parsons, R.A. Ballingall, C.T. Elliott: IEE Conf. Pub. **321**, 78 (1990)
34.36 J. Schuster, E. Bellotti: J. Electron. Mater. **43**(8), 2808 (2014)
34.37 L.G. Hipwood, C.L. Jones, D.C. Walker, J. Shaw, P. Abbott, R.A. Catchpole, M. Ordish, C.D. Maxey, H.W. Lau, P. Knowles, M.C. Wilson: Proc. SPIE **6542**, 65420I (2007)
34.38 L.G. Hipwood, I.M. Baker, C.L. Jones, C. Maxey, H.W. Lau, J. Fitzmaurice, M. Wilson, P. Knowles: Proc. SPIE **6940**, 69400G (2008)
34.39 Y. Nemirovski, A. Unikovsky: J. Vac. Sci. Technol. **B10**, 1602 (1992)
34.40 M.A. Kinch, R.L. Strong, C.A. Schaake: J. Electron. Mater. **42**(11), 3243–3251 (2013)
34.41 R.L. Strong, M.A. Kinch, J.M. Armstrong: Proc. SPIE **8704**, 870420 (2013)
34.42 I.M. Baker, C.D. Maxey, H. Weller, P. Thorne: Proc. SPIE **8542**, 85421A (2012) doi:10.1117/12.981850
34.43 P. Capper, T. Tung, L. Colombo: In: *Narrow-Gap II-IV Compounds for Optoelectronic and Electromagnetic Applications*, ed. by P. Capper (Chapman Hall, London 1997) p. 30
34.44 I.M. Baker, G.J. Crimes, J.E. Parsons, E.S. O'Keefe: Proc. SPIE **2269**, 636 (1994)
34.45 S.M. Johnson, J.A. Vigil, J.B. James, C. Cockrum, W. Konkel, M. Kalisher, R. Risser, T. Tung, W. Hamilton, W. Ahlgren: J. Electron. Mater. **22**, 835 (1993)
34.46 M.V. Blackman, D.E. Charlton, M.D. Jenner, D.R. Purdy, J.T. Wotherspoon, C.T. Elliott, A.M. White: Electron. Lett. **23**, 978 (1987)
34.47 S. Margalit, Y. Nemirovsky, I. Rotstein: J. Appl. Phys. **50**, 6386 (1979)
34.48 A. Kolodny, I. Kidron: IEEE Trans. Electron Devices **ED-27**, 37 (1980)
34.49 L.O. Bubulac, W.E. Tennant, R.A. Riedel, T.J. Magee: J. Vac. Sci. Technol. **21**, 251 (1982)
34.50 L.O. Bubulac, C.R. Viswanathan: J. Cryst. Growth **123**, 555 (1992)
34.51 J. Syz, J.D. Beck, T.W. Orient, H.F. Schaake: J. Vac. Sci. Technol. **A7**, 396 (1989)
34.52 M.A. Kinch: Proc. SPIE. **4369**, 566 (2001)
34.53 J. White, R. Pal, J.M. Dell, C.A. Musca, J. Antoszewski, L. Faraone, P. Burke: J. Electron. Mater. **30**(6), 762 (2001)
34.54 I.M. Baker, R.A. Ballingall: Proc. SPIE **510**, 210 (1985)
34.55 P. Tribulet, J.-P. Chatard, P. Costa, S. Paltrier: J. Electron. Mater. **30**(6), 574 (2001)
34.56 J. Wenisch, H. Bitterlich, M. Bruder, P. Fries, R. Wollrab, J. Wendler, R. Breiter, J. Ziegler: J. Electron. Mater. **42**(11), 31861 (2013)
34.57 O. Gravrand, G. Destefanis, S. Bisotto, N. Baier, J. Rothman, L. Mollard, D. Brellier, L. Rubaldo, A. Kerlain, V. Destefanis, M. Vuillermet: J. Electron. Mater. **42**(11), 3349 (2013)
34.58 O.K. Wu, T.J. deLyon, R.D. Rajavel, J.E. Jensen: Molecular beam epitaxy of HgCdTe. In: *Narrow-Gap II-IV Compounds for Optoelectronic and Electromagnetic Applications*, ed. by P. Capper (Chapman Hall, London 1997)
34.59 C.D. Maxey, J.P. Camplin, I.T. Guilfoy, J. Gardener, R.A. Lockett, C.L. Jones, P. Capper, M. Houlton, N.T. Gordon: J. Electron. Mater. **32**(7), 656 (2003)
34.60 J.M. Arias, J.G. Pasko, M. Zandian, S.H. Shin, G.M. Williams, L.O. Bubulac, R.-E. DeWames, W.E. Tennant: Appl. Phys. Letts. **62**, 976 (1993)
34.61 J. Bajaj: Proc. SPIE **3948**, 42 (2000)

34.62 T. Tung, M.H. Kalisher, M.H. Stevens, P.E. Herning: Mater. Res. Soc. Symp. Proc. **90**, 321 (1987)
34.63 C.C. Wang: J. Vac. Sci. Technol. **B9**, 740 (1991)
34.64 G.N. Pulz, P.W. Norton, E.E. Krueger, M.B. Reine: J. Vac. Sci. Technol. **B9**, 1724 (1991)
34.65 P.W. Norton, P. LoVecchio, G.N. Pultz, J. Hughes, T. Robertson, V. Lukach, K. Wong: Proc. SPIE **2228**, 73 (1994)
34.66 T. Tung: J. Cryst. Growth **86**, 161 (1988)
34.67 M. Carmody, A. Yulius, D. Edwall, E. Piquette, R. Jacobs, D. Benson, A. Stoltz, J. Markunas, A. Almeida, J. Arias: J. Electron. Mater. **41**(10), 2719 (2012)
34.68 J.M. Arias, S.H. Shin, E.M. Gertner: J. Cryst. Growth **86**, 362 (1988)
34.69 J.B. Varesi, R.E. Bornfreund, A.C. Childs, W.A. Radford, K.D. Maranowski, J.M. Peterson, S.M. Johnson, L.M. Giegerich, T.J. de Lyon, J.E. Jensen: J. Electron. Mater. **30**(6), 566 (2001)
34.70 G. Bostrup, K.L. Hess, J. Ellsworth, D. Cooper, R. Haines: J. Electron. Mater. **30**(6), 560 (2001)
34.71 K. Vural, L.J. Kozlowski, D.E. Cooper, C.A. Chen, G.L. Bostrup, C. Cabelli, J.M. Arias, J. Bajaj, K.W. Hodapp, D.N.B. Hall, W.E. Kleinhans, G.G. Price, J.A. Pinter: Proc. SPIE **3698**, 24 (1999)
34.72 R. Driggers, R. Vollmerhausen, J.P. Reynolds, J. Fanning, G.C. Holst: Opt. Eng. **51**(6), 063202 (2012)
34.73 K. McEwen, D. Jeckells, S. Bains, H. Weller: Proc. SPIE **9451**, 94512D (2015)
34.74 C.M. Sparrow: Astrophys. J. **44**, 76–87 (1916)
34.75 Y. Reibel, N. Péré-Laperne, L. Rubaldo, T. Augey, G. Decaens, V. Badet, L. Baud, J. Roumegoux, A. Kessler, P. Maillart, N. Ricard, O. Pacaud, G. Destéfanis: Proc. SPIE **9451**, 94512E (2015)
34.76 E.C. Piquette, W. McLevige, J. Auyeung, A. Wong: Proc. SPIE **9154**, 91542H (2014)
34.77 P. Thorne, H. Weller, L.G. Hipwood: Proc. SPIE **8353**, 83532J (2012)
34.78 M.A. Kinch: Proc. SPIE **9070**, 907032 (2014) doi:10.1117/12.2051335
34.79 J.M. Robinson, M. Kinch, D. Littlejohn, K. Jeppson: Proc. SPIE **9100**, 91000I (2014)
34.80 W.E. Tennant, D.J. Gulbransen, A. Roll, M. Carmody, D. Edwall, A. Julius, P. Dreiske, A. Chen, W. McLevige, S. Freeman, D. Lee, D.E. Cooper, E. Piquette: J. Electron. Mater. **43**(8), 3041 (2014)
34.81 L. Pillans, R.M. Ash, L. Hipwood, P. Knowles: Proc. SPIE **8353**, 83532W (2012)
34.82 C.T. Elliott: Proc. SPIE **3436**, 763 (1998)
34.83 C.T. Elliott, N.T. Gordon, A.M. White: Appl. Phys. Lett. **74**, 2881 (1999)
34.84 C.T. Elliott, T. Ashley: Electron. Lett. **21**, 451 (1985)
34.85 T.N. Casselman, G.R. Chapman, K. Kosai: US Workshop Phys. Chem. MCT II-IV Compd., Dallas (1991)
34.86 R.D. Rajavel, D.M. Jamba, J.E. Jensen, O.K. Wu, C. Le Beau, J.A. Wilson, E. Patten, K. Kosai, J. Johnson, J. Rosbeck, P. Goetz, S.M. Johnson: J. Electron. Mater. **26**, 476 (1997)
34.87 R.D. Rajavel, D.M. Jamba, J.E. Jensen, O.K. Wu, P.D. Brewer, J.A. Wilson, J.L. Johnson, E.A. Patten, K. Kasai, J.T. Caulfield, P.M. Goetz: J. Electron. Mater. **27**, 747 (1998)
34.88 M.B. Reine, A. Hairston, P. O'Dette, S.P. Tobin, F.T. Smith, B.L. Musicant, P. Mitra, F.C. Case: Proc. SPIE **3379**, 200 (1998)
34.89 W.E. Tennant, M. Thomas, L.J. Kozlowski, W.V. McLevige, D.D. Edwall, M. Zandian, K. Spariosu, G. Hildebrandt, V. Gil, P. Ely, M. Muzilla, A. Stoltz, J.H. Dinan: J. Electron. Mater. **30**(6), 590 (2001)
34.90 W. Cabanski, R. Breiter, R. Koch, K.-H. Mauk, W. Rode, J. Ziegler, H. Schneider, M. Walther, R. Oelmaier: Proc. SPIE **4369**, 547 (2001)
34.91 M.F. Vilela, K.R. Olsson, E.M. Norton, J.M. Peterson, K. Rybnicek, D.R. Rhiger, C.W. Fulk, J.W. Bangs, D.D. Lofgreen, S.M. Johnson: J. Electron. Mater. **42**(11), 3231 (2013)
34.92 J.P.G. Price, C.L. Jones, L.G. Hipwood, C.J. Shaw, P. Abbot, C.D. Maxey, H.W. Lau, J. Fitzmaurice, R.A. Catchpole, M. Ordish, P. Thorne, H.J. Weller, R.C. Mistry, K. Hoade, A. Bradford, D. Owton, P. Knowles: Proc. SPIE **6940**, 69402S (2008)
34.93 P. Abbott, L. Pillans, P. Knowles, R.K. McEwen: Proc. SPIE **7660**, 766035 (2010)
34.94 J.D. Beck, C.-F. Wan, M.A. Kinch, J.E. Robinson: Proc. SPIE **4454**, 188 (2001)
34.95 I.M. Baker, S. Duncan, J. Copley: Proc. SPIE **5406**, 133 (2004)
34.96 J. Rothman, L. Mollard, S. Gout, S. Bonnefond, J. Wlassow: J. Electron. Mater. **40**(8), 1757 (2011)
34.97 J.D. Beck, R. Scritchfield, P. Mitra, W. Sullivan III, A.D. Gleckler, R. Strittmatter, R. J. M. Martin: Advanced Photon counting techniques, SPIE Conf. Series, 8033 (2011)
34.98 M.A. Kinch, I.M. Baker: HgCdTe electron avalanche photodiodes. In: *Mercury Cadmium Telluride – Growth, Properties and Applications*, ed. by P. Capper, J. Garland (Wiley, Chichester 2011) p. 493
34.99 J.D. Beck, R. Woodal, M. Scritchfield, M. Ohlson, L. Wood, P. Mitra, J. Robinson: J. Electron. Mater. **37**, 1334 (2008)
34.100 J. Rothman, L. Mollard, S. Bosson, G. Vojetta, K. Foubert, S. Gatti, G. Bonnouvrier, F. Salveti, A. Kerlain, O. Pacaud: J. Electron. Mater. **41**(10), 2928 (2012)
34.101 G. Finger, I.M. Baker, D. Alvarez, D. Ives, L. Mehrgan, M. Meyer, J. Stegmeier: Optical and IR telescope instrumentation and detectors, Scientific Detector Workshop, Florence (2013)
34.102 D. Atkinson, D. Hall, C. Baranec, I. Baker, S. Jacobson, R. Riddle: Proc. SPIE **9154**, 915419 (2014)
34.103 I.M. Baker, P. Thorne, J. Henderson, J. Copley, D. Humphreys, A. Millar: Proc. SPIE **6206**, 620608 (2006)
34.104 I.M. Baker, D. Owton, K. Trundle, K. Thorne, P. Storie, P. Oakley, J. Copley: Proc. SPIE **6940**, 69402L (2008)
34.105 D. Marr: *Vision* (W.H.Freeman, New York 1982)
34.106 M. Masie, P. McCarley, J.P. Curzan: Proc. SPIE **1961**, 17 (1993)
34.107 P. McCarley: Proc. SPIE. **3698**, 716 (1999)
34.108 C.R. Baxter, M.A. Massie, P.L. McCarley, M.E. Couture: Proc. SPIE **4369**, 129 (2001)

34.109 N.T. Gordon: Emission devices. In: *Narrow-Gap II-IV Compounds for Optoelectronic and Electromagnetic Applications*, ed. by Peter Capper (Chapman Hall, London 1997) p. 486
34.110 P. Bouchut, G. Destefanis, J.P. Chamonal, A. Million, B. Pelliciari, J. Piaguet: J. Vac. Sci. Technol. B **9**, 1794 (1991)
34.111 T. Ashley, C.T. Elliott, N.T. Gordon, R.S. Hall, A.D. Johnson, G.J. Pryce: Infrared Phys. Technol. **36**, 1037 (1995)
34.112 R. Zucca, J. Bajaj, E.R. Blazewski: J. Vac. Sci. Technol. A **6**, 2725 (1988)
34.113 J.R. Meyer, I. Vurgaftman: Reduced-dimensionality HgTe-CdTe for the infrared. In: *Infrared Detectors and Emitters: Materials and Devices*, Electronic Materials, Vol. 8, ed. by P. Capper, C.T. Elliott (Kluwer, New York 2001)

35. Optoelectronic Devices and Materials

Stephen J. Sweeney, Jayanta Mukherjee

Unlike the majority of electronic devices, which are silicon based, optoelectronic devices are predominantly made using III–V semiconductor compounds such as GaAs, InP, GaN, and GaSb, and their alloys due to their direct-band gap. Understanding the properties of these materials has been of vital importance in the development of optoelectronic devices. Since the first demonstration of a semiconductor laser in the early 1960s, optoelectronic devices have been produced in their millions, pervading our everyday lives in communications, computing, entertainment, lighting, and medicine. It is perhaps their use in optical-fiber communications that has had the greatest impact on humankind, enabling high-quality and inexpensive voice and data transmission across the globe. Optical communications spawned a number of developments in optoelectronics, leading to devices such as vertical-cavity surface-emitting lasers, semiconductor optical amplifiers, optical modulators, and avalanche photodiodes. In this chapter, we discuss the underlying theory of operation of some important optoelectronic devices. The influence of carrier–photon interactions is discussed in the context of producing efficient and high-performance emitters and detectors. Finally, we discuss how the semiconductor band structure can be manipulated to enhance device properties using quantum confinement and strain effects, and how the addition of dilute amounts of elements such as nitrogen and bismuth is having a profound effect on the next generation of optoelectronic devices.

35.1 **Introduction to Optoelectronic Devices** 898
35.1.1 Historical Perspective 898
35.2 **Light-Emitting Diodes and Semiconductor Lasers** 900
35.2.1 Carrier–Photon Interactions in Semiconductors 900
35.2.2 Direct- and Indirect-Gap Semiconductors 900
35.2.3 Emission and Absorption Rates and the Einstein Relations 901
35.2.4 Population Inversion 902
35.2.5 Gain in Semiconductors 902
35.2.6 Density of States 902
35.2.7 Optical Feedback in a Fabry–Perot Laser Cavity 905
35.2.8 Wave Guiding 906
35.2.9 Carrier Confinement 907
35.2.10 Current Confinement 908
35.2.11 Laser Threshold and Efficiency 909
35.2.12 Carrier Recombination Processes 910
35.2.13 Temperature Sensitivity and T_0 912
35.3 **Single- and Multimode Lasers** 913
35.3.1 DFB and Wavelength Tuneable Lasers 913
35.3.2 Vertical Cavity Surface-Emitting Lasers (VCSELs) 915
35.3.3 High Power Lasers 917
35.3.4 Quantum Cascade Lasers 920
35.3.5 Emerging Semiconductor Materials for Modern Semiconductor Lasers 921
35.4 **Optical Amplifiers** 922
35.4.1 An Introduction to Optical Amplification 922
35.4.2 Semiconductor Optical Amplifiers (SOAs) 922
35.5 **Modulators** 923
35.5.1 Modulator Theory 923
35.5.2 Polarization-Insensitive Modulators 924
35.5.3 High-Speed High-Power QCSE Modulators 925
35.5.4 The Electro-Optic Effect 926
35.6 **Photodetectors** 927
35.6.1 Photodetector Requirements 927
35.6.2 Photodetection Theory 927
35.6.3 Detectors with Internal Gain 928
35.6.4 Avalanche Photodetectors 928
35.7 **Conclusions** 929
References 930

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_35

35.1 Introduction to Optoelectronic Devices

In this chapter, we introduce the underlying theory and operating principles of semiconductor optoelectronic devices. There exist today a plethora of optoelectronic devices that are used in a multitude of applications. These devices include sources such as light-emitting diodes (LEDs) and laser diodes, photodetectors, optical amplifiers, and optical modulators. With such devices, one can generate, modulate, detect, and switch photons in an analogous way to electrons in an electrical circuit. We begin this chapter by considering the underlying physical interactions between electrons and photons that occur in semiconductors and how they may be harnessed to produce a wide variety of devices. At the time of writing, optoelectronic devices have found their way into many different aspects of modern life whether it be the ubiquitous indicator LEDs on hi-fi systems, televisions, computers, solid-state lighting and countless other items or in the bar-code scanning systems at the supermarket, compact disk (CD), digital versatile disk (DVD) and blu-ray systems at home, the laser printer in the office, or when using a telephone or watching cable television. Over the past decade or so there has been an information explosion whereby information from all over the world can be quickly accessed by anyone with an Internet-enabled device. In all of these applications, it is a semiconductor-based optoelectronic device that forms an essential part of the system.

One of the major advantages of semiconductor devices is their small size. For example, a typical edge-emitting laser measures approximately $500 \times 250 \times 100\ \mu m^3$. Several thousands such devices can be made from a single wafer. Thus, even when packaged, these form very compact sources of coherent radiation. Other types of laser, such as the gas laser, simply cannot be compared with the semiconductor laser in terms of size, modulation rates, flexibility of application, and power consumption. Furthermore, semiconductor devices can be tailored to meet the exacting requirements of an application by simply altering the composition of the various layers forming the structure.

Of the uses of semiconductor devices outlined above, their use in telecommunications stands out as having the largest impact on modern life. Digital-based data transmission allows information to be transmitted over large distances with a much lower degradation in signal quality compared with older, analog-based systems. Optical telecommunications are ideal for use in digital systems, enabling data transmission rates well in excess of 10 Gbit/s using short optical pulses of < 100 ps in duration. Even at this exceptionally fast rate of data transmission, bit error rates of better than 1 in every 10^9 bits can be routinely achieved. Such capabilities are a direct consequence of significant research and development work that has gone into producing semiconductor devices for the emission and detection of light. In the next section, we give a brief history of this development before proceeding to discuss the key elements of physics and technology related to the device operation.

35.1.1 Historical Perspective

The development of optoelectronic devices began in the early 1960s with the development of the LED and soon thereafter, the semiconductor laser. *Holonyak* had been experimenting with the alloy GaAsP to produce visible light via spontaneous emission (SE) and was successful in producing visible (red) LEDs [35.1]. Stimulated emission was first predicted by *Albert Einstein* in his famous 1917 paper, *Zur Quantentheorie der Strahlung* (On the quantum theory of radiation) [35.2]. In 1961, the possibility of obtaining stimulated emission in semiconductors was discussed by *Bernard* and *Duraffourg* who, for the first time, derived the condition for lasing action in semiconductor materials [35.3]. This became known as the Bernard–Duraffourg condition, as discussed later in this chapter. In the following year, the first reports of lasing action in semiconductor materials were published by four independent groups, *Nathan* et al. [35.4], *Hall* et al. [35.5], *Quist* et al. [35.6], and *Holonyak* et al. [35.1]. These lasers were based upon GaAs which, with its direct-band gap, made it suitable for use as an optical source. The first devices consisted of simple p–n homojunctions which have a single interface between the n- and p-doped regions. With a wavelength of near to 900 nm, these lasers emitted in the near-infrared region of the electromagnetic spectrum. Following on from his success with LEDs, Holonyak produced visible (red) semiconductor lasers. The first semiconductor lasers had very high threshold current densities, J_{th} (defined later in this chapter), and they could only be operated under pulsed conditions. This made such devices impractical but started a period of intensive research into producing the first continuous-wave (CW) room-temperature semiconductor-based laser. After much investment into developing the growth technology of liquid-phase epitaxy (LPE), it became possible to produce a high-quality double heterostructure consisting of GaAs sandwiched between higher band gap $Al_xGa_{1-x}As$ layers. The double heterostructure brought two key advantages over the homojunction; firstly, the lower band-gap GaAs region formed a reservoir for the carriers (carrier confinement) where they could recombine across

the band gap, and secondly, the higher refractive index of GaAs with respect to $Al_xGa_{1-x}As$ provided better confinement of the optical field. These two improvements resulted in a significant reduction in J_{th} of two orders of magnitude and enabled CW operation at room temperature to be achieved for the first time [35.7, 8]. Almost 30 years later, in 2000, Alferov and Kroemer were co-recipients of the Nobel prize for physics for their pioneering work on the development of semiconductor heterostructures. In the late 1970s, further improvements in semiconductor growth technology led to the development of molecular-beam epitaxy (MBE) and vapor-phase epitaxy (VPE) which enabled increasingly thin layers to be grown reproducibly. The layer thicknesses of the order of less than 100 Å became achievable and introduced the regime in which quantum-confinement effects could be harnessed. These quantum well (QW) structures [35.9] brought about further improvements in laser performance, including a further increase in carrier confinement, narrower line width, and extended wavelength tunability for a given material composition. The splitting of the degeneracy of the valence band (VB), also resulted in a reduced density of states at the top of the VB. These improvements led to a substantially reduced J_{th}. In 1986, *Adams*, and also *Yablonovitch* and *Kane*, independently predicted that the introduction of biaxial strain into the active region of QWs would result in a further improvement in the laser characteristics [35.10, 11]. The proposed benefits of strain included a further reduction in the density of states at the top of the VB (for compressive strain) and the ability to tailor the symmetry of the carrier distribution to that of the laser beam. The introduction of tensile strain made it possible to produce a semiconductor laser with transverse magnetic (TM) polarization, something that was not previously possible. Strained-layer QWs are now included in the majority of commercially available semiconductor laser products. More recent material developments include quantum dot QD lasers, in which carriers are restricted in all three dimensions. There is a major effort to develop QD lasers with better performance than existing QW lasers with QD lasers now in commercial production. More than 20 years ago the first theoretical prediction [35.12] showed that using three-dimensionally (3-D)-confined structures with an atomic-like discrete density of states in the active region of semiconductor laser should allow the development of devices with low threshold current density and very high thermal stability. Since then, QD lasers have been demonstrated with record low room-temperature threshold current densities, $< 20\,A/cm^2$ [35.13]. Work is ongoing to produce temperature-insensitive QD lasers, particularly at 1.3 μm.

In the late 1970s and early 1980s, single-mode lasers such as the distributed feedback (DFB) laser were also developed, which have made long-haul optical communications possible [35.14]. The development of such advanced semiconductor laser devices spawned research into other semiconductor optoelectronic devices such as monolithic tunable lasers (discussed in detail by [35.15]), the semiconductor optical amplifier (SOA), optical modulators, and advanced photodetectors. These devices are discussed in the latter part of this chapter.

In parallel with the developments mentioned above, other technologies were being investigated such as the vertical-cavity surface-emitting laser (VCSEL) originally proposed by *Iga* in 1977 and demonstrated by his group in 1979 (for a comprehensive review, see [35.16]). VCSELs promise low-cost single-mode lasers which can easily be made into arrays. There remains intense activity to develop VCSELs for operation at the important telecommunications wavelength of 1.55 μm. VCSELs are discussed in later in this chapter. Other devices of note include the quantum cascade laser (QCL) developed by *Capasso* et al. in 1994 [35.17]. The development of QCLs was driven by the need to produce lasers emitting in the mid-infrared (mid-IR) ($\approx 3-12\,\mu m$) for gas sensing and environmental monitoring applications. For conventional interband lasers as discussed above, this requires narrow-band gap materials that are much less developed than materials such as GaAs and InP. QCLs get around this problem by utilizing *intraband* transitions involving only one type of carrier, for example, electrons (thus they are unipolar devices), and can be grown using conventional GaAs- or InP-based materials. In such devices, the carrier cascades through several QWs, giving rise to many photons per injected electron, quite unlike conventional interband devices. Clearly, QCLs offer great possibilities for the future development of mid-IR lasers and perhaps, with suitable materials such as GaN, they may allow emission in the near-IR (for a review see [35.18]).

Unfortunately, it is impossible to go into detail about the developments of the full range of optoelectronic devices in only one chapter of a book. However, we hope that this chapter provides a basic grounding on the relevant physics of the most important optoelectronic devices allowing the interested reader to investigate further, starting with the bibliography.

35.2 Light-Emitting Diodes and Semiconductor Lasers

35.2.1 Carrier–Photon Interactions in Semiconductors

In its simplest form, a semiconductor can be considered as a system containing two energy bands populated with electrons. In such a system, there exist three possible processes by which the electrons can move between the two bands:

1. The spontaneous emission
2. Absorption
3. Stimulated emission processes, as shown schematically in Fig. 35.1.

SE occurs when an electron in the conduction band (CB) with energy E_2 recombines with a hole in the VB at a lower energy E_1. The difference in energy (E_2–E_1) is emitted in the form of a photon. This process, as its name implies, is random and photons may be emitted in any direction with arbitrary polarization. The probability of SE is proportional to the density of electrons in the CB (n) and to the density of holes in the VB (p). In the absorption process, an electron in the VB is promoted to the CB by the absorption of a photon with an energy equal to the optical band gap. The probability of absorption is proportional to the photon density, the density of electrons in the VB, and the density of empty states in the CB. The stimulated emission process occurs when a photon of energy $E_2 - E_1$ interacts with an electron in the CB, causing it to recombine with a hole in the VB, thereby generating a photon. The photon released by this process has both identical energy, phase and momentum to that of the incident photon.

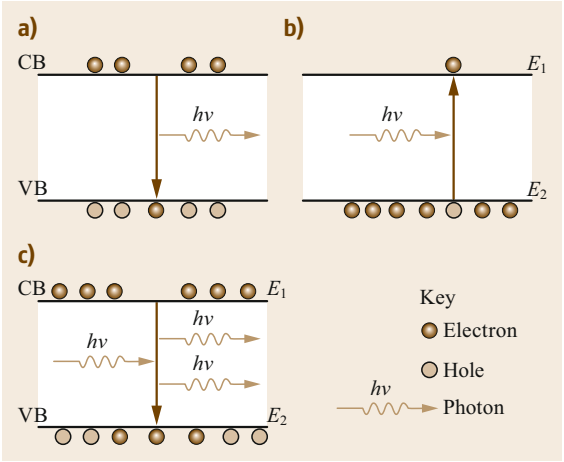


Fig. 35.1a–c The three electron–photon interactions in semiconductors: **(a)** SE, **(b)** stimulated absorption, and **(c)** stimulated emission

The probability of SE is therefore proportional to n, p and the photon density. In contrast to SE, stimulated emission produces photons that are essentially identical and therefore very pure monochromatic light is generated.

At room temperature in an undoped semiconductor, there are far fewer electrons in the CB than in the VB. As a consequence, absorption is far more likely than emission. To achieve an optical gain, the probability of stimulated emission must exceed that of absorption. By applying either an electrical current (electrical pumping) or by injecting light with photon energies greater than the transition energy (optical pumping), it is possible to inject electrons into the CB and holes into the VB. Electrical pumping is used for practical devices.

35.2.2 Direct- and Indirect-Gap Semiconductors

Semiconductors may exist in two basic forms where the band gap is either direct or indirect. Whether or not the band gap is direct or indirect has a profound influence on their suitability for use in optoelectronic devices. Figure 35.2 illustrates the difference between a direct- (e.g., GaAs) and an indirect-band gap (e.g., Ge) semiconductor. It can be seen that for the direct-band gap material, the CB energy minimum occurs at the same k-value as the VB maximum (direct). In contrast, for the indirect-band gap material, the CB minimum lies at a different k-value to the VB maximum. For any electron transition, the total energy and momentum must

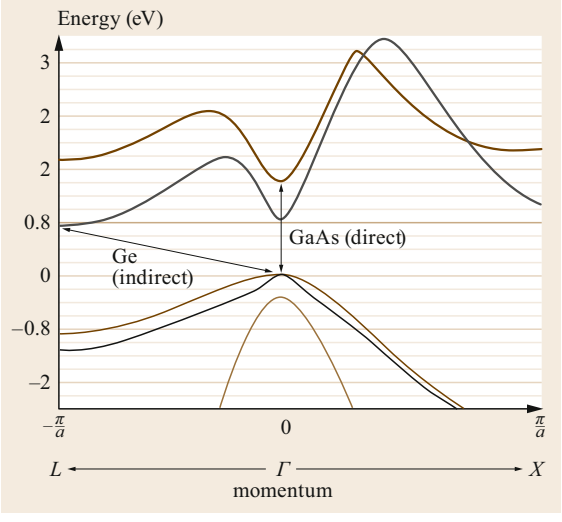


Fig. 35.2 Band structure of a direct-gap semiconductor (GaAs) and an indirect semiconductor (Ge)

be conserved. The photon wavevector has a magnitude given by $2\pi/\lambda$, where λ is typically $\approx 1\,\mu$m, while the magnitude of the electron wavevector ranges between $-\pi/\alpha$ and π/α within the first Brillouin zone, where α is the crystal lattice spacing, typically ≈ 1 Å. Clearly, therefore, the photon wavevector is much smaller than the possible electron wavevectors and, hence, if a photon interacts with an electron, the transition must occur with virtually no change in wavevector; hence, only vertical transitions are allowed. It is clear that such transitions can occur in the direct-band gap semiconductor. However, in the case of indirect semiconductors, such transitions are only possible via the interaction of a phonon. Since this becomes a three-particle interaction, the transition probability is significantly reduced. Stimulated emission is inhibited in indirect-band gap semiconductors, and it is for this reason that direct-band gap semiconductors (such as GaAs, InP, and related alloys) are primarily used to produce photoemissive devices such as LEDs and lasers. Indirect semiconductors, such as elemental germanium or silicon, are therefore impractical for the production of such devices. However, these are very effective semiconductor materials for detection, as will be discussed later in this chapter.

35.2.3 Emission and Absorption Rates and the Einstein Relations

The populations of electrons and holes in the CB and VB, respectively, are governed by both the densities of states and the Fermi–Dirac occupation probabilities. In a semiconductor, electron–electron and hole–hole scattering rates are typically ≈ 100 fs while electron–hole recombination times are ≈ 1 ns. We can therefore assume that the electrons and holes are in thermal equilibrium with themselves and can be described using a Fermi–Dirac distribution. It is useful to define two quasi-Fermi levels. These are referred to as F_c and F_v, corresponding to the energy at which the occupation probability equals 1/2 for *electrons* in the CB and VB, respectively. In thermal equilibrium, $F_c = F_v$. The corresponding energy-dependent Fermi functions are

$$f_c(E) = \frac{1}{1+\exp\left(\frac{E-F_c}{k_B T}\right)} \quad \text{and} \quad f_v(E) = \frac{1}{1+\exp\left(\frac{E-F_v}{k_B T}\right)} , \tag{35.1}$$

where k_B is the Boltzmann constant and T is the absolute temperature. By considering each of the processes illustrated in Fig. 35.1, one can obtain mathematical expressions relating each of the processes to the another. If we label states in the VB as 1 and states in the CB as 2, the *absorption rate*, r_{12}, for photons of energy $h\nu$ is given by

$$r_{12} = B_{12}\rho_v f_v (1-f_c)\rho_c P(h\nu) , \tag{35.2}$$

where ρ_c and ρ_v are the CB and VB densities of states, respectively. $P(h\nu)$ is the photon density at energy $h\nu$ and B_{12} is the Einstein coefficient for the absorption process. In a similar way, the stimulated emission rate, r_{21}, may be written as

$$r_{21} = B_{21}\rho_c f_c (1-f_v)\rho_v P(h\nu) , \tag{35.3}$$

where B_{21} is the Einstein coefficient for the stimulated emission process. For the SE rate r_{21}^{spon}

$$r_{21}^{\text{spon}} = A_{21}\rho_c f_c (1-f_v)\rho_v . \tag{35.4}$$

A_{21} is the Einstein coefficient for the SE process. Note that, since SE does not require a photon to initiate the process, r_{21}^{spon} does not depend on $P(h\nu)$. Under steady-state conditions, the total upward transition rate must equal the total downward transition rate, thus

$$r_{12} = r_{21} + r_{21}^{\text{spon}} . \tag{35.5}$$

Thus, by combining Equations (35.2)–(35.4), we obtain

$$\begin{aligned} &B_{12}\rho_v f_v (1-f_c)\rho_c P(h\nu) \\ &\quad = B_{21}\rho_c f_c (1-f_v)\rho_v P(h\nu) \\ &\qquad + A_{21}\rho_c f_c (1-f_v)\rho_v . \end{aligned} \tag{35.6}$$

In thermal equilibrium ($F_c = F_v$),

$$P(h\nu) = \frac{A_{21}}{B_{12}\exp\left(\frac{h\nu}{k_B T}\right) - B_{21}} . \tag{35.7}$$

The standard expression for black-body radiation is given by Planck's law as

$$P(h\nu) = \frac{8\pi^3 n^3 (h\nu)^2}{(hc)^3}\frac{1}{\exp\left(\frac{h\nu}{k_B T}\right)-1} . \tag{35.8}$$

Here, n is the refractive index of the semiconductor. For simplicity, here we assume that the medium is nondispersive. From these two expressions of $P(h\nu)$, we obtain the result that

$$B_{12} = B_{21} = B \tag{35.9}$$

and

$$A_{21} = B\frac{8\pi^3}{(hc)^3} n^3 (h\nu)^2 . \tag{35.10}$$

35.2.4 Population Inversion

The quantity of greatest importance to semiconductor laser operation is the net stimulated emission rate, which may simply be calculated as $r_{\text{stim}} = r_{21} - r_{12}$. Positive values of r_{stim} mean that an optical wave will grow in intensity as it travels through the semiconductor, while a negative r_{stim} implies that the optical wave would be reduced in intensity. From (35.2) and (35.3), r_{stim} can be written as

$$\begin{aligned} r_{\text{stim}} &= r_{21} - r_{12} \\ &= B_{21}\rho_c f_c\,(1-f_v)\,\rho_v P(h\nu) \\ &\quad - B_{12}\rho_v f_v\,(1-f_c)\,\rho_c P(h\nu)\,; \end{aligned} \tag{35.11}$$

therefore,

$$r_{\text{stim}} = r_{21} - r_{12} = B\rho_c\rho_v\,(f_c - f_v)\,P(h\nu)\,. \tag{35.12}$$

The ratio of SE to net stimulated emission is given by

$$\begin{aligned} \frac{r_{\text{spon}}}{r_{\text{stim}}} &= \frac{A_{12}\rho_c\rho_v f_c f_v}{B\rho_c\rho_v f_c(f_c - f_v)P(h\nu)} \\ &= \frac{A_{21} f_c f_v}{BP(h\nu)(f_c - f_v)}\,. \end{aligned} \tag{35.13}$$

Combining (35.13) with (35.10) leads to

$$\frac{r_{\text{spon}}}{r_{\text{stim}}} = \frac{8\pi^3 n^3 (h\nu)^2}{(hc)^3 P(h\nu)} \frac{f_c(1-f_v)}{(f_c - f_v)}\,. \tag{35.14}$$

Using the expressions for f_c and f_v (35.1), this can be simplified to

$$\frac{r_{\text{spon}}}{r_{\text{stim}}} = \frac{8\pi^3 n^3 (h\nu)^2}{(hc)^3 P(h\nu)} \frac{1}{\left[1 - \exp\left(\frac{h\nu - (F_c - F_v)}{k_B T}\right)\right]}\,, \tag{35.15}$$

and consequently

$$\begin{aligned} r_{\text{stim}} &= r_{\text{spon}} \frac{(hc)^3 P(h\nu)}{8\pi^3 n^3 h\nu^2} \\ &\quad \times \left[1 - \exp\left(\frac{h\nu - (F_c - F_v)}{k_B T}\right)\right]. \end{aligned} \tag{35.16}$$

From this expression, we see that, when $F_c - F_v = h\nu$, $r_{\text{stim}} = 0$. Thus, when the quasi-Fermi-level splitting equals the photon energy, the absorption and stimulated emission rates cancel. At this injection level, the semiconductor is effectively transparent. This is known as the Bernard–Duraffourg condition. For photons of energy $h\nu > F_c - F_v$, there will be absorption. However, for photons of energy $h\nu < F_c - F_v$ there is a gain. This defines the condition necessary to achieve population inversion in the semiconductor and lasing action can occur when $F_c - F_v > E_g$.

35.2.5 Gain in Semiconductors

If $\alpha(h\nu)$ is the rate of gain/loss per unit length and $P(h\nu)$ is the photon density, then the rate at which the photon density increases/decreases per unit length is simply $\alpha(h\nu)P(h\nu)$. The corresponding rate of increase of photon density per unit time is then

$$r_{\text{stim}} = P(h\nu)\alpha(h\nu)\frac{c}{n}\,, \tag{35.17}$$

where c/n is the speed of light within a semiconductor of refractive index, n. By equating our two expressions for r_{stim} (35.16) and (35.17), we find that

$$\begin{aligned} \alpha(h\nu) &= r_{\text{spon}} \frac{h^3 c^2}{8\pi^3 n^2 h\nu^2} \\ &\quad \times \left[1 - \exp\left(\frac{h\nu - (F_c - F_v)}{k_B T}\right)\right]. \end{aligned} \tag{35.18}$$

Under conditions of low pumping or at high photon energies ($h\nu > F_c - F_v$), (35.18) can be approximated to

$$\begin{aligned} \alpha(h\nu) &= -r_{\text{spon}} \frac{h^3 c^2}{8\pi^3 n^2 (h\nu)^2} \\ &\quad \times \exp\left(\frac{h\nu - (F_c - F_v)}{k_B T}\right), \end{aligned} \tag{35.19}$$

from which we have

$$\begin{aligned} r_{\text{spon}} &= -\alpha(h\nu) \frac{8\pi^3 n^2 (h\nu)^2}{h^3 c^2} \\ &\quad \times \exp\left(\frac{(F_c - F_v) - h\nu}{k_B T}\right). \end{aligned} \tag{35.20}$$

Typical plots of the absorption and SE rates can be seen in Fig. 35.3. From Fig. 35.3a, it can be seen that, with increasing pumping level (increasing quasi-Fermi-level splitting), the semiconductor moves from purely absorption and becomes a gain medium. With increasing pumping, the spread of photon energies over which the gain occurs also increases, with the peak gain moving toward higher photon energy. In Fig. 35.3b, the SE rate is plotted as a function of photon energy. From this, it can be seen that the emission spectrum is broad and corresponds to the output spectrum of an LED. With increasing pumping level, the integrated SE increases, while the peak shifts to higher photon energies. In practical devices, current-induced Joule heating may lead to a slight decrease in the band gap, which reduces the overall blue shift.

35.2.6 Density of States

As a semiconductor laser is put under the forward bias, the quasi-Fermi levels in the CB and VB move toward

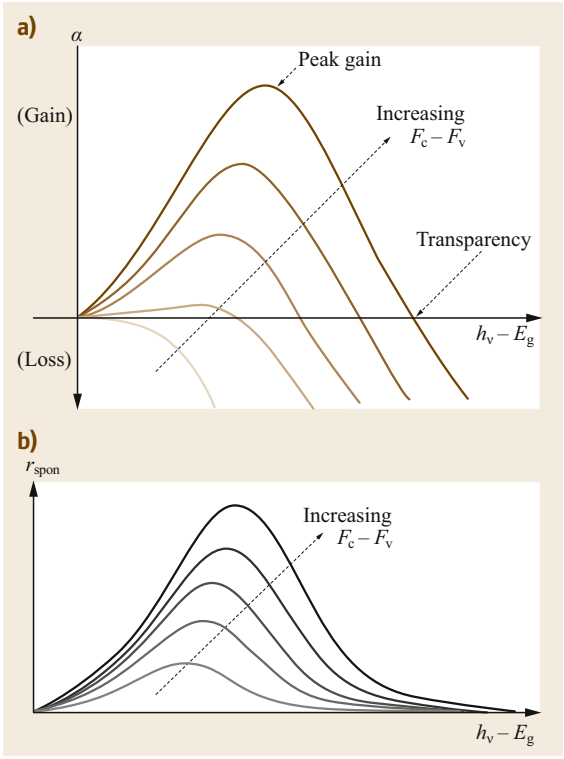


Fig. 35.3 (a) Gain/loss and (b) SE spectra as a function of increasing quasi-Fermi-level splitting

higher energies. The separation of the quasi-Fermi levels is dependent on the density of states $\rho(E)$ in the CB and VB. At laser transparency, where the quasi-Fermi-level separation equals the photon energy, the transparency carrier density, n_{tr}, and hence the threshold carrier density, n_{th}, depend upon the number of states in a given energy range. By reducing the number of available states, n_{th}, and consequently I_{th} (J_{th}), can be lowered. The density of states can be modified by changing the band structure. This can be achieved by reducing the dimensionality, thereby limiting the motion of the carriers, and also by introducing strain into the active region of the laser, as shall be described shortly.

Dimensionality

In conventional bulk semiconductor materials, the carriers have unrestricted motion in all directions and can be described by a 3-D density of states

$$\rho_{3\text{-D}} \propto m_{\text{eff}}^{\frac{3}{2}} \sqrt{E} \,, \tag{35.21}$$

where m_{eff} is the effective mass. This is illustrated in Fig. 35.4a, where the shaded region indicates the carrier population as a function of energy. In such (bulk) devices there are relatively few states near to the band

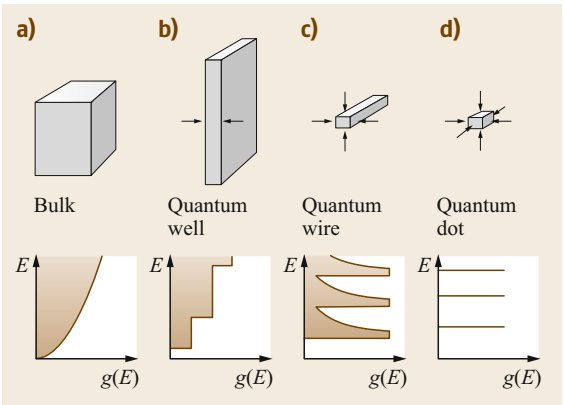


Fig. 35.4a–d Density of states in systems of differing dimensionality: (a) bulk, (b) QW, (c) quantum wire, and (d) QD

edges and hence the quasi-Fermi levels move toward higher energies (where there is a larger concentration of available states) in order to achieve the sufficient gain. This results in a broad output spectrum corresponding to radiative transitions over a wide energy range. By restricting the motion of the carriers along one dimension, as in QW structures, the density of states becomes independent of energy, with many states at the band edges. In a QW, the density of states is dependent only on the effective mass, m_{eff}, and the QW width, L_z as

$$\rho_{2\text{-D}} \propto \frac{m_{\text{eff}}}{L_z} \,. \tag{35.22}$$

In bulk structures, at the Brillouin zone center of the VB, the light- and heavy-hole bands are degenerate. Consequently, as the laser is pumped, states that do not contribute to the gain of the lasing mode also become filled. In a QW, the energy states form discrete levels. For an infinite square well, the confinement energy of the n-th energy level E_n is dependent on the QW width L_z and the growth-direction effective mass m_{eff}^z, and at the zone center can be simply expressed analytically as

$$E_n = \frac{\hbar^2}{2m_{\text{eff}}^z} \left(\frac{n\pi}{L_z} \right)^2 . \tag{35.23}$$

Thus, in a QW, the degenerate heavy- and light-hole bands become split (due to the difference in m_{eff}^z), resulting in a further reduction in the density of states at the VB Brillouin zone center. Therefore, in a QW laser, a higher proportion of the filled states can contribute to the gain of the lasing mode.

Further improvements in carrier confinement can be achieved by moving toward quantum wire or even QD structures. In quantum wire lasers, the carriers

are further confined, allowing the freedom of movement along one direction only (Fig. 35.4c). In principle, such structures can be realized, for example, by growth along lithographically defined V-grooves and lasers have been successfully produced in this manner [35.19], although the complexity of the processing has meant that they have yet to become a commercial product. In a QD laser, the carriers are confined still further so that they are restricted in all three dimensions (Fig. 35.4d). Thus, an ideal QD is analogous to an atom. By completely restricting the carrier motion, one may theoretically expect to achieve temperature-insensitive operation [35.12]. Initial attempts to produce QD lasers concentrated on lithographic processes, but it was not until the successful demonstration of self-assembled growth (the Stranski–Krastanov technique) that QD lasers became a reality. Record low room-temperature threshold current densities ($< 20\,\mathrm{A/cm^2}$) have been achieved using QD lasers [35.13]. However, due to the low gain, long lasers are required which increases the threshold current.

The Influence of Strain

A detailed discussion of strain in semiconductors is given in Chap. 20 of this book. However, due to the enormous importance of strain in optoelectronic devices, in this section, we discuss the influence of strain on device characteristics. For many years, manufacturers of semiconductor lasers struggled to produce dislocation- and strain-free material. In 1986, it was suggested independently by *Adams* and by *Yablonovitch* and *Kane* that the incorporation of strain into the active layer of semiconductor lasers would lead to improved performance [35.10, 11]. Many of these benefits have since been demonstrated [35.20–22]. The majority of the lasers currently manufactured in the world today are based upon strained layers.

In a strained-layer structure, the lattice constant of the epilayer material α_{epi} is different from the lattice constant of the substrate α_{sub} on which it is grown. When a thin epilayer of this material is grown onto the much thicker substrate, the lattice constant of the grown epilayer becomes equal to that of the substrate, α_{sub}, and the epilayer is in a state of biaxial stress. The resulting net strain in the (x–y) plane of the layer ε is given by

$$\varepsilon_{\parallel} = \varepsilon_{xx} = \varepsilon_{yy} = \frac{(\alpha_{\mathrm{sub}} - \alpha_{\mathrm{epi}})}{\alpha_{\mathrm{epi}}} , \qquad (35.24)$$

where it is assumed that the substrate is much thicker than the epilayer. Hence, $\alpha_{\mathrm{epi}} > \alpha_{\mathrm{sub}}$ results in compressive strain, while $\alpha_{\mathrm{epi}} < \alpha_{\mathrm{sub}}$ produces tensile strain. Figure 35.5 shows the three possible types of epilayer growth:

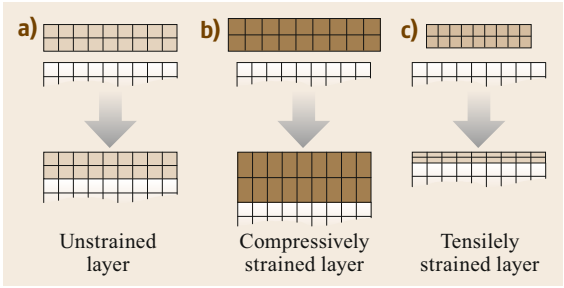


Fig. 35.5a–c Illustration of (a) unstrained, (b) compressively strained, and (c) tensile-strained layers. For the strained layers the epilayer expands (contracts) in the plane while expanding (contracting) in the growth direction, corresponding to compressive (tensile) strain respectively

1. Compressively strained
2. Tensile strained
3. Unstrained.

The sign convention used here is that positive strain indicates that the layer is under biaxial compression, while a negative strain corresponds to a layer being in a state of biaxial tension. For compressive strain, when the lattice constant of the epilayer is reduced in the plane, the lattice constant in the growth direction, $\alpha_{\perp}$, is increased. The strain in the growth direction $\varepsilon_{\perp}$ ($= \varepsilon_{zz}$) is governed by Poisson's ratio σ and is given by

$$\varepsilon_{\perp} = \varepsilon_{zz} = -\frac{2\sigma\varepsilon_{\parallel}}{1-\sigma} . \qquad (35.25)$$

In tetrahedral semiconductors, $\sigma \approx 1/3$; therefore, upon substituting this into (35.25), it can be seen that $\varepsilon_{\perp} \approx -\varepsilon_{\parallel}$. In order to understand how strain affects the band structure, it is useful to split the strain into an axial component $\varepsilon_{\mathrm{ax}}$

$$\varepsilon_{\mathrm{ax}} = \varepsilon_{\perp} - \varepsilon_{\parallel} \approx -2\varepsilon_{\parallel} , \qquad (35.26)$$

and a hydrostatic component, $\varepsilon_{\mathrm{vol}}$ ($= \Delta V/V$, the fractional change in volume of the epilayer)

$$\varepsilon_{\mathrm{vol}} = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} \approx \varepsilon_{\parallel} . \qquad (35.27)$$

The axial and hydrostatic strain components affect the band structure as described in the next section. In direct-band gap semiconductors, while strain has only a small effect on the CB, it greatly alters the VB structure.

The maximum layer thickness that can be grown under a state of strain is determined by the critical thickness h_{c}. Above this thickness, it becomes energetically favorable to relieve the strain energy through the formation of dislocations. Thus, in order to grow high-quality (high-Q) strained material, it is necessary to

keep the thickness of each layer below h_c. As the strain in a layer is increased, the corresponding critical thickness decreases. The maximum strain–thickness product for InGaAs/GaAs materials is ≈ 200 Å% [35.23]. It is however possible to use alternating layers of compressive and tensile strain to produce zero-net-strain structures that can prevent the formation of dislocations. Strained-layer semiconductor lasers are often fabricated with strain-compensated wells and barriers when there are many strained QWs, thereby providing a large strain in the wells while reducing the overall strain in the structure.

The introduction of strain into the active region of semiconductor diode lasers brings about several changes in the band structure. The initial predictions of Adams and Yablonovitch and Kane concentrated on the benefits of compressive strain. It has since been shown that tensile strain can also enhance laser performance [35.24–26]. In the previous section, it was shown that the strain in an epilayer can be split into purely axial and hydrostatic components, ε_{ax} and ε_{vol}, respectively. The axial component of the strain splits the cubic symmetry of the semiconductor and hence the degeneracy of the heavy- and light-hole VBs at the Brillouin zone center is lifted (it should be noted that the VB will already be nondegenerate in a QW due to the effects of quantum confinement). In addition, the VB dispersion becomes anisotropic. In the case of compressive strain, the heavy-hole band is moved to higher energy with respect to the light-hole band, while the opposite is true for tensile strain. Also, for compressive strain, the *in-plane* (x–y) mass becomes lighter than the mass in the growth (z) direction. Again, the opposite is true for tensile strain. The hydrostatic component of the strain causes an increase in the mean band gap for compressive strain and a decrease in the mean band gap for tensile strain. These effects on the band structure of a strained *bulk-like* material are shown schematically in Fig. 35.6.

Due to the much reduced in-plane highest hole band mass, the density of states in a compressively strained laser is lower than that for an unstrained device. As a result, n_{tr} is decreased, which not only lowers the amount of unwanted SE [35.27] but also reduces carrier-density-dependent nonradiative processes such as Auger recombination [35.28]. Also, due to the increased heavy- and light-hole band splitting, carrier spillover into higher subbands can be substantially reduced. Consequently, compressively strained lasers have been shown to have much lower threshold current densities than similar unstrained devices [35.20, 29].

To understand the improvement in the performance of tensile-strained lasers, it is necessary to examine the character of the VB. The heavy-hole band consists

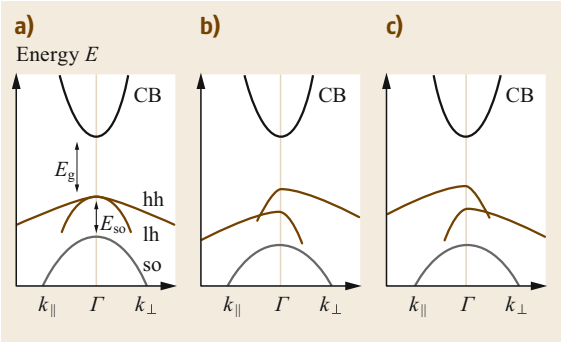


Fig. 35.6a–c Effects of strain for a bulk-like direct-band gap semiconductor for **(a)** unstrained, **(b)** tensile strain, and **(c)** compressive strain

of $1/2\,x$-like states and $1/2\,y$-like states, but contains no states with z-like character. In contrast, the light-hole band consists of $1/6\,x$-like states, $1/6\,y$-like states, and $2/3\,z$-like states. y-like states contribute to the transverse electric (TE) gain where the electric field oscillates across the facet (y-axis) and z-like states contribute to the TM gain, where the electric field oscillates along the growth direction. x-like states contribute only to SE, as the electric field vector lies in the propagation direction of the lasing mode. This is illustrated in Fig. 35.7. In bulk, unstrained QW and compressively strained QW lasers, the lasing mode is always TE-polarized as $g_{th}(\mathrm{TE}) < g_{th}(\mathrm{TM})$. With the inclusion of tensile strain in the QWs of semiconductor lasers it became possible for the first time to produce a device with TM polarization [35.25].

From Fig. 35.6, it can be seen that the in-plane mass of a tensile-strained laser is relatively large, resulting in an increased density of states. However, from Fig. 35.7, it can be seen that the top of the VB in a tensile-strained device contains $2/3\,z$-like states, which can contribute to the TM gain. Only one-third of the injected carriers will *not* produce gain in the lasing mode, compared with two-thirds in unstrained (bulk-like) and half in compressively strained QW devices. Hence, tensile-strained lasers can also exhibit a lower threshold current [35.25]. However, if only a moderate amount of tensile strain is incorporated into the QW, the threshold current will increase due to the heavy- and light-hole subbands becoming degenerate.

35.2.7 Optical Feedback in a Fabry–Perot Laser Cavity

The probability of stimulated emission depends not only on maintaining population inversion, but also on the photon density. In a laser, the photon density is sustained by the use of optical feedback. In a gas laser,

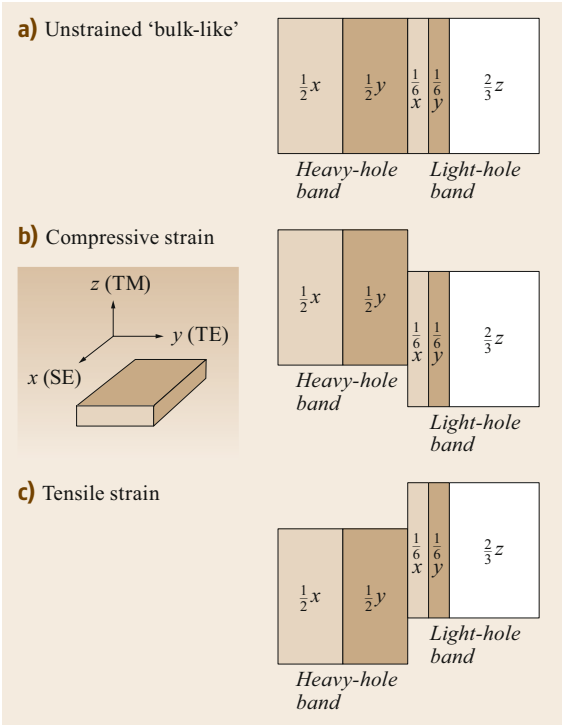


Fig. 35.7a–c Schematic illustration of the VB alignment for **(a)** unstrained bulk-like, **(b)** compressive strain, and **(c)** tensile strain

optical feedback is realized by the use of partially metallized mirrors, which form a Fabry–Perot cavity around the gain medium. When the light reaches the end of the cavity, a certain fraction (R) is reflected back into the gain medium, where it again becomes available to stimulate the emission of further photons. In a standard edge-emitting semiconductor laser, the same principle is used but the mirrors are formed by cleaving the semiconductor along a crystallographic plane (most commonly the (110) plane), forming laser facets. This is illustrated in Fig. 35.8. By using as-cleaved facets, the resulting reflectivity, R, is typically $\approx 30\%$ due to the refractive-index step between the laser material and (usually) air. This can be either increased or decreased with suitable facet coatings. Semiconductor lasers usually have roughened sides as a result of the fabrication process. This has the desirable effect of reducing the reflectivity of the sides of the laser and hence reduces the probability of generating unwanted transverse modes.

The application of optical feedback using a Fabry–Perot cavity has the effect of producing Fabry–Perot resonances or modes in the output spectrum of the laser. From Fig. 35.8, it can be seen that the optical field will form a standing wave when the cavity length is equal to

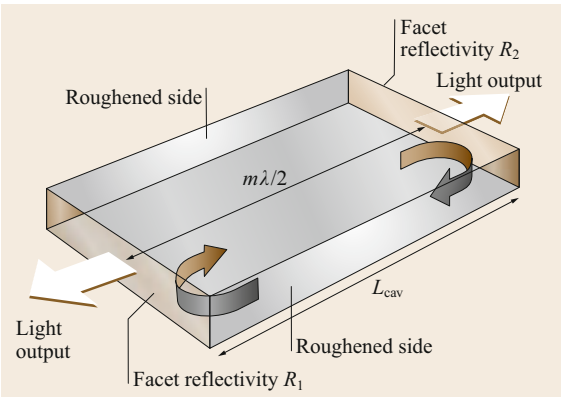


Fig. 35.8 Schematic diagram of a Fabry–Perot edge-emitting laser

an integer number of half-wavelengths, given by

$$\frac{m\lambda}{2} = nL_{cav} \quad m = 1, 2, 3, \ldots , \tag{35.28}$$

where m is the (integer) number of half-wavelengths, λ is the lasing wavelength (as measured), n is the refractive index of the laser cavity, and L_{cav} is the cavity length (typically $250\,\mu m \leq L_{cav} \leq 1500\,\mu m$). It can be seen that the maximum wavelength that the cavity can support is $\lambda = 2nL_{cav}$. If (35.21) is rearranged for λ and differentiated with respect to m, the mode spacing, $\Delta\lambda$ is found to be

$$\Delta\lambda = \frac{\lambda^2}{2nL_{cav}} . \tag{35.29}$$

Thus, the possible modes of a semiconductor laser consist of a comb of many hundreds of delta functions spaced $\Delta\lambda$ apart, as illustrated in Fig. 35.9. For a typical 1.5 μm laser with $L_{cav} = 500\,\mu m$ and $n \approx 3.2$, $\Delta\lambda \approx 0.7\,nm$. The wavelength at which lasing occurs is therefore determined by both the material gain curve and the mode spacing. The mode which lies closest to the peak of the material gain curve at threshold will lase (Fig. 35.9), although other modes near to the gain peak may also be present in the output spectrum of an edge-emitting Fabry–Perot laser. With increasing bias however, the relative strengths of these modes will decrease with respect to the dominant lasing mode. This produces a very narrow and spectrally pure output, in stark contrast with the SE spectrum of an LED.

35.2.8 Wave Guiding

As discussed earlier in this chapter, the stimulated emission process is not only dependent on the carrier density, but also on the photon density within the laser

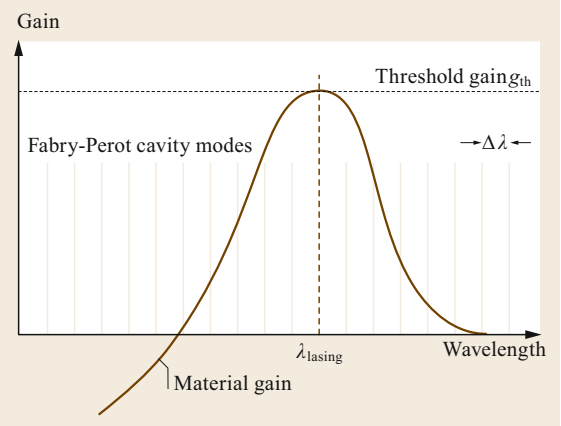


Fig. 35.9 Illustration of gain in a semiconductor laser. The cavity mode that lies closest to the material gain peak at threshold will lase

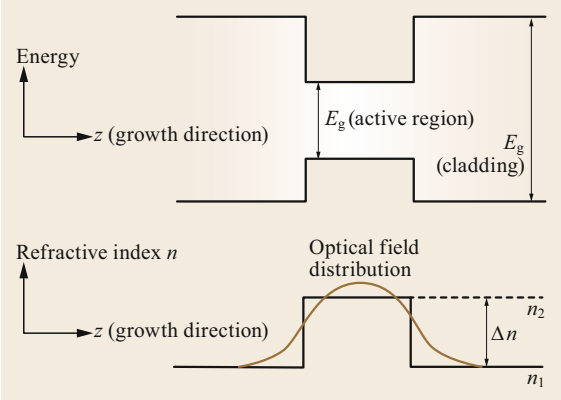


Fig. 35.10 Wave guiding in a double heterostructure

cavity. In particular, it is a requirement that the photon density is maintained as high as possible so that the probability of stimulated emission is maximized while still allowing sufficient radiation to escape. As the optical field propagates along the laser cavity, it can extend in both the growth (z) and transverse (y) directions. Hence, it is necessary for the optical field to be confined along two directions. In a semiconductor laser, wave guiding can be accomplished by two methods:

- Index guiding
- Gain guiding.

Index guiding is achieved by surrounding the active region with a material of larger band gap and, hence, lower refractive index. In the *growth* direction of an edge-emitting double-heterostructure semiconductor laser, the optical field is always index guided. Figure 35.10 shows an example of index guiding in a typical double-heterostructure laser together with the associated refractive-index profile in the growth direction. Index guiding confines most of the optical field within the active region due to total internal reflection at the cladding heterojunctions with only a small fraction of the field evanescently decaying into the higher band gap layers. The degree of confinement depends upon the refractive-index step between the active region and the surrounding layers.

In the transverse direction, the optical field can be confined by either index- or gain-guiding mechanisms. Gain guiding arises from a weak current-induced change in refractive index across the active region caused by the nonuniform carrier distribution. Index-guided structures can be considered as either strongly index-guided or weakly index-guided, corresponding to either a large or small refractive-index step in the transverse direction. Gain-guided structures are easier to fabricate than index-guided structures but are less effective in confining the optical field when compared with index-guided structures [35.30].

35.2.9 Carrier Confinement

In addition to confining the optical mode, laser characteristics can be greatly enhanced by improving the confinement of the electrons and holes, thereby increasing the probability of radiative recombination. The first laser structures [35.1, 4–6] were based upon single p–n junctions (homojunctions). In such structures, the injected electrons and holes recombine close to the junction. However, due to the nonuniform carrier distribution, as the electrons diffuse out of the gain region, there is a variation in gain with position. In addition, due to the small difference in refractive index at the edge of the gain region, a large amount of the optical field spread into absorbing regions. Thus, the combination of a poor carrier and field confinement gives rise to large threshold current densities $\approx 10^5\,\mathrm{A/cm^2}$ in homojunction lasers. The development of the double-heterostructure laser [35.7, 8] brought about a large decrease in J_{th} of two orders of magnitude. This results from the greatly improved carrier confinement due to the difference in the band gap between the active and cladding layers, which forms a potential barrier to the carriers. A further improvement in carrier confinement is obtained with QW structures. In a QW laser, the injected carriers rapidly diffuse into the thin ($\approx 25-150\,\text{Å}$) QW layers [35.9]. Due to the large potential barriers between the well and barrier material, it is difficult for the carriers to subsequently escape. In addition to improved carrier confinement, QW lasers offer many other benefits over bulk lasers as discussed later in this chapter. Figure 35.11 illustrates the improvement in carrier confinement obtainable on going from ho-

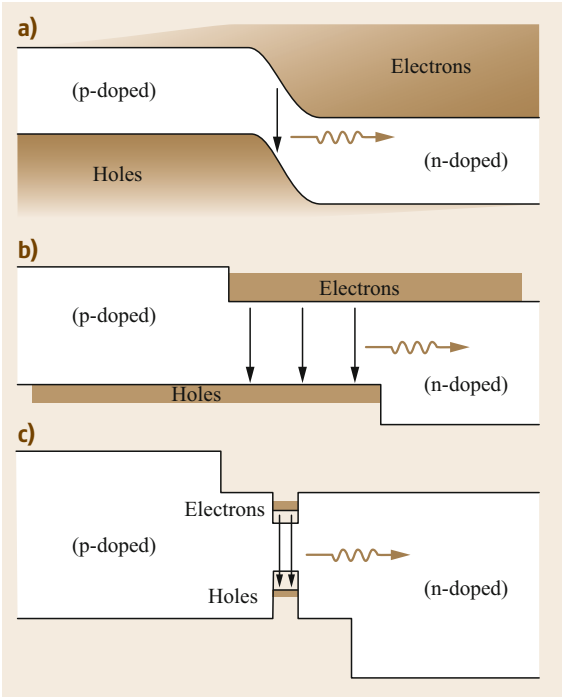


Fig. 35.11a–c Advances in carrier confinement **(a)** homojunction, **(b)** double heterostructure, and **(c)** QW

mojunctions to double heterostructures and QW laser structures.

35.2.10 Current Confinement

The threshold current of semiconductor lasers can be further decreased by reducing the area of the device that is being electrically pumped. Reducing the threshold current can also suppress current-heating effects and extend the operating lifetime of the laser. Current confinement can also help to produce a stable fundamental mode. There are several fabrication techniques and associated laser structures designed to increase current confinement [35.30]. The ridge structure, as shown in Fig. 35.12a, is a weakly index-guided type of laser. In this structure, following the epitaxial growth of the complete layer structure, the wafer is etched to produce a ridge that is $\approx 2-10\,\mu m$ wide. The wafer is then coated with a dielectric (usually SiO_2) and metallized. The weak index guiding is provided by the large refractive-index step between the active region material and the dielectric. The current is directly injected into the ridge and therefore travels through a relatively narrow region of the device. The two main other structures are strongly index-guided buried heterostructure (BH) devices. In double-channel planar buried heterostructure (DCPBH) lasers (Fig. 35.12b), following the growth of the layer structure, two channels, $\approx 3\,\mu m$ deep and $\approx 10\,\mu m$ wide, are chemically etched, forming a mesa. After cleaning, n- and p-doped current blocking layers are grown followed by the cladding and contacting layers. The current confinement is provided by the reverse-biased p–n junctions that surround the mesa.

The final structure considered is the semi-insulating planar buried heterostructure (SIPBH), as can be seen in Fig. 35.12c. In a similar way to the DCPBH devices, SIPBH lasers are based upon an etched mesa. In an SIPBH device, the current confinement is provided by semi-insulating layers that surround the mesa. In these structures, after wet-chemically etching the wafer to form the mesa (typically $1-2\,\mu m$ wide and $2-3\,\mu m$ high), an Fe-doped layer is grown, followed by an n-doped antidiffusion layer, p-doped cladding and contacting layers. The resistivity of the Fe-doped regions surrounding the mesa is typically $\approx 10^8\,\Omega cm$ thereby ensuring that the current only flows through the active region. Buried heterostructure lasers tend to have lower threshold currents than ridge structures due to the combination of excellent current and optical field confinement. For high-speed applications such as telecommunications, SIPBH devices outperform DCPBH lasers due to their much lower parasitic capacitance.

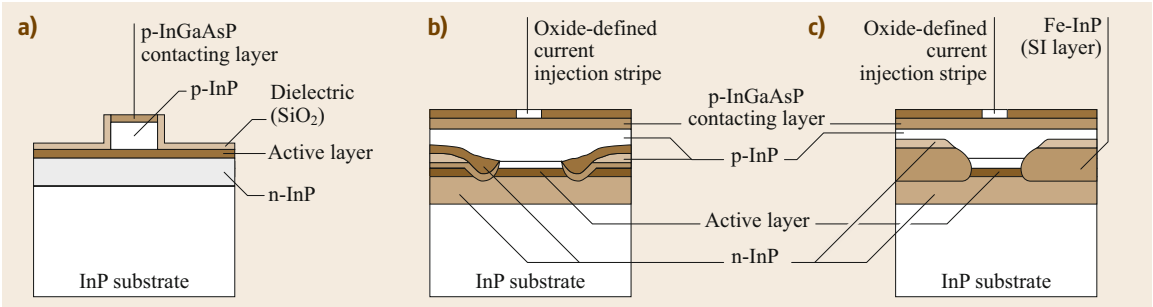


Fig. 35.12a–c Schematic diagrams of typical long-wavelength InP lasers **(a)** ridge structure, **(b)** double-channel planar buried heterostructure (DCPBH), and **(c)** semi-insulating planar buried heterostructure (SIPBH)

35.2.11 Laser Threshold and Efficiency

In Sect. 35.2.4, the condition for material transparency was derived. In a laser structure, the threshold is reached when the total gain is equal to the optical losses as the wave travels along the cavity plus the (useful) loss of light from the facets (mirror loss). If we consider the laser cavity shown in Fig. 35.8, the threshold gain condition, g_{th} is given by

$$g_{th} = \frac{1}{\Gamma}[\alpha_m + \alpha_i] = \frac{1}{\Gamma}\left[\frac{1}{2L_{cav}} \ln\left(\frac{1}{R_1 R_2}\right) + \alpha_i\right], \quad (35.30)$$

where R_1 and R_2 are the facet reflectivities, α_m is the mirror loss, and α_i is the internal loss per unit length in the *entire* structure, including scattering losses, free-carrier absorption, and inter-VB absorption [35.31]. Γ is the optical confinement factor, which is equal to the fraction of the optical field intensity that overlaps the active region. This therefore accounts for the fact that not all of the optical field will give rise to the gain. Note that g_{th} is the threshold gain per unit length of the device. For a typical 1000 µm long 1.5 µm QW device ($\Gamma \approx 0.02$) with as-cleaved facets ($R_1 = R_2 \approx 0.3$) and $\alpha_i = 10\,\mathrm{cm}^{-1}$, $g_{th} \approx 1100\,\mathrm{cm}^{-1}$. The carrier density required to reach g_{th} is termed n_{th}, and the corresponding current (density) is termed I_{th} (J_{th}). Figure 35.13 shows how the light output and carrier density increase as current is injected into the laser structure. Below threshold, increasing the current causes the Fermi-level splitting to increase, thereby increasing the electron and hole (carrier) densities in the active region of the laser. In this regime, the laser is acting as an LED and SE is produced. However, the electrical-to-optical conversion efficiency is relatively low ($\approx 1\%$). In practice, non-radiative recombination processes (as discussed in the next section) cause the carrier density to have a non-linear dependence on current below the laser threshold. At threshold, the stimulated lifetime decreases, causing each additional injected carrier to undergo stimulated emission. These carriers rapidly transfer to the VB with the production of a photon via stimulated emission giving rise to an abrupt increase in the light output. Thus, at threshold, the carrier density becomes pinned at n_{th} since every additional carrier quickly undergoes stimulated emission. Hence, above threshold, the intrinsic *differential* quantum efficiency, η_d, of the lasing process can be as high as 100% in an ideal laser. However, effects including carrier leakage, absorption, and self-heating can reduce the overall measured differential quantum efficiency.

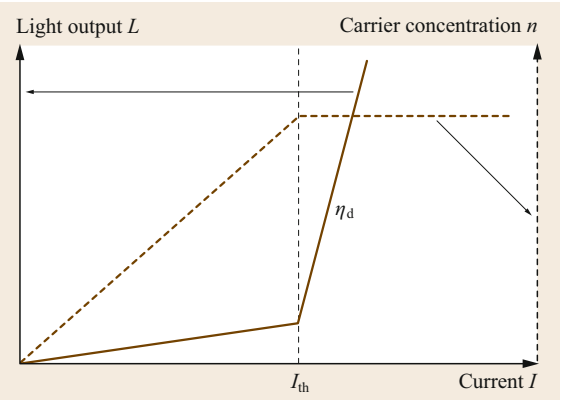


Fig. 35.13 Variation of the light output (*solid curve*) and carrier density (*dashed curve*) as a function of the injection current. Above threshold, the light output increases abruptly while the carrier density becomes pinned

The differential quantum efficiency, η_d, is defined as the ratio of the incremental number of photons emitted from the facets to the incremental number of carriers injected into the laser (above threshold), thus

$$\eta_d = \frac{dL}{dI}\frac{e}{h\nu}, \quad (35.31)$$

where L is the optical power emitted from both facets, e is the electronic charge, and $h\nu$ is the photon energy (the lasing energy). Thus, η_d can simply be determined from the light–current characteristic (providing that all of the stimulated light is collected), η_d can more generally be defined from

$$\eta_d = \eta_i\left(\frac{\alpha_m}{\alpha_i + \alpha_m}\right), \quad (35.32)$$

η_i is the *internal* quantum efficiency and accounts for the fact that, in a real laser, not all of the injected carriers will result in the production of a photon. This may be due to inefficient injection of carriers into the active region of the laser (dependent on both the laser materials and the geometry of the laser). It is convenient to rewrite (35.32) in terms of the inverse efficiency so that

$$\frac{1}{\eta_d} = \frac{1}{\eta_i}\left(\frac{\alpha_i + \alpha_m}{\alpha_m}\right) = \frac{1}{\eta_i}\left(\frac{\alpha_i}{\alpha_m} + 1\right) = \frac{1}{\eta_i}\left[\frac{2\alpha_i L_{cav}}{\ln\left(\frac{1}{R_1 R_2}\right)} + 1\right]. \quad (35.33)$$

If R_1 and R_2 are known (easily calculated), by measuring η_d for lasers with several different cavity length taken from the same wafer and plotting a graph of $1/\eta_d$ versus L_{cav}, one can obtain a value for both η_i (from the intercept) and α_i (from the slope/intercept).

From Fig. 35.13, it is clear that the *overall* efficiency of the laser will depend on both the differential quantum efficiency η_d, which we wish to maximize, and the threshold current I_{th}, which we would like to minimize. In the discussion thus far, the threshold current has been described only in terms of an SE current. In practice, there exist many other recombination paths that contribute to the laser threshold. These are discussed in the next section.

35.2.12 Carrier Recombination Processes

When carriers are injected into a semiconductor laser, there are many possible recombination paths. In an ideal semiconductor laser, above threshold, the carrier density pins, as illustrated in Fig. 35.13. Therefore, above threshold, any recombination path which depends upon n will also become pinned. However, the threshold current itself is determined by the different mechanisms by which carriers recombine in a real laser structure. In this section, the important radiative and nonradiative recombination mechanisms are briefly discussed. The threshold current of a semiconductor laser is defined as the current required to provide enough gain to reach the threshold gain level g_{th}. Due to differences between laser structures, it is frequently more useful to compare the threshold current density J_{th} between devices, which accounts for the overall area of the device that is being pumped, thereby allowing a useful comparison to be made between different device structures. In this chapter, both are used interchangeably when discussing laser characteristics. At threshold, the total current I can be expressed as the sum of the current paths as

$$I = eV(An + Bn^2 + Cn^3) + I_{leak}\,, \qquad (35.34)$$

where e is the electron charge, V is the active region volume, and n is the carrier density (where we assume that the electron (n) and hole (p) densities are equal, $n = p$).

The An term is due to recombination at defects and is considered to be negligible in high-quality material. It does however make significant contributions to some laser materials, most notably InGaAsN/GaAs where it can account for up to 50% of I_{th} even in the best 1.3 μm devices [35.32].

The Bn^2 term corresponds to the radiative recombination of electrons and holes giving rise to SE, as observed from LEDs, and semiconductor lasers operated below threshold. Radiative recombination occurs (for interband processes) when an electron in the CB recombines with a hole in the VB, resulting in the SE of a photon as illustrated in Fig. 35.14. For lasing to occur, it is necessary to have spontaneous recombination to initiate the stimulated emission process. However, because SE is a random process, emitting photons in all directions over a broad energy range, only a small fraction of the photons can couple with the laser gain. Hence, for efficient laser operation, it is desirable that the amount of unwanted SE is reduced.

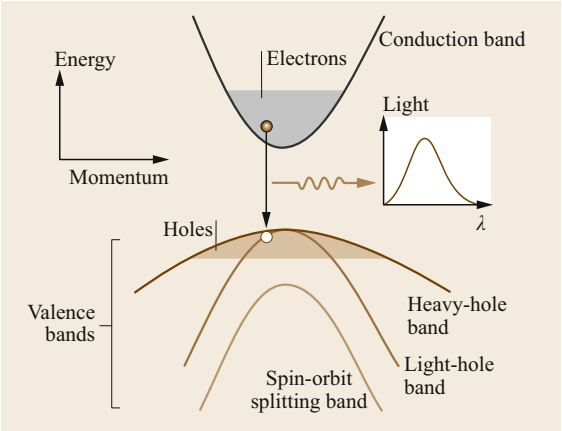


Fig. 35.14 Spontaneous emission in a bulk semiconductor

The Cn^3 term describes *nonradiative* Auger recombination. In an Auger recombination process, the energy of a recombining electron–hole pair that would normally produce a photon is instead given to a third carrier (electron or hole) which is excited further into its respective band. Thus, in an Auger recombination process, three carriers are involved, the initial electron–hole pair plus an additional electron or hole. The two main Auger recombination processes are illustrated in Fig. 35.15, where the closed circles represent electrons and open circles represent holes. In the conduction–hole–conduction–conduction (CHCC) process, the energy and momentum of a recombining CB electron and VB hole are transferred directly to a second CB elec-

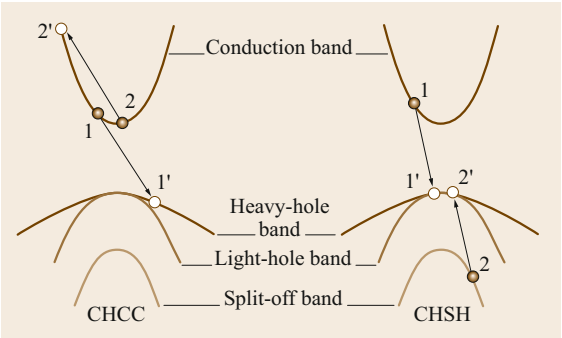


Fig. 35.15a,b The two primary Auger recombination processes: **(a)** CHCC producing hot electrons and **(b)** CHSH producing hot holes

tron, which is then excited into a higher energy CB state. In the direct conduction–hole–spin–hole (CHSH) process, the energy and momentum are transferred to a second VB hole, which is then excited into the spin split-off band. In addition to these direct Auger recombination processes, there may be phonon-assisted processes whereby the conservation of momentum and energy of the carriers is relaxed by the emission or absorption of a phonon. These are however of less importance in 1.3 and 1.5 μm lasers [35.33]. It should be noted that Auger-generated carriers are very energetic and initially lie many k_BT more than E_g above their respective band edges. While these carriers quickly scatter back to thermal equilibrium, there is a finite probability that they will scatter from the active region into the cladding layers, thus forming an additional Auger-generated leakage current.

Auger recombination processes are very sensitive to the band gap, E_g, and are severely detrimental to laser performance. For emission wavelengths below 1 μm, the probability of an Auger recombination process occurring is small due to the near-vertical transitions that are required on an energy–momentum diagram, which decreases the probability of the process occurring (*inset* Fig. 35.16). However, as E_g decreases, the transitions become less vertical and the probability of finding a state for the *hot* Auger carrier increases. Auger recombination, therefore, begins to influence the performance of semiconductor lasers and by 1.5 μm it dominates their behavior at room temperature, forming ≈ 80% of the threshold current [35.28]. This is illustrated in Fig. 35.16 where the normalized variation of the measured threshold current with lasing energy was determined by performing high-pressure measurements on several lasers over the nominal wavelength range 1.3–1.6 μm. With increasing lasing energy (band gap), the threshold current decreases strongly due to the lowered contribution of Auger recombination. Thus, long-wavelength lasers are more susceptible to Auger recombination. Auger recombination is also largely responsible for the high-temperature sensitivity of the threshold current of 1.3 and 1.5 μm semiconductor lasers. In Fig. 35.17, we plot the temperature dependence of the threshold current (*circles*) for a 1.3 μm InGaAsP laser. The squares correspond to the measured radiative current as determined from SE measurements [35.28]. At low temperature, the threshold current follows the radiative current very closely, however, above ≈ 160 K, the threshold current increases strongly due to thermally activated Auger recombination. This is primarily due to the n^3 dependence of the Auger current and also because the Auger coefficient C is thermally activated [35.34]. Suppressing Auger recombination is challenging and work is ongoing on new semiconductor materials such as the III–V-bismides which have the potential to reduce or eliminate Auger recombination processes through band structure engineering, as discussed later in this chapter.

I_{leak} is associated with heterobarrier leakage and current spreading around the active region. With increasing temperature there is an increasing probability that electrons in the CB and holes in the VB will have enough thermal energy to escape from the QWs. This is termed heterobarrier leakage. Here, these carriers may

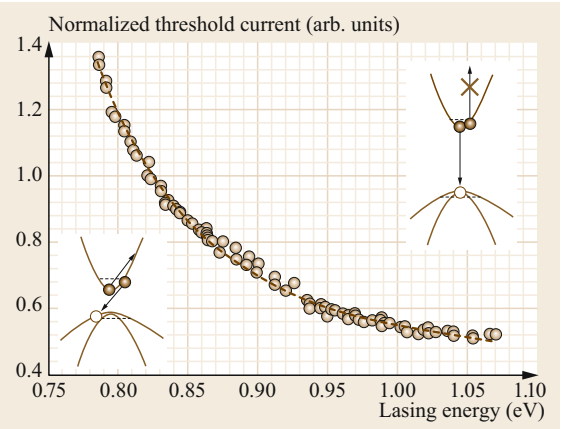


Fig. 35.16 Variation of the threshold current with lasing energy over the wavelength range 1.2–1.6 μm. With increasing wavelength (decreasing energy), the threshold current increases swiftly due to the increased probability of Auger recombination, as illustrated in the *insets*

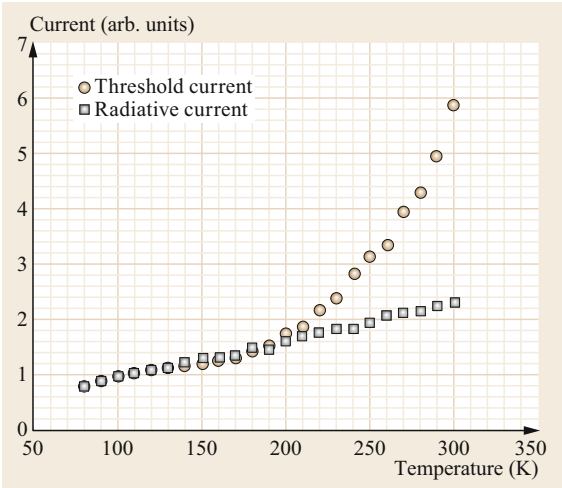


Fig. 35.17 Threshold current (*circles*) and radiative current (*squares*) versus temperature for a 1.3 μm InGaAsP laser. The threshold current increases more strongly with temperature than the radiative current due to the thermally induced onset of Auger recombination

recombine radiatively (at shorter wavelengths) or non-radiatively. If they have sufficient energy, they may even escape into the cladding layers. The probability of carriers escaping into the cladding layers depends upon the position of the quasi-Fermi levels with respect to the cladding band edge, as illustrated in Fig. 35.18. With increasing temperature, the thermal spread of carriers broadens and the population of unconfined carriers at threshold increases, thereby increasing the threshold current. For diffusion-driven leakage of electrons, the leakage current I_{leak} can be described via an activation energy E_a, equal to the difference in energy between F_c and the lowest (Γ, L, or X) CB edge of the cladding layer. Thus,

$$I_{\text{leak}} = I_0 \exp\left(-\frac{E_a}{k_B T}\right), \tag{35.35}$$

where k_B is the Boltzmann constant, T is the temperature, and I_0 is independent of temperature. As with Auger recombination, carrier leakage is strongly temperature dependent and can lead to a large change in the threshold current with temperature. For this reason, much effort has been put into designing lasers which minimize carrier leakage. In III–V semiconductors, due to the large hole effective mass, F_v moves toward the VB at a much slower rate than F_c moves toward the CB edge. For InGaAs(P)-based materials as used in 1.3 and 1.5 μm lasers, the QW to barrier/SCH band offset in the CB is much lower than the VB offset. Additionally, holes have a much lower mobility than electrons. Consequently, the overflow of electrons is generally considered to be more significant compared to the overflow of holes in this materials systems. Carrier leakage is a particular problem for visible semiconductor lasers for which the Al-containing layers may be close to, or indeed are, indirect, whereby the CB X-minima are at a lower energy than the CB Γ-minima. Thus, electrons may thermally escape from the separate confinement heterostructure (SCH) into the cladding, forming a leakage current. An example of this is shown in Fig. 35.19 where the threshold current is plotted as a function of temperature for a 670 nm Al(GaInP) laser. The strong increase in threshold current with temperature is due to the thermal leakage of electrons from the GaInP QW via the SCH region into the X-minima of the AlGaInP barrier layers. At room temperature, leakage accounts for ≈ 20% I_{th} at room temperature, rising to ≈ 70% at 80 °C at this wavelength [35.35]. Due to the smaller activation energy, shorter wavelength devices suffer still further from carrier leakage and this, in practice, sets the lower wavelength limit for lasers produced using this material system.

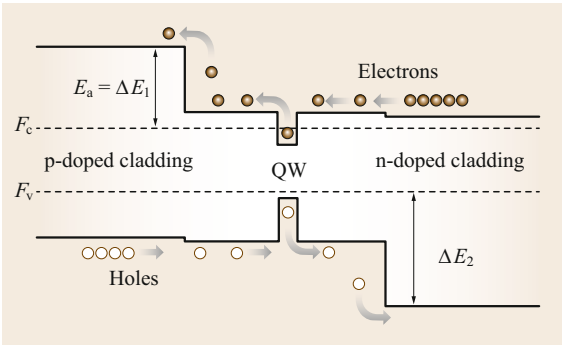


Fig. 35.18 Carrier leakage in a QW semiconductor laser

35.2.13 Temperature Sensitivity and T_0

As described in the previous section, the threshold current of lasers operating in the visible and in the 1.3–1.5 μm range is very sensitive to the ambient temperature. From a commercial perspective, this is very unsatisfactory as these lasers are required to provide a constant light output for a given drive current above threshold all year round. However, throughout the year, the ambient temperature can typically vary significantly. Hence, further cost results from having to use expensive temperature-control electronics to stabilize the temperature and hence light output of these lasers. This can be seen in Fig. 35.20 where the power–current curves are plotted for a 1.3 μm laser over the temperature range 20–70 °C. Clearly, by operating the laser at a constant bias current, the light output decreases considerably as the temperature decreases. In extreme cases, the current may no longer be sufficient

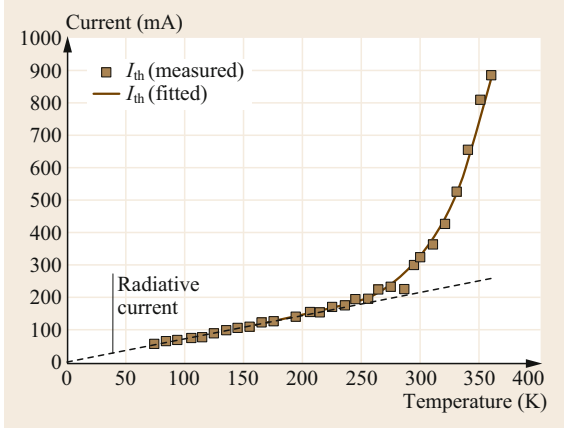


Fig. 35.19 Threshold current versus temperature for a 670 nm GaInP QW laser. The strong temperature sensitivity above 250 K is due to electron leakage into the X-minima of the AlGaInP barrier layers

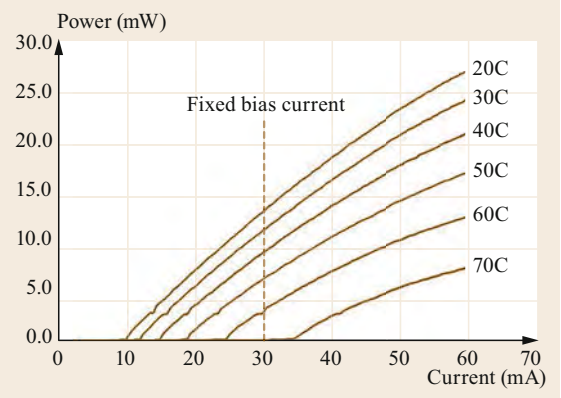


Fig. 35.20 Typical curves of power versus current for a 1.3 μm laser as a function of temperature. The large increase in threshold current with increasing temperature gives rise to a strong loss of power for a fixed bias current

to reach laser threshold, at which point the device will no longer lase and the output would only be SE (as shown in Fig. 35.20, for the 70 °C curve). In practical situations, this therefore necessitates the use of temperature-control feedback electronics. These add considerable cost (many times the cost of the laser itself) to the system. This has become known as the T_0 problem. Around room temperature and over a limited range of temperature, the threshold current of many semiconductor lasers was observed to increase approximately exponentially with temperature [35.36]. Hence, it is usual to write that $I_{th} = I_0 \exp(T/T_0)$, where I_0 is a constant and T_0 is the characteristic temperature of the threshold current given by

$$\frac{1}{T_0} = \frac{1}{I_{th}} \frac{dI_{th}}{dT} = \frac{d \ln(I_{th})}{dT}, \tag{35.36}$$

T_0 is usually expressed in units of Kelvin. It follows that a low value of T_0 corresponds to a high-temperature sensitivity and vice versa. Hence, a high T_0 is desirable. Experimentally, T_0 can be determined from measurements of I_{th} from

$$T_0 = \frac{T_2 - T_1}{\ln\left[\frac{I_{th}(T_2)}{I_{th}(T_1)}\right]}, \tag{35.37}$$

where $I_{th}(T_1)$ and $I_{th}(T_2)$ are the threshold currents at T_1 and T_2, respectively, and $T_2 > T_1$. In practice, over a wider temperature range, T_0 is itself temperature dependent due to the temperature dependence of the recombination mechanisms that occur in the lasers. However, T_0 is a parameter often used to investigate the temperature dependence of the threshold current. For a typical 1.3 or 1.5 μm QW laser, $T_0 \approx 50$–60 K. This is due to Auger recombination, as discussed above. It can be shown that, for an ideal QW laser, the maximum $T_0 = 300$ K around room temperature, where the threshold current is entirely due to radiative recombination [35.37]. Furthermore, by going to QD lasers, due to the 3-D-confinement effect, an infinite T_0 has been predicted [35.12]. Recently reported experimental work has shown that this may be possible by using p-doped QD layers, which exhibit temperature-insensitive operation around room temperature [35.38].

35.3 Single- and Multimode Lasers

35.3.1 DFB and Wavelength Tuneable Lasers

In an ideal laser, the output spectrum would be very narrow with no side modes. In practice, the output spectrum of Fabry–Perot devices such as those described earlier in this chapter consists of several competing modes. This is due to the comb of Fabry–Perot modes, which overlap with the laser gain spectrum (shown earlier in Fig. 35.9). There are a number of laser devices that can be operated with high spectral purity. Perhaps the most important of these is the DFB laser. In a Fabry–Perot laser, the end facets provide optical feedback to give rise to provide a sufficient gain to achieve threshold. Because the reflectivity of the end facets is only slightly wavelength dependent (for a given polarization), this gives rise to several longitudinal modes. These are clearly visible in the output amplified SE spectrum of a Fabry–Perot laser; indeed, this is one way in which the gain can be measured [35.39]. In contrast, in a DFB laser, there is an additional layer called the guiding layer (Fig. 35.21a,c) grown close to the active layer. The guiding layer is essentially a grating consisting of a corrugated layer of a dissimilar refractive-index semiconductor defined using lithographic and etching techniques. If the grating is defined only near the laser facets, it is called a distributed Bragg reflected (DBR) laser (Fig. 35.21b). The corrugations in the material provide a periodic variation in the refractive index along the laser cavity and hence the optical feedback is *distributed* along the length of the cavity.

Both DBR and DFB lasers essentially have the same operating principle. The optical radiation generated in the active layer satisfying the Bragg condition will be reflected back efficiently. This condition is set

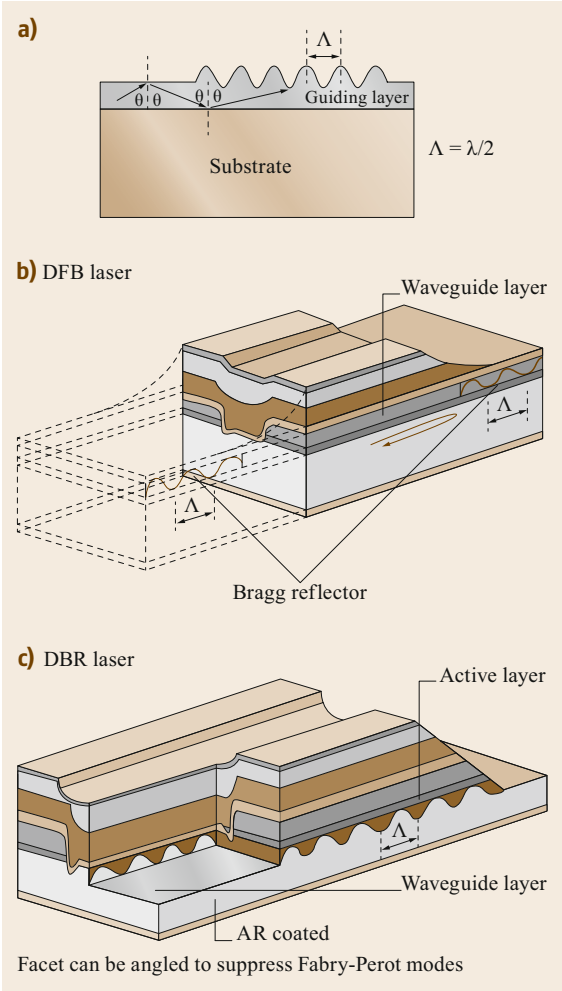


Fig. 35.21 **(a)** Schematic of a DBR based waveguide design on a substrate with grating period Λ, **(b)** a DBR laser design with a grating structure defined to selectively reflect the desired wavelength for oscillating in the active layer. **(c)** A DFB laser design where the grating covers the entire cavity. Both lasers shown have a buried heterostructure lateral current confinement design. (After [35.40, 41])

by waves scattering off successive corrugations reflecting and constructively interfering. For a guiding layer with the refractive index η, the Bragg condition is given by [35.40, 42]

$$2\Lambda \sin\theta = \frac{m\lambda_0}{\eta} \,. \tag{35.38}$$

The minimum requirement for a positive optical feedback within the cavity is scattering at $\approx 180°$ with respect to the direction of the propagating mode resulting in $\theta \approx 90°$. One can include the angular dependence in the definition of the modal effective index η_{eff} for the guided mode giving

$$\Lambda = \frac{m\lambda_0}{2\eta_{\text{eff}}} \,, \tag{35.39}$$

that is, the grating period required for the laser to operate λ_0 at depending on the order of diffraction m.

The grating structure can be produced via photolithography in which holographic techniques are used for projecting a periodic interference pattern on the surface to be etched [35.40–42]. For many near-infrared lasers, the requirement for the first-order grating period is in the 100–300 nm range, requiring resolutions available only with advanced lithographic equipment. To make fabrication easier, higher order diffraction grating structures have been investigated, although their efficiency tends to be lower due to lower order diffraction loss. It can be shown that a DFB laser has two resonant frequencies around the desired cavity resonance where the cavity oscillations can occur [35.43]. This is undesirable and is generally remedied by using a $\lambda_0/4$ phase-shifted design [35.44]. DBR lasers can also provide single-frequency operation with a high degree of stability and tunability. As the gain and Bragg reflection regions are separate, independent control of the carrier density in the grating region is possible which can be used to alter the effective index of the waveguide using the free-carrier plasma effect [35.42], for fine frequency tuning.

An added benefit of the grating in a DFB laser is that it may be used to artificially reduce the temperature sensitivity of the laser threshold current and lasing wavelength. Unlike Fabry–Perot lasers, which primarily emit at the mode closest to the gain peak, as discussed above, DFB lasers can be made to emit at one fixed mode. The peak of the gain spectrum moves approximately at the same rate as that of the band gap ($\approx 0.5\,\text{nm/K}$), while the temperature dependence of the DFB mode follows the thermally induced change in the refractive index ($\approx 0.1\,\text{nm/K}$) which is much less temperature sensitive. Thus, DFB lasers are far less susceptible to thermally induced mode hops. Furthermore, by carefully designing the peak of the gain spectrum to lie at a shorter wavelength than the DFB wavelength, one can effectively *tune* the DFB laser with increasing temperature resulting in, over a limited temperature range, a relatively temperature-independent threshold current. This has been demonstrated as a means of achieving cooler-less 1.3 μm lasers, albeit at the expensive of higher threshold currents.

DFB lasers are the mainstay of optical communications systems due to their single-mode behavior with high side-mode suppression and narrow line widths ($< 0.1\,\text{nm}$). In an optical-fiber communication system, this allows for the simultaneous transmission of sev-

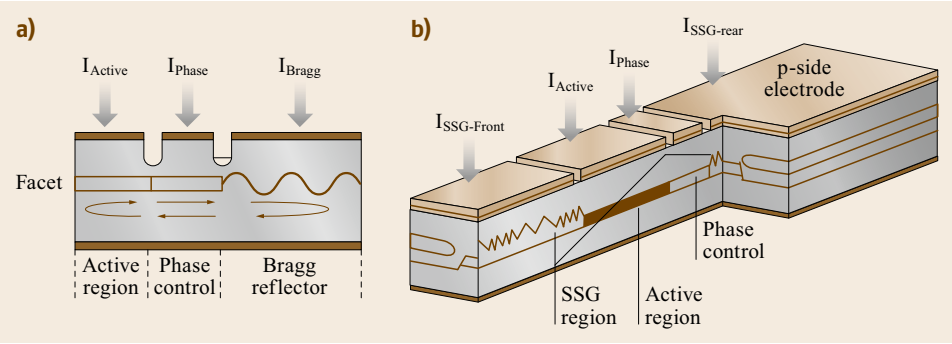


Fig. 35.22
(a) A three section laser design for independent gain, phase and mode control. **(b)** A superstructure grating (SSG-DBR) tuneable laser

eral data channels with closely spaced wavelengths around the fiber attenuation minimum of 1.55 μm. This is known as dense wavelength-division multiplexing (DWDM).

The DBR design can be adapted to design efficient and widely tuneable lasers on a single semiconductor chip. Figure 35.22a shows a three-section laser design where each section can be pumped independently. This allows wavelength tuning via phase tuning or tuning of the effective grating period (both via refractive-index tuning) by independent electrical pumping of the respective sections. A combination of both allows a continuous tuning of about 1% around the central wavelength [35.45, 46]. This limitation of tuning range is imposed by the extent to which the effective refractive index can be altered via current injection, that is, (mainly) via the free-carrier plasma effect. In this effect, the free carriers under mutual Columbic interaction alter the dielectric polarization of the material, thereby altering its refractive index [35.45, 47]. An order of magnitude increase in the tunability range can be achieved by using a superstructure grating (SSG) design, Fig. 35.22b. An SSG can be approximated as an array of frequency-independent scattering centers spaced at Λ. The SSG-DFB laser works using the Vernier effect between the reflection peaks of the two SSGs which are designed to have slight offset in their reflection peaks [35.45, 47]. SSG-DFB lasers and related design variations are used as local oscillators in modern coherent detection systems forming an integral part of today's optical communication networks.

35.3.2 Vertical Cavity Surface-Emitting Lasers (VCSELs)

We note from the discussion in Sect. 35.2.7 that if the frequency spacing between the adjacent longitudinal modes is wider than the gain bandwidth, single longitudinal mode operation is possible. Equation (35.29) suggests that this can be achieved by having a very short cavity. In a VCSEL, the cavity length is made very short and the laser beam is emitted perpendicular to the wafer, that is, in the same direction as that of the injected current. Because the VCSEL emits from its surface, the optical cavity over which the optical field may experience the gain is very short. Consequently, in order to reach the threshold gain level, the optical field needs to make several passes through the active region. This requires highly reflective end mirrors. At a typical III–V semiconductor–air interface, the refractive-index contrast gives rise to a reflectivity of $\approx 30\%$. This is enough to achieve a sufficient gain to reach threshold in an edge-emitting laser, which is typically many hundreds of micrometers long. However, in a VCSEL, the cavity length is typically of the order of the wavelength of light in the material, requiring mirror reflectivities $> 99\%$. This is achieved by sandwiching the active region between two multilayer dielectric (Bragg) mirrors. The short cavity and small active volume require a high-Q cavity to keep the lasing threshold current low [35.40, 48].

The active layer in a VCSEL is sandwiched between a stack of Bragg mirrors, that is, alternate layers of high and low refractive-index materials, each of which is designed to be a quarter wavelength thick at the desired laser wavelength, Fig. 35.23a,b. The Bragg mirrors provide very high reflectivity of the light back into the active region, allowing the VCSEL to generate a sufficient gain in spite of the small gain region optical path length. Constructive and destructive interferences within the Bragg stack result in a Bragg resonance, which gives rise to a reflectivity that increases with the number of layer pairs (one high index high η_{high} and one low index low η_{low}). The resultant reflectivity for the intra-cavity field can be approximated by

$$r_{Bragg} \approx \tanh\left[2m\left(\frac{\eta_{high}-\eta_{low}}{\eta_{high}+\eta_{low}}\right)\right] = \tanh\left[\frac{2m\Delta\Lambda}{\lambda_0}\right], \quad (35.40)$$

if the entry and exit materials for the radiation are considered the same [35.49], where m is the number of

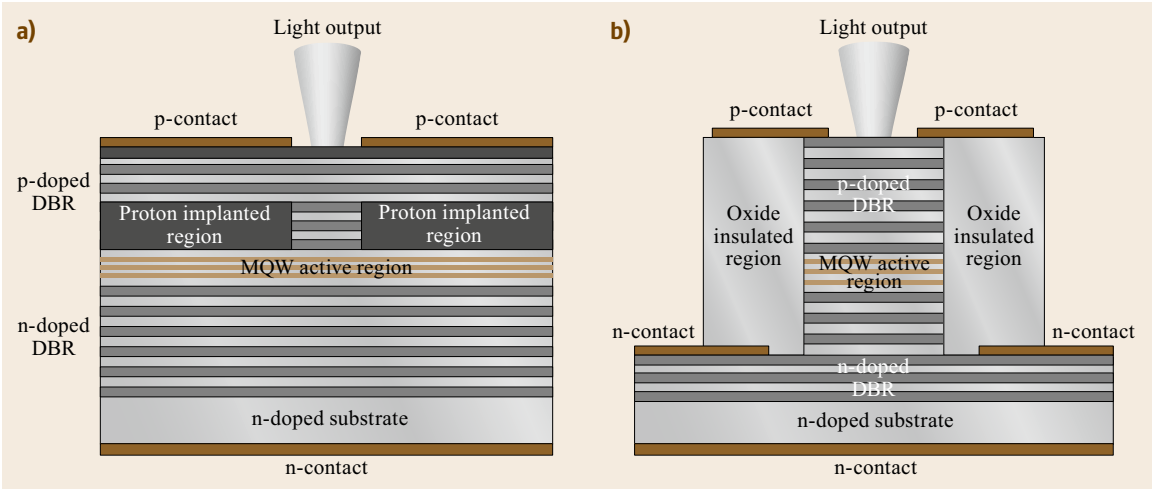


Fig. 35.23a,b Cross-sections of two common cylindrical VCSEL designs for lateral current confinement **(a)** a proton-bombarded top-bottom contact design and **(b)** an oxide confined side-contact design. The MQW active region sandwiched between the DBR stacks forming the cavity is also shown. The VCSEL being a cylindrical design emits a circular beam profile

layer pairs. For $\eta_{\text{high}} - \eta_{\text{low}} = 0.3$, greater than 98% intensity reflectance ($R = r^2_{\text{Bragg}}$) can be achieved for 30 layers in an AlGaAs/GaAs-based Bragg stack for resonance wavelength around $\approx 850\,\text{nm}$. For longer wavelengths, a higher number of (thicker) layer pairs are generally required [35.49]. A Bragg stack with a sufficient number of layer pairs results in a flat and broad reflectivity spectrum (called a stop band) with a width $\Delta\Lambda$ around the cavity resonance wavelength λ_0 it is designed for, see Fig. 35.24b. The top Bragg stack usually has lower number of layer pairs for the output coupling of the generated laser beam.

This emission wavelength from a VCSEL (or for that matter any single-frequency laser) is selected via a dynamic competition between the cavity resonance and material gain peak. For a high-Q cavity, the emission wavelength is closer to the cavity resonance than the gain peak (mode/frequency pulling [35.50]). Thus, in general, the VCSEL cavity is designed so that there is a fixed detuning between the material and the cavity resonances at room temperature so that at the operating temperature the cavity resonance aligns with the peak gain and the desired wavelength is emitted with the maximum gain available at the cavity resonant mode [35.51–53].

In addition to the advantage of being an inherently single-frequency laser, VCSELs offer several other advantages, including ease of fabrication and on-wafer testing of single chips, ease of fabrication of planar arrays (hence lower costs), and their symmetric circular beam profile which makes coupling of the generated light to an optical fiber easier and more efficient (lower coupling loss) compared to edge-emitting lasers. VCSELs can be fabricated using conventional planar integrated circuit fabrication techniques and can provide emission over a wider aperture if high order spatial modes can be tolerated in the application. Optical interconnects based on narrow-aperture VCSELs are used in many levels of modern light-wave communication systems. Note that the resonant cavity LED (RCLED) is a related device that provides surface emission in a VCSEL-like structure. However, such devices have a lower number of DBR layer pairs in the Bragg stack and do not achieve a lasing threshold and thus emit narrow-band SE. RCLEDs are, however, useful for applications such as plastic optical fiber (POF)-based visible light data links [35.54].

Other forms of surface-emitting laser have also been designed where an angled facet is used to direct the laser beam perpendicular to the plane of the epistructure while using conventional current injection designs as in edge-emitting semiconductor lasers [35.40, 41, 55]. Laser designs based on photonic band-gap materials (i. e., photonic crystal lasers [35.56]) also employ vertical emitting designs (Photonic Crystal Surface Emitting Lasers, PCSELs) with carrier injection schemes similar to conventional VCSELs.

High-quality VCSELs have already been produced for emission at 650 and 850 nm (for plastic and silica fiber communications), for oxygen sensing using 760 nm VCSELs and high-power devices for emission at 980 nm [35.57]. However, there has been considerable difficulty in producing VCSELs at the telecommunication wavelengths of 1.3 and 1.55 μm. This is

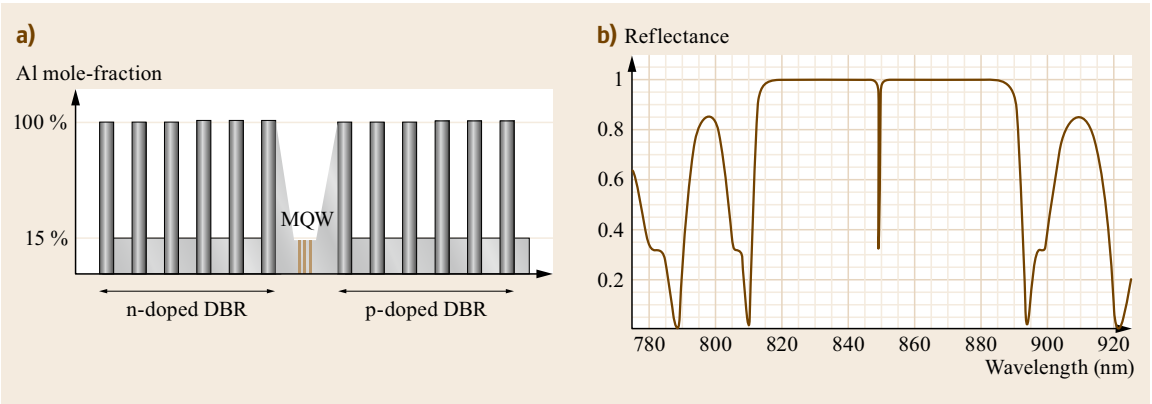


Fig. 35.24 **(a)** MQW active region in a AlGaAs/GaAs-based VCSEL sandwiched between two AlGaAs/GaAs Bragg stacks which are differently doped for efficient carrier transport. **(b)** Typical stop-band characteristics in a VCSEL around the resonance wavelength

primarily due to the fact that GaAs/AlGaAs is the preferred system for making highly reflective DBRs due to the high refractive-index contrast between GaAs and AlAs. Such DBRs are therefore compatible with the GaAs substrates as used in devices below $\approx 1\,\mu m$. However, for longer wavelength devices, which are primarily based on InP substrates, it is difficult to form highly reflective DBRs using InGaAsP/InP due to their low refractive-index contrast. Alternative methods of producing long-wavelength VCSELs are inherently process-intensive, using techniques such as wafer fusion in which the active layer and InGaAsP/InP DBR are first grown on an InP substrate, while a GaAs/AlAs DBR is grown separately on a GaAs substrate. The active layer is then fused onto the GaAs/AlAs DBR, and the InP substrate is then removed [35.58]. The wafer fusion process is understandably both costly and difficult, and therefore methods of growing monolithic 1.3 and $1.55\,\mu m$ VCSEL structures using only a single GaAs substrate are particularly attractive. This has stimulated a great deal of research into producing GaAs-based active regions emitting at these wavelengths. In recent years, there has been some success in achieving this by two approaches. The first approach is to use InAs/GaAs QDs as the active region. QD lasers with very low threshold current density ($< 20\,A/cm^2$) emitting at $1.3\,\mu m$ have already been realized [35.13]. However, there are few reports in the literature on QD-based VCSELs. This is largely due to the difficulty in achieving a sufficient gain from the QDs and can result in the need for several layers of QDs. In spite of this, QDs are showing promise for use in temperature-insensitive edge-emitting lasers and high-power lasers.

The second approach has been in the use of the so-called dilute nitrides. In 1997, *Kondow* et al. originally proposed the use of InGaAsN/GaAs as an active material to achieve long-wavelength emission on GaAs [35.59]. This is due to the unusual band-gap bowing that occurs when small concentrations (≈ 2–5%) of nitrogen are added to GaAs. By growing the alloy InGaAsN, one may achieve long-wavelength emission while maintaining a low or zero strain. Although material quality remains an issue, there have been several successful demonstrations of edge-emitting lasers and VCSELs based upon this material emitting at $1.3\,\mu m$ [35.60–62]. The push towards $1.55\,\mu m$ emission has been largely hampered by material quality issues although at the time of writing this chapter, there have been the first reports of low-threshold $1.5\,\mu m$ edge-emitting lasers based upon InGaAsN/GaAs [35.63]. Other researchers have produced close-to-$1.5\,\mu m$ edge-emitting and VCSEL devices with the pentenary material InGaAsNSb/GaAs [35.64]. It is speculated that antimony may act as a surfactant to improve the quality of growth. The dilute nitride approach does at present appear to offer the best possibility of obtaining monolithic VCSELs emitting at the technologically important wavelength of $1.55\,\mu m$.

35.3.3 High Power Lasers

Direct and efficient conversion of electrical energy into coherent light provides a wealth of applications such as in solid-state and fiber laser pumping, medical applications, optical communication via free-space energy transfer, and high quality material processing (cutting and welding) allowing the process to be controlled with high spatial and temporal resolution. Thus, interest in scaling the output power to the kilowatt range and beyond has been there from the very beginning of semiconductor lasers. There is a limit to the max-

imum optical power that can be extracted from an edge-emitting stripe-geometry laser. This limit is set by the *thermal runaway* power load and secondly the optical power load at the facet which can be up to few $\approx$ MWcm^{-2}, beyond which the facet succumbs to the generated heat and undergoes catastrophic optical damage (COD) [35.65, 66]. With an increase in carrier injection (pumping), the optical output power from the laser increases until the generated heat (via unused injected power) causes thermal escape of the carriers from the active region (thermal runaway [35.46]), resulting in a drop in the optical power upon further current injection. At sufficiently high injection, the device no longer lases owing to the fact that the threshold current itself increases (due to joule heating) with injection until such point that the threshold current exceeds the injected current level. Thermal runaway is a reversible phenomenon, unlike COD, which irreparably damages the laser [35.46]. COD is sensitive to the quality of the facet and modern facet passivation approaches mean that COD is less problematic than it once was. Thermal runaway generally occurs at a much lower power load than facet damage in diode lasers and thus, under some circumstances, self-limits the device from undergoing COD. High-power diode lasers consequently require wide aperture designs, firstly to achieve a higher gain volume for high-power generation, and secondly to withstand and efficiently dissipate the heavy thermal load generated by the device. Additionally, they are often operated under quasi-CW conditions (100 µs–1 ms pulse length) to limit thermal effects and to extract the maximum possible power.

The distinction between low- and high-power semiconductor lasers is not very well defined. It may depend on the type of diode laser and the application for which it has been designed. Generally, 50 mW (or more) CW for single-mode, single-frequency lasers and 500 mW (or more) CW for broad-area multimode lasers and laser arrays is generally considered to be high power [35.46]. For many applications, however, the power requirement from diode lasers is quite low. For example, CD players use a 3 mW range GaAs laser emitting at $\lambda = 780$ nm, supermarket scanners use visible ($\lambda = 670$ nm) 3 mW AlGaInP lasers, and in optical fiber telecommunications 10–50 mW (in-fiber) InGaAsP lasers ($\lambda = 1.3$ or 1.55 µm) are generally used, all emitting in single mode. However, an important class of applications requires much higher output powers either in pulsed or CW operation. Operation in a single spatial mode or at a single frequency is often desirable, but raw power can also be very useful. One example where raw power is most important is in diode-pumped yttrium–aluminium garnet (YAG)-based solid-state lasers, where the pumping laser array fabricated from AlGaAs material emits kilo-watts of incoherent quasi-CW pulses. The wavelength of the AlGaAs laser array is well tuned to the absorption spectrum of the YAG crystal (806–810 nm), providing much higher overall efficiency than flash-lamp pumping. Applications where both high-power and near-diffraction-limited beam quality (i. e., coherent emission) are required include many communication and most scientific applications, where transmission over long distances and/or the ability to focus the light to a small spot of high power density is desired, such as pumping erbium-doped fiber amplifiers (EDFAs). In general, the power available from incoherent sources exceeds the power available from coherent sources by more than an order of magnitude. However, the brightness or radiance (power/beam divergence) of coherent, diffraction-limited sources is significantly higher. Tremendous advances have been made in scaling the output power of diode lasers and arrays, thanks to modern crystal growth, device cooling, and facet passivation technologies.

Over the last two decades, the output power from diode lasers has increased by an order of magnitude [35.46, 67]. The electro-optic (wall-plug) conversion efficiency measured under laboratory conditions is now $> 70\%$ in GaAs-based lasers, which has reduced the thermal load immensely [35.68–70]. Under pulsed conditions, more than 150 W of optical power output has already been demonstrated from a 100 µm emitting stripe geometry laser [35.69]. Under CW operation, optical power in the range of tens of watts has been demonstrated before thermal runaway occurs [35.46]. Further improvements are being reported each year [35.49, 70]. Multi-kW level raw power delivery is already possible from modern stacked laser diode arrays [35.71].

A diode laser array is essentially a multistripe semiconductor laser with stripes fabricated in close proximity to each other (Fig. 35.25). Lateral current confinement can either be achieved via a mesa-stripe (gain guided) or via proton-bombarded (index-guided) design [35.71–73]. It is common to have 1 cm wide arrays (called diode laser bars) in array stacks. The proximity of the individual stripes results in an overlap of oscillating intracavity field distributions from adjacent emitters (Fig. 35.25b) and with appropriate design these can be phase locked, resulting in a combined, coherent wavefront being emitted from the whole facet [35.71–73]. Vertical stacks of the cm bars operate under active, for example, liquid-based cooling (Fig. 35.25c).

Optically pumped semiconductor slab/disk lasers (also known as VECSELs or OPSELs, see below) [35.74] are the only technology competing with stacked diode laser arrays. Vertical external cavity

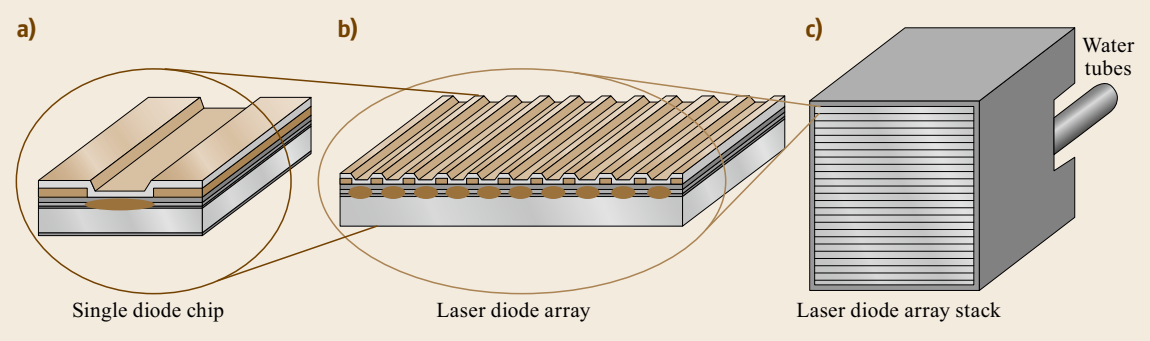


Fig. 35.25a–c Schematic for a high-power diode laser: **(a)** single chip, **(b)** array of diode lasers, **(c)** a two-dimensional stack (requires active cooling)

lasers or VECSELs are essentially optically pumped large aperture VCSELs which are designed with a pump absorbing layer in place of the top DBR stack (output coupling side), see Fig. 35.26a. Hence, they are also called optically pumped surface emitting lasers (OPSELs) or just disk or slab lasers. Diode laser array stacks are used for optical pumping while an external mirror completes the cavity [35.75].

The advantage of VECSELs over a stack of edge emitting laser bars is a symmetric beam profile, a narrow line width and the possibility of intracavity frequency doubling [35.75]. The scalability of the aperture allows the generation of high powers while maintaining the circular beam shape and narrow line width. However, efficient thermal management is essential for VECSEL operation due to the generally poor thermal conductivity of the bottom DBR.

Despite the developments in high-power diode laser technology, achieving high spatial coherence (beam quality) still remains an open challenge. Beam quality of high-power diode lasers (both edge emitting and large aperture VECSELs) is affected by nonlinear filamentation of the spatial modes and is far behind from what is achievable from modern solid state and gas lasers. Poor beam quality affects the focusing ability (brightness) of the beam and the coupling efficiency to optical fibers. A nonzero positive phase–amplitude coupling in the semiconductor gain medium is the primary reason behind the filamentation instability [35.46, 73]. Thermal effects also play a key role [35.46, 73, 74, 76, 77]. Currently, a master oscillator power amplifier (MOPA) design or a derivative of it offers some promise in achieving high-spatial and temporal coherence from a broad area diode laser [35.46, 73]. However, the coherence properties start deteriorating at higher pumping due to thermal effects [35.74, 76–80]. The working principle behind a MOPA design is as follows. The MOPA design consists of a narrow aperture ridge waveguide section with an in-built grating (a DBR laser-type design) acting as the master oscillator which injects a spatially and spectrally filtered beam into an adiabatically tapered section that amplifies and scales the optical power, see Fig. 35.26b. Some variants of the MOPA design do not use a grating when spectral purity is not important for the application. Beam spoiling grooves are sometimes used to attenuate any backward

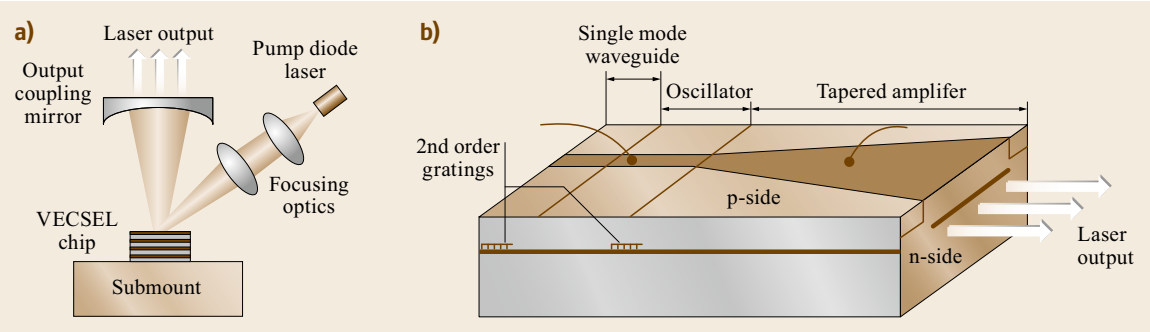


Fig. 35.26 **(a)** Schematic of a vertical external cavity surface-emitting laser (VECSEL) showing the VECSEL chip, the external cavity and the pumping scheme **(a)** a master oscillator power amplifier (MOPA) laser design for high brightness (see text for details)

propagating wave in the tapered region for enhancing the beam stability [35.46, 73].

High brightness, high-power diode laser sources are gaining interest for applications previously dominated by solid-state lasers, because of their efficiency, compactness, and high reliability. Among others, laser projection displays, free-space communication, frequency doubling, and direct material processing are rapidly emerging areas, where high output power together with diffraction-limited beam quality is either the key requirement or strongly enhance system performance. Production of high-power diode lasers with improved brightness will allow them to compete with the very best solid state and gas lasers at much lower cost [35.46, 67], and may enable future generations of compact, inexpensive, and portable laser systems.

35.3.4 Quantum Cascade Lasers

The semiconductor lasers that we have discussed so far emit wavelengths largely determined by the band gap of the available (direct-band gap) semiconductor materials since the devices rely on so-called interband transitions. This restricts the use of semiconductors as a lasing medium in large parts of the mid- to far-infrared spectrum, not covered by conventional direct semiconductor materials or for which processes, such as Auger recombination, which become increasingly important for narrow band gaps, are problematic and limit device performance [35.81, 82]. A QCL uses intraband transitions (instead of interband transitions used by conventional lasers) to generate photons in the mid- and far-infrared spectrum, including in the terahertz (THz) range. These intraband transitions are electronic transitions within the QWs, and the emitted photon energy depends on the QW width and heterojunction band offsets. Therefore, the desired emission wavelength can be varied just by varying the well width. Thus, QCLs are uni-polar lasers unlike conventional semiconductor lasers which are bi-polar devices. Almost all III–V semiconductor-based QCLs use electrons in the CB as the charge carriers; however, it is also possible to use holes as charge carriers in a VB-based device, as is the case of SiGe/Si-based QCLs [35.83]. The QCL structure was originally proposed in 1971 [35.84] and was first demonstrated only in 1994 [35.85], mainly due to the complex structure and accurate growth technology required to implement a QCL design in practice. Since QCLs are intraband and unipolar devices, the physics of gain and lasing is different from what we have discussed above for interband lasers and a detailed discussion is beyond the scope of this chapter. Nevertheless, QCLs form an important class of modern semiconductor lasers so we will present a qualitative outline of their working principles.

The cavity and waveguide design used in both conventional semiconductor lasers and QCLs are essentially similar except for the doping (both cladding layers are n-doped in the majority of QCLs) [35.56, 85, 86]. The main difference in QCLs lies in the active layer which employs intraband transitions only in one band. Since-electron charge carrier in CB-based devices form the vast majority of QCLs, we will focus on these. This can be seen in Fig. 35.27 which shows a section of the active layer which employs MQWs in two stages: an injector region and an active region all under an electric field. The closely spaced QWs and narrow barriers in the injector region result in the formation of mini bands in a superlattice [35.56, 85, 86]. Quantum-mechanical tunneling-assisted transport of the injected electrons through the minibands into the adjacent active region creates an effective population inversion in the active QW. This leads to an intraband transition of the electron from the upper level to lower level of the QW resulting in photon emission. Subsequently, the electron thermalizes to a lower state and tunnels into the next injector region cascading through the structure (hence the name of the laser) producing many more photons. A typical QCL has 20–80 such stages, each having several MQW structures. Typically, there are several hundred to over a thousand epi-layers in QCLs. QCLs have been demonstrated from just below 3 up to 67 μm, opening-up the THz range and its associated applications [35.86, 87]. High-power and DFB QCLs have also been demonstrated [35.56, 86]. It is interesting to note that the *quantum* efficiency of QCLs can be far above 100% owing to the fact that one electron produces many photons. However, this is not breaking any law of physics, as the power (wall-plug) efficiency is typically $< 50\%$ owing to the large voltages that are required to produce the electric field across the relatively thick structure which allows the cascading process to occur.

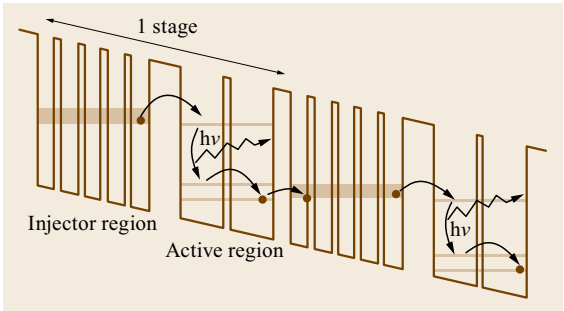


Fig. 35.27 Schematic of the CB design for two successive stages inside the active region of a quantum cascade laser (QCL)

35.3.5 Emerging Semiconductor Materials for Modern Semiconductor Lasers

The lasing characteristics and reliability of semiconductor lasers largely depend on materials used in their design. The band gap and the lattice constant both play a vital part and the maturity of the growth technology of the chosen semiconductor. While the band gap for most types of laser largely determines the emission wavelength, the lattice constant determines whether the semiconductor material can be grown lattice matched on the substrate of choice, hence reducing the density of defects. The latter directly affects the performance and reliability of the device. As discussed earlier, a small amount of strain can be beneficial if kept within a thin QW layer [35.40, 55]. However, in general, the overall semiconductor laser structure should be relatively strain free to minimize defect formation. A simple SQW-SCH active region design requires a minimum of five epi-layers on the substrate to make a laser structure (typically, there are more than 10 layers). One can see that this immediately limits the number of semiconductor alloys with the desired band gap and lattice constants which can be used in a typical laser design. While many new semiconductor alloys have been proposed in recent years to fill the existing gaps, there are only a few with mature enough growth technology which can be used for commercial devices that require high reliability. Figure 35.28 lists some semiconductor alloys with their band gaps and lattice constants [35.88, 89].

Materials falling on a vertical line are lattice matched to each other. There is a great impetus toward achieving efficient semiconductors active regions grown on silicon substrates which will allow on-chip integration of lasers with silicon-based optical waveguide components for all-optical communications and data processing. Being an indirect-band-gap semiconductor, silicon itself is a poor emitter of light; hence, the integration of high optical quality direct-band gap semiconductors on cheap silicon substrates is highly desirable. The primary limiting issue to achieve this is the lattice mismatch between silicon and standard III–V materials such as InP- or GaAs-based alloys (Fig. 35.28), which makes it difficult to grow high-Q alloys directly on silicon.

There have been tremendous advancements in hybrid approaches to this where strain relaxed (metamorphic) growth techniques have been used to grow relatively defect-free working laser structures [35.90, 91], through the use of QD [35.92] and exotic QW [35.93] active regions and through wafer fusion approaches [35.91]. This remains an active area of ongoing research as noted in a number of reviews on this topic [35.94, 95]. Entirely new materials for semiconductor lasers continue to being investigated to help push the wavelength limits of semiconductor lasers and to improve the efficiency and temperature sensitivity of devices. Recently, for example, GaAsBi-based semiconductor lasers have been demonstrated [35.90, 94–99]. Such alloys provide a route to overcome the ill-effects of Auger recombination which have limited the maximum achievable power and efficiency of near- and mid-infrared semiconductor lasers and have the potential to significantly reduce the energy demands of optical fiber communications [35.90, 94–98].

Table 35.1 lists some of the common active and cladding layer compositions used in modern semiconductor lasers together with the substrate material and the emisson wavelength range. The list is incomplete as many materials and compositions still under research are not included [35.100]. The values of mole fractions indicated in Table 35.1 are either for direct-band gap or strain balanced designs. Semiconductor lasers emitting in the range 0.35−0.92 μm find applications in holographic storage, image recorders, fax machines, printers, solid state and fiber laser pumping, barcode readers, blu-ray disks and other disk read–write devices, material processing, medical therapeutics, range finders and aerospace. 0.9−1.10 μm semiconductor lasers are used as optical fiber amplifier pumps, in material processing and medicine. Approximately, 1.1−1.65 μm lasers find their use in optical fiber communication sys-

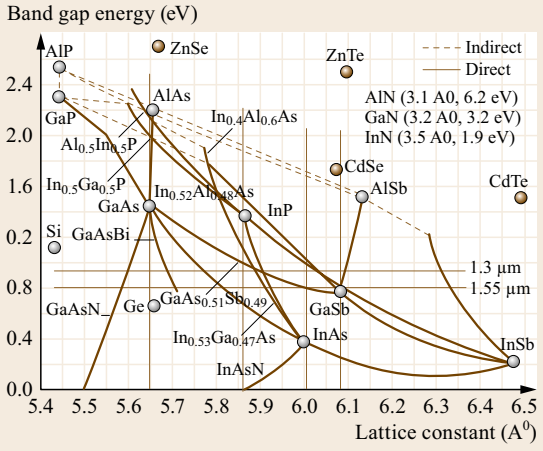


Fig. 35.28 Band gap energies and lattice constants of common semiconductor alloys used for making semiconductor lasers (After [35.88]). Open symbols refer to indirect-band gap materials while closed circles are direct-band-gap materials. AlN, GaN, and InN indicated separately. Alloys are indicated by the connecting curves (*solid* are direct-band-gaps and *dashed* are indirect-band gaps)

Table 35.1 List of some example active and cladding layer compositions used in modern semiconductor lasers. The substrate material and approximate emisson wavelength range is also indicated. (After [35.40, 42])

Cladding layer/Active layer	Substrate	Emission wavelength range (μm)
$GaN/In_xGa_{1-x}N$	Sapphire, GaN, SiC	≈ 0.35–0.50
$Al_xGa_{1-x}N/In_xGa_{1-x}N$	Sapphire, GaN, SiC	≈ 0.40–0.60
$Cd_xZn_{1-x}S/CdS$	CdS	≈ 0.35–0.50
$(Al_xGa_{1-x})_{0.5}In_{0.5}P/Ga_yIn_{1-y}P$	GaAs	≈ 0.60–0.72
$Al_xGa_{1-x}As/GaAs$	GaAs	≈ 0.60–0.92
$Al_xGa_{1-x}As/In_yGa_{1-y}As$	GaAs	≈ 0.90–1.10
$InP/Ga_xIn_{1-x}As_yP_{1-y}$	InP	≈ 1.1–1.65
$Al_xGa_{1-x}As/In_yGa_{1-y}As_{1-x}N_x$	GaAs	≈ 1.24–1.30
$Al_xGa_{1-x}Sb_{1-y}As_y/In_xGa_{1-x}Sb_{1-y}As_y$	GaSb or InAs	≈ 1.7–4.0

tems, transceivers, and Raman amplifiers. Lasers of the range 1.24–1.30 μm are also used in aerospace and military range finders, and 1.7–4.4 μm lasers are used in spectroscopic sensing of humidity, gas impurities, drugs and optical fiber, and free-space communications [35.56].

35.4 Optical Amplifiers

35.4.1 An Introduction to Optical Amplification

As a light pulse propagates through an optical communication system, the pulse becomes attenuated until eventually it is necessary to regenerate the pulse to keep the signal above the background noise level. This was initially achieved by detecting the pulse using a photodetector, and then using the detected electronic pulse to trigger a laser giving a fresh output pulse. An alternative technique is to use direct optical amplification. This has been achieved for instance by doping optical fibers with rare-earth elements, such as erbium, forming an EDFA. Population inversion is achieved by exciting the erbium atoms at 980 nm or 1.48 μm, where the excited carriers decay to a level from which they can recombine to give stimulated emission at 1.55 μm, amplifying weak signals in the fiber. Amplification can also be achieved within conventional silica fibers using the Raman effect; by pumping the fiber with a high-power pump laser (total power ≈ 1 W) at a fixed frequency above the signal frequency, Raman scattering gives rise to gain at a lower frequency (the difference is the phonon energy). Thus, by pumping an optical fiber at 1.45 μm, one may achieve gain at 1.55 μm. The main advantage of Raman amplification compared with EDFAs is the fact that the gain curve may be dynamically tuned by judiciously using different wavelength pump lasers. The disadvantages of Raman amplification that the overall gain is generally lower than an EDFA and its cross-amplification can be a problem. For some applications, semiconductor optical amplifiers (SOAs) may be preferred.

35.4.2 Semiconductor Optical Amplifiers (SOAs)

The structure of an SOA is very similar to that of a laser, but with one significant difference: the reflectivity of the end facets, $R \approx 0$. This is achieved through the use of multilayer antireflection coatings together with angled facets or cavities. This suppresses lasing within the cavity and eliminates reflected signals in the optical system. The signal emerging from an optical fiber is, in general, randomly polarized and therefore it is very desirable that any optical amplifier provides the gain that is independent of the direction of polarization of the light. In a bulk heterostructure device, the TE and TM material gain are equal. However, the optical confinement factor Γ for TE light polarized in the plane of the heterostructure is slightly larger than that for TM light polarized perpendicular to the heterostructure plane, leading to a larger overall TE gain. As discussed earlier in this chapter, in a normal QW system, the quantum confinement brings the heavy-hole band to the top of the VB and light polarized in the TE mode is amplified considerably more strongly than light polarized in the TM mode. This problem can be overcome using strained-layer techniques. It is possible to grow the wells or the barriers with a small amount of tensile strain. This raises the light-hole band and the strain can be adjusted to increase the device TM gain until it is just equal to that of the TE gain. Work to date indicates that the relative gains in the two modes are sensitive to the magnitude of the amplifier current and so the gains are equal over a limited range.

Strained SOAs

Another more promising approach is to introduce both compressive- and tensile-strained wells alternately into the active region. The compressive wells provide gain predominantly to the TE mode and the tensile wells predominantly to the TM mode. The structure can then be designed to make the two gain modes equal. The initial work by *Tiemeijer* et al. from Philips concentrated on 1.3 μm devices [35.101]. They found that a combination of four compressive wells and three tensile wells each with 1% strain gave the TE and TM gains within 1 dB of each other over a wide wavelength band and over an order of magnitude change in the amplifier current. This structure is an excellent example of the flexibility in device design that is afforded by the introduction of strained-layer techniques. Other applications of such a structure include two-polarization or two-frequency lasers, polarization control elements, and nonabsorbing strain-overcompensated mirrors in high-power lasers.

35.5 Modulators

35.5.1 Modulator Theory

Although it is possible to modulate the output from a semiconductor laser by directly modulating the injected current, in some applications requiring high-speed or low-power switching, it is advantageous to run the laser source CW and modulate the light subsequently [35.102]. This is basically because, unlike lasers, semiconductor modulators are diodes operated in the reverse bias. They therefore draw little current and, because there is no carrier injection, there is much less chirp.

Electroabsorption modulators make use of the fact that the presence of a large electric field adds a perturbation to the Schrödinger equation, which leads to a decrease in the band-gap energy E_g. Thus, photons with energies just below the band gap become more strongly absorbed when a field is applied, leading to a decrease in light intensity. Electroabsorption modulators can benefit enormously from the use of multiple QWs, where the electroabsorption effect is more than one order of magnitude greater than that in comparable bulk structures.

In QWs the modulation of the energy levels by an electric field applied perpendicular to the plane of the wells is called the quantum-confined Stark effect (QCSE). For a given applied voltage, it causes more change, $\Delta\alpha$, to the absorption coefficient α close to the band edge than in equivalent bulk structures for three main reasons. Due to the quantum confinement, there is a step-like increase in the density of states at the band edges, as described above, and so a faster rate of increase in absorption coefficient with energy. Thus, $\Delta\alpha/\Delta E_g$, the change in absorption coefficient with the band gap, is larger. Secondly, the steep rise in the absorption coefficient with energy at the band edge is further enhanced in QW structures due to excitonic effects. An exciton consists of an electron–hole pair bound together by electrostatic attraction and has associated with it a sharp line absorption spectrum which is superimposed on the QW absorption. Figure 35.29 shows the measured photocurrent spectra at low temperature as a function of the applied reversed bias for two long-wavelength modulator structures. The exciton peak is clearly visible at the absorption edge, particularly for low bias values. Excitons also exist in bulk material at low temperatures but not at room temperature. In QWs, however, the electrons and holes are confined into the same region of space by the well and so the electrostatic binding energy is sufficiently large that they are still able to exist at room temperature. Finally, the actual energy of the band edge is largely determined by the effect of quantum confinement, which depends not only on the width of the QW but also on its shape. Since the shape of the well can be strongly modified by an applied electric field, so can the electronic energies at the band edges, as can be seen from the shift in absorption edge with the applied bias, as shown in Fig. 35.29. Thus, ΔE_g is enhanced by quantum confinement.

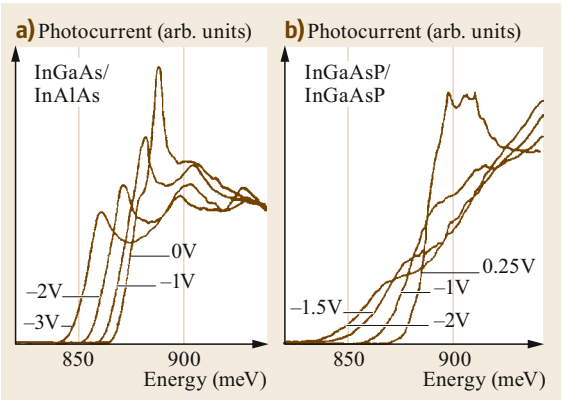


Fig. 35.29a,b Measured photocurrent spectra at 2 K of an InGaAs/InGaAlAs **(a)** and an InGaAsP/InGaAsP **(b)** electroabsorption modulator at several bias voltages. (After [35.58])

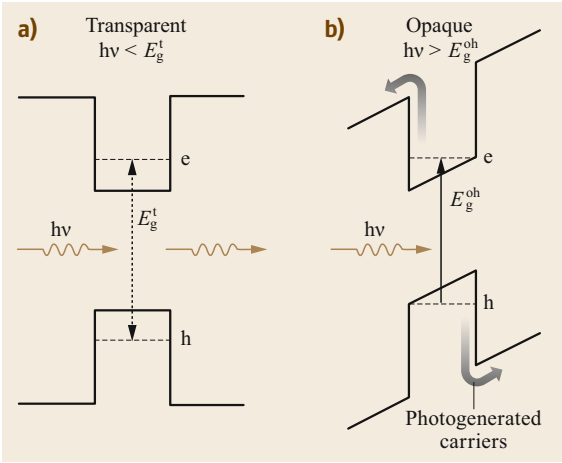


Fig. 35.30a,b The band alignment of a QW in **(a)** flat band (*zero field*) where the optical gap equals $E_g^t > h\nu$ (*transparent*) and **(b)** for a large field across the QW, where the optical gap equals $E_g^0 < h\nu$ (*opaque*)

The effect of an electric field applied perpendicular to the plane of a QW is to change the shape of the well from the rectangular well, as shown in Fig. 35.30a, to the triangular shape, as shown in Fig. 35.30b. Solutions to the Schrödinger equation show that the lowest confined states are closer together in the triangular situation than the rectangular one and so the effective absorption edge is moved to lower energies. Because the electron ground state is shifted toward one side of the QW, and the hole ground state toward the opposite side, the overlap of the electron and hole wavefunctions is decreased by the application of an electric field. The result is to decrease the absolute value of the absorption coefficient and to decrease the exciton absorption peak, as shown in Fig. 35.29. However, if one considers an energy just below the band-edge exciton energy at the zero field, at about 850 meV in Fig. 35.29, the absorption coefficient at that energy can be greatly increased by the application of an electric field.

For a modulator, a very important parameter is the ratio of the light intensity transmitted in the *on* state to that transmitted in the *off* state when the field is applied. This is given simply by

$$P = \exp(\Delta\alpha L)\,, \tag{35.41}$$

where L is the distance the light travels through the absorbing medium. In QW structures, although $\Delta\alpha$ is relatively large, if the light travels in a direction perpendicular to the plane of the wells, the length L is small; of the order of tens of nm. One way to overcome this problem is to use the QCSE modulator in the waveguide configuration; that is, to pass the light parallel to the plane of the QWs in a waveguide structure similar to that for a laser as described previously. In this case, the effective absorption per unit length will be decreased by the optical confinement factor (Γ) to $\Gamma\alpha$, but L can easily be increased much more; to tens or even hundreds of μm. Such a waveguide configuration can achieve a large on–off ratio, but it has two potential problems. First, it can have a large insertion loss because it is difficult to launch the light into the waveguide unless it is integrated with the laser source and also there can be light scattering and absorption along the length of the guide even in the on state. Second, the modulator will be dependent on the polarization of the light beam traveling along the waveguide, as described below.

35.5.2 Polarization-Insensitive Modulators

As was explained when considering strained-layer electronic properties, the degeneracy of the heavy- and light-hole bands, which exists in bulk material, is split by confinement effects in a QW. This is because the quantum confinement energy is proportional to l/m^* and so the first light-hole VB lh_1 lies below the first heavy-hole VB hh_1. As a consequence, the wavefunctions that make up the edge of the VB are derived from the heavy-hole band and so only interact with TE-polarized photons as described above. Thus, while photons of energy close to the band edge and polarized in the TE mode are modulated as the band edges move together with the applied electric field, photons of the same energy, but polarized in the TM mode, are relatively unaffected. This is in general unacceptable if the photons are arriving from a system, such as an optical fiber, which may deliver photons of any polarization in a manner which is sensitive to extraneous influences.

The problem of polarization sensitivity in modulators, as with the polarization-insensitive amplifier, may be dealt with by engineering the band structure with a judicious use of in-built strain within the QWs. As described above, the light-hole band may be raised in energy with respect to the heavy-hole band by the introduction of tensile strain. Thus, with tensile strain, it is possible to shift the heavy- and light-hole bands until the confined states are once again degenerate, as shown in Fig. 35.31a. Such a system, although it may be polarization insensitive at zero net field, can again become polarization sensitive as a net field is applied, as is shown by the movement of the band edges in Fig. 35.32a. This effect arises since the heavy-hole confined level, being closer to its band edge, is more sensitive to the distortion of that band edge by the applied field. Once again, however, it is predicted that this effect can be compensated by further band-structure

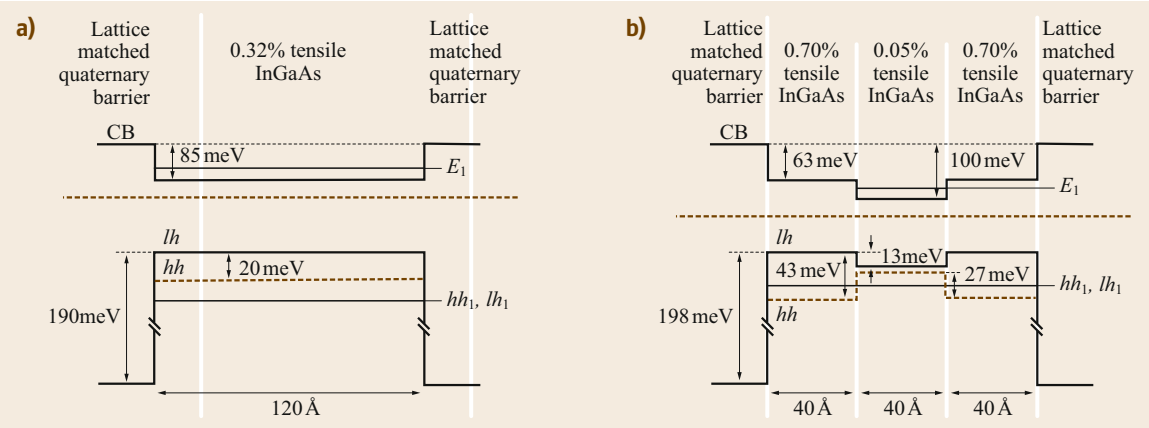


Fig. 35.31a,b Calculated CB and VB profiles for **(a)** conventional tensile QW and **(b)** multi-strain-stepped QW. (After [35.41])

refinements brought about by employing three zones of different strain within the same well. This is illustrated in Fig. 35.25b. In this multistrain well, the tensile strain is larger in the outer thirds of the well. Thus, when a field is applied and the hole wavefunctions are moved toward the edge of the well, the heavy-hole level sees a smaller effective well width and its change in energy can be made the same as for the light hole. Figure 35.32b shows that excellent matching of the movements of the heavy- and light-hole bands can be achieved with the three-zone well over the full range of practical electric fields, up to 100 kV/cm.

35.5.3 High-Speed High-Power QCSE Modulators

Strained-layer QCSE modulators have been successfully integrated with strained-layer lasers to produce high-speed low-chirp sources operating at 1.55 μm for optical communications. The laser and the modulator consist of the same wave-guide structure but are used in the forward and reverse bias, respectively. In this case, there are no polarization problems because if, for example, compressively strained wells are used the laser will produce the TE-polarized output, which is efficiently modulated by the modulator. One problem that is encountered, however, is that while the modulator is in the absorbing state, a significant density of photogenerated carriers may accumulate in the QWs. This has two effects. Firstly, the carriers tend to fill the band-edge states, thus reducing the field in the well, and in addition bleaching the absorption process. Secondly, they may take a long time to escape, so that efficiency of switching to the nonabsorbing state is dependent on how long the modulator was absorbing and what carrier density had accumulated. In practice, it is found that the escape rate of photogenerated holes

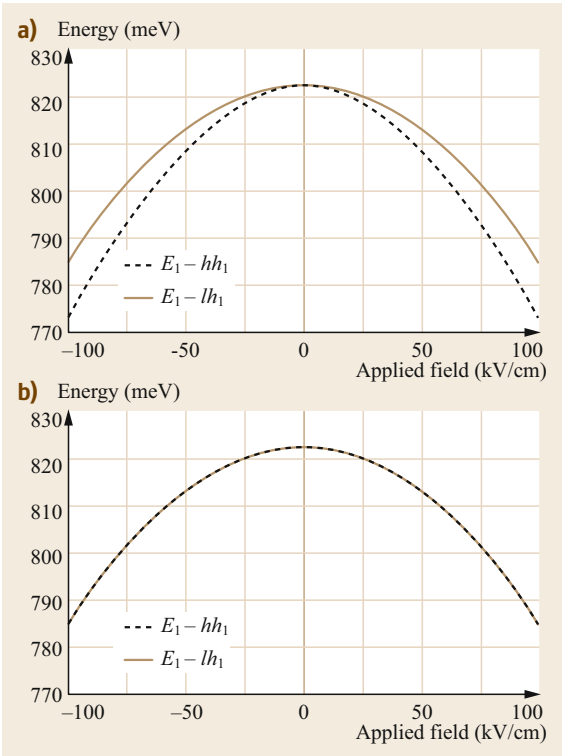


Fig. 35.32a,b Calculated quantum confined Stark shifts of the E_1–hh_1 (*dashed lines*) and E_1–lh_1 (*solid lines*) transitions for structures **(a)** and **(b)** in Fig. 35.25. (After [35.41])

from the QWs is slower than that for electrons and is the limiting process. This problem may be addressed by changing from the standard InGaAsP material system to InGaAlAs, which has a much smaller VB offset than in the equivalent InGaAsP layers [35.101]. Alternatively, the problem may be reduced by band-structure engineering in InGaAsP using strain.

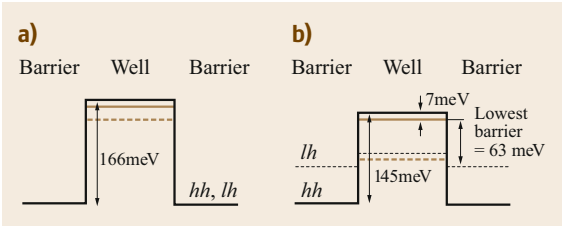


Fig. 35.33a,b Calculated VB edge energies and band alignments for **(a)** lattice-matched modulator and **(b)** strained modulator, as in Fig. 35.28. The *thick horizontal lines* indicate the highest hh confined state

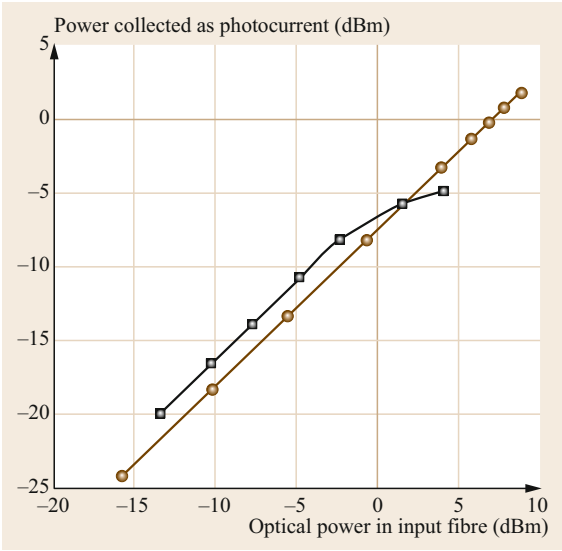


Fig. 35.34 Total power absorbed in lattice-matched (*open squares*) and strained (*closed squares*) electroabsorption modulator as a function of power in the input fiber. (After [35.43])

Figure 35.33a shows the VB structure of a lattice-matched InGaAsP-based electroabsorption modulator designed to operate around 1.5 μm. The heavy-hole well was calculated to be 166 meV deep, and the structure showed power saturation, as illustrated by the open squares in Fig. 35.34. However, when a 1% compressive strain was introduced into the well and 1% tensile strain into the barriers between the wells, the calculated excitation energy from the heavy-hole states in the well to light-hole states in the barrier was reduced to 63 meV, as illustrated in Fig. 35.33b. Since the strain in the barrier has little effect on the well depth for heavy holes, the heavy-hole wavefunctions and quantum confined Stark effect should be comparable in both structures, but now thermal excitation out of the well by phonon scattering to the light-hole level in the barrier is greatly enhanced. The improvement in the power-saturation characteristics of the strained modulator is shown by the solid squares in Fig. 35.34; a considerable improvement in the 10 GHz switching performance was also observed.

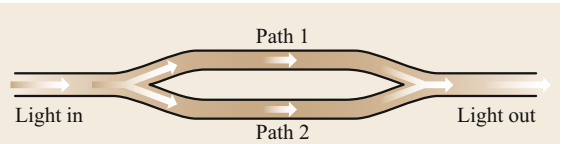


Fig. 35.35 Schematic of a Mach–Zender interferometer

35.5.4 The Electro-Optic Effect

The real part of the refractive index can be described in terms of virtual transitions between the VB and CB and is therefore related to the imaginary part of the refractive index. Hence, when an electric field applied to a semiconductor leads to a change in the absorption characteristics as described above, there is also an accompanying change in the real part of the refractive index. This results in a phase change of the light reaching the end of the active region of the modulator. While this causes unwanted chirp in the electroabsorption modulators described above, it can be used to advantage in modulators and switches that employ interference effects. A good example of such a device is the modulator based on a Mach–Zender interferometer. In this structure, the incoming wave is divided into two equal components at a Y-junction and then recombined at a similar junction, as shown in Fig. 35.35. If the two arms of the interferometer are exactly equivalent then the two beams will arrive in-phase and constructively recombine at the output. However, if the refractive index for light traveling down one path is increased by the electro-optic effect so that its phase at the output is changed by π, then the two beams will destructively interfere at the output and no light will be transmitted. Band-structure considerations for the electro-optic effect are ones with which we are already familiar. The linewidth enhancement factor α_1 is defined by

$$\alpha_1 = -\frac{4\pi}{\lambda}\left(\frac{\frac{dN}{dg}}{\frac{dg}{dn}}\right), \qquad (35.42)$$

and provides a measure of the ratio of the change in the real part N of the refractive index to the change in the imaginary part g of the refractive index, each as a function of carrier density n. Thus, for modulators based on the electro-optic effect, we would wish to have a band structure that results in a large α_1, while for electroabsorption modulators, we require α_1 to be as small as possible.

35.6 Photodetectors

In the first part of this chapter, we dealt with optoelectronic devices designed to produce light for use in applications including displays, indicators, data storage, and communications. The latter two of these applications also require the ability to detect the light once it has been created. Much of the explosion of interest in semiconductor optoelectronic devices over the past 20 or so years has been particularly aimed at the communications market. This has put high demands on the ability to detect small optical signals being transmitted at very fast data rates ($> 40\,\mathrm{Gbit/s}$). In this section, we consider the important aspects for detectors and describe some of the approaches that have been used.

35.6.1 Photodetector Requirements

There are many attributes important to the design of photodetectors. These can broadly be split into the following categories:

- Sensitivity
- Speed
- Noise
- Physical size (footprint)
- Reliability
- Temperature sensitivity
- Ease of use
- Finally cost.

When deciding on a particular photodetector, one must decide which of these factors is of most importance to the particular application of interest.

35.6.2 Photodetection Theory

The primary requirement of the band structure of semiconductor optical detectors is that the band gap of the active region is less than the photon energy to be absorbed. The electron–hole pair formed when the photon is absorbed is then separated by an applied or built-in field that exists in the active region. If, under the influence of the field, the electron and hole move apart by a distance, x, then the charge induced to flow in the external circuit connected to the detector is x/l. In a simple photoconductive detector consisting of a bulk semiconductor with conducting electrodes, l is the distance between the electrodes. In a p–n or p–i–n diode detector, l is the width of the depletion region.

Several factors need to be taken into account when considering the band structure of a detector. Firstly, the absorption depth of the light to be detected must be matched to the width of the active region, so that there is a high probability that the photogenerated carriers will be produced within the active region or less than the minority-carrier diffusion length from it. Figure 35.36 shows the absorption coefficient as a function of energy for direct-gap GaAs and for indirect-gap Si. An absorption coefficient of $10^3\,\mathrm{cm}^{-1}$, typical for Si, corresponds to an absorption depth of $10\,\mu\mathrm{m}$, while $10^4\,\mathrm{cm}^{-1}$ for GaAs corresponds to a $10\,\mu\mathrm{m}$ absorption depth. One must also consider the effect of the band structure on the dark current in the device, since this sets the lower limit on the strength of the signal that can be detected. There are three important effects that can give rise to the generation of electron–hole pairs even when no light is falling on the device. These are:

(a) Thermal generation
(b) Band-to-band tunneling
(c) Avalanche breakdown.

Processes (a) and (b) are illustrated in Fig. 35.37. Thermal generation, which occurs even at the zero elec-

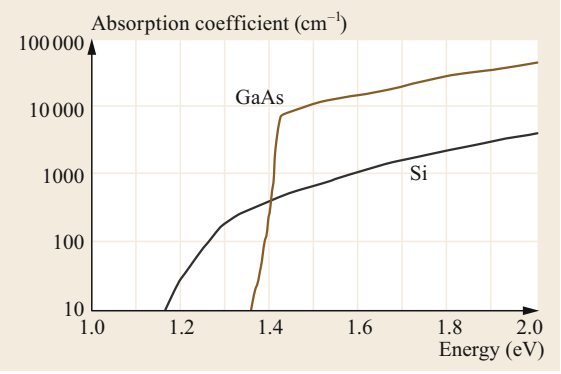


Fig. 35.36 Absorption coefficient as a function of energy for direct-gap GaAs and indirect-gap Si

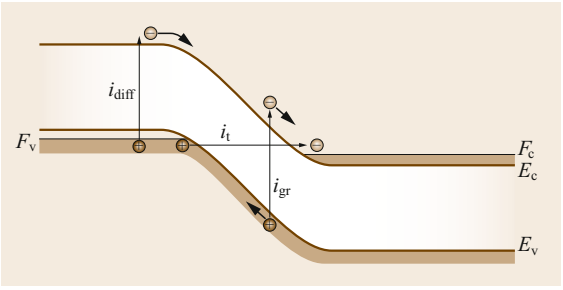


Fig. 35.37 Mechanisms contributing to the dark current in a reverse-biased p–i–n diode. Thermal generation has contributions from diffusion of minority carriers into the field region i_{diff} and from thermal generation within the field region i_{gr}. Zener (band-to-band) tunneling is indicated by the horizontal arrow i_t

tric field, can itself be conveniently divided into two parts. First, there is the part i_{diff} due to the diffusion of minority carriers into the field region. Second, there is the current i_{gr} due to thermal generation within the field region. The thermal generation current increases strongly with temperature but is relatively independent of the applied field, although i_{gr} increases slightly with a reverse bias since this causes the depletion width to increase. Process (b), band-to-band tunneling, is indicated by the horizontal arrow i_t in Fig. 35.37. At large reverse fields, electrons at the top of the VB see a triangular-shaped potential barrier through which it is possible for them to tunnel into the CB. This quantum-mechanical tunneling is relatively independent of temperature but is strongly dependent on the applied field. Process (c), avalanche breakdown, which occurs at very large electric fields, has to be avoided except in avalanche photodetectors (APDs), which will be described in some detail below. Avalanche breakdown involves carriers with initial kinetic energies greater than the band-gap energy E_g away from the CB minimum. This means that electrons in wide-band gap III–V semiconductors will be at energies such that they are scattering between the central Γ minimum, the three X minima at the edges of the Brillouin zone in the ⟨100⟩ directions, and the four L minima at the edges of the Brillouin zone in the ⟨111⟩ directions. Hydrostatic pressure measurements [35.103] showed that the effective ionization energy ⟨E⟩ could be related to the average CB energy simply by

$$\langle E\rangle = \frac{1}{8}(E_\Gamma + 3E_X + 4E_L)\,, \qquad (35.43)$$

where E_Γ, E_X, and E_L are the energies of the Γ, X, and L minima above the top of the VB, respectively.

The first two effects giving rise to a dark current, described in (a) and (b) above, increase almost exponentially with decreasing band gap E_g. Therefore, there exists an optimum value for the band gap E_g to be used to detect photons of a particular energy E_{ph}. Considering E_g to be the adjustable parameter, which can be determined by the alloy composition, as E_g is reduced below E_{ph}, the absorption coefficient increases and light becomes more efficiently absorbed within the active region, of the diode. However, once all the photons are absorbed, reducing E_g further adds nothing to the photocurrent but does continue to increase the dark current. Therefore, it is clear that the optimum band gap is just below that of the photon energy to be detected. The exact difference E_{ph}–E_g depends upon the shape of the absorption edge of the material, the thickness of the active region and its position within the device.

35.6.3 Detectors with Internal Gain

In a photoconductive detector, a gain can actually be achieved by adding impurities that trap one type of carrier, forming a center with a low-capture cross-section for the other type of carrier. Assuming the trapped carrier is a hole, as the free electron reaches the anode, another will be injected from the cathode, and this process will continue until an electron recombines with the trapped hole. Under these circumstances, x exceeds l and a gain of x/l is achieved: $x/l = \tau_r/t_t$, where t_t is the transit time of the electron between the electrodes and τ_r is the electron–hole recombination lifetime. The response time of such a system is $\tau_r = t_t x/l$ so that, as the gain increases, the speed of the device decreases. A related effect is obtained in a phototransistor. If, for example, we consider a photon being absorbed in the base or collector region of an n–p–n transistor, then the presence of the hole, which will be effectively trapped in the base region, will cause the transit of electrons between the emitter and the collector for the recombination lifetime of the hole. The gain of the system is then equal to the gain of the transistor, which is related to τ_r/t_t, where t_t is now the transit time of the electron from the emitter to collector.

35.6.4 Avalanche Photodetectors

Interestingly, silicon as well as direct-band gap materials can be used to produce efficient detectors, but for different reasons. In the direct-band gap materials, photons with energies greater than the band gap are strongly absorbed by the creation of electron–hole pairs, leading to the generation in the detector of a primary photocurrent or photovoltage. In silicon, photons with energies just above the band gap are relatively weakly absorbed, but the electrons generated can be greatly amplified by the process of avalanche multiplication. This occurs when a large reverse voltage is applied to the p–n or p–i–n junction of the detecting diode. Electrons drifting in the field can gain kinetic energy in excess of the band-gap energy E_g. It is then possible for the hot carrier to give up its energy and momentum in an impact ionization process, resulting in the generation of an electron–hole pair. The generated carriers can, in turn, drift in the electric field and themselves generate further carriers. The band structure of silicon is particularly well suited for avalanche multiplication since impact ionization can occur for hot electrons with conservation of energy and momentum when the electron's kinetic energy is almost exactly equal to E_g [35.104]. By contrast, a hot hole in silicon requires an energy close to $3E_g/2$ before it can generate an electron–hole pair. Thus, in silicon, a large avalanche multiplication can be

achieved with each electron that is optically generated and swept across the avalanche region. This situation is illustrated in Fig. 35.38a. Very large gains can be obtained, but all of the carriers are swept out of the active region in a time equal or less than $t_e + t_h$, where t_e and t_h are the transit times taken by electrons and holes, respectively, to cross the active region.

In contrast, it is found that, in the majority of direct-band gap materials, holes and electrons are able to produce impact ionization at the same electric field. As a result, above this field, avalanching occurs for both types of carrier as they move in opposite directions and the carrier density escalates uncontrollably, leading to complete avalanche breakdown. This situation is illustrated in Fig. 35.38b. Gain, but also breakdown, occurs when the field is large enough to create an impact ionization process just as the electron is about to leave the active region. If exactly the same happens for the hole, the situation can continue ad infinitum, leading to an infinite gain, but a current pulse that does not switch off. Of course, in fact, impact ionization is a statistical process and the current pulse might stop after one electron transit or run away as described, leading to an unstable noisy detection system.

Silicon is therefore clearly a very useful material for the production of APDs. Unfortunately, however, the detection bandwidth of silicon limits it from the visible spectrum up to $\approx 1\,\mu m$ in the near-infrared IR. It is of significant technological importance to have fast high-gain low-noise detectors above this wavelength, in particular in the range $1.3-1.6\,\mu m$ for optical communications. The mid-infrared is also of importance due to the large number of gas absorption bands in this range. Due to the effects discussed above, the performance of APDs in the near-mid-infrared region is poor when compared with silicon APDs. Recently, *Adams* proposed that APDs based upon InGaAsN may give rise to improved characteristics for detection in the near- and mid-IR [35.105]. As discussed earlier in this chapter, InGaAsN is already showing great promise for long-wavelength lasers due to the unusually large band-gap bowing brought about by small concentrations of nitrogen in GaAs. This clearly makes it a potential candidate

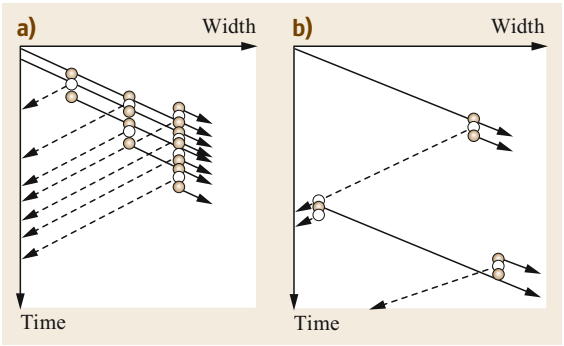


Fig. 35.38a,b Avalanche multiplication as a function of position and time where **(a)** the electron multiplication rate (*solid line*) is much greater than that for holes (*dashed line*). **(b)** When the electron and hole multiplication rates are equal, the current pulse continues for a long time

for the manufacture of long-wavelength detectors. However, in addition to the large band-gap bowing, InGaAsN has an unusually large CB effective mass derived from the dilute amount of nitrogen in the lattice. This has several effects of benefit to APDs. Firstly, the high effective mass inhibits electrons from gaining kinetic energy under an applied field. Secondly, the increased mass, and hence increased density of states, will increase the probability of electron scattering. As the electrons heat up and their energy approaches that of the N levels, they will be strongly scattered. Furthermore, if any electrons manage to reach the top of the E^- band, they have to be scattered across, or tunnel through, the energy gap to the next conduction (E^+) band before they can gain more energy. All of these effects prevent electrons from avalanching. However, since the VB is largely unaffected by the presence of nitrogen, the *holes* may be accelerated by the field and avalanche. It has been proposed that such devices will have a multiplication of $\approx 10^4$ compared with < 10 in the best currently available APDs at these wavelengths. Furthermore, it is also predicted that the noise factor of these devices could be reduced by two orders of magnitude. Although yet to be demonstrated in practice, InGaAsN may in the future provide both *high-Q detectors* and emitters in the infrared.

35.7 Conclusions

This chapter introduced some of the most important optoelectronic devices in use today. It has discussed how these devices were developed based upon highly innovative work which has transformed semiconductor growth technology over the past 40 years. The requirements for lasing action in semiconductors were discussed as well as the ways in which low-dimensional structures and strain may be used to improve the properties of devices such as lasers (both edge-emitting and surface-emitting), modulators, amplifiers, and detectors. The chapter considered current state-of-the-art semiconductor laser devices and discussed emerging future directions in semiconductor optoelectronic materials and devices.

Acknowledgments. It is a pleasure to acknowledge the many people with whom the authors have worked with over the years. In particular, we would like to thank the staff and students, past and present, at the University of Surrey, and at our UK and international academic and commercial partners.

References

35.1 N. Holonyak Jr., S.F. Bevacqua: Appl. Phys. Lett. **1**, 82 (1962)
35.2 A. Einstein: Phys. Z. **18**, 121 (1917)
35.3 M.G.A. Bernard, G. Duraffourg: Phys. Status Solidi **1**, 699 (1961)
35.4 M.I. Nathan, W.P. Dumke, G. Burns, F.H. Dill Jr., G. Lasher: Appl. Phys. Lett. **1**, 62 (1962)
35.5 R.N. Hall, G.E. Fenner, J.D. Kingsley, T.J. Soltys, R.O. Carlson: Phys. Rev. Lett. **9**, 366 (1962)
35.6 T.M. Quist, R.H. Rediker, R.J. Keyes, W.E. Krag, B. Lax, A.L. McWhorter, H.J. Zeiger: Appl. Phys. Lett. **1**, 91 (1962)
35.7 H. Nelson, J.I. Pankove, F. Hawrylo, G.C. Dousmanis: Proc. IEEE **52**, 1360 (1964)
35.8 Z.I. Alferov, V.M. Andreev, D.Z. Garbuzov, Y.V. Zhilyaev, E.P. Morozov, E.L. Portnoi, V.G. Trofim: Sov. Phys. Semicond. **4**, 1573 (1971)
35.9 R. Dingle, C.H. Henry: Quantum Effects in Heterostructure Lasers, US Patent (Application), 3982207 (1976)
35.10 A.R. Adams: Electron. Lett. **22**, 249 (1986)
35.11 E. Yablonovitch, E.O. Kane: J. Light. Technol. **LT-4**, 504 (1986)
35.12 Y. Arakawa, H. Sakaki: Appl. Phys. Lett. **40**, 939 (1982)
35.13 G. Liu, A. Stintz, H. Li, K.J. Malloy, L.F. Lester: Electron. Lett. **35**, 1163 (1999)
35.14 Y. Sakakibara, K. Furuya, K. Utaka, Y. Suematsu: Electron. Lett. **16**, 456 (1980)
35.15 M.-C. Amann, J. Buus: *Tunable Laser Diodes* (Artech House, Boston 1998)
35.16 K. Iga: IEEE J. Sel. Top. Quant. Electron. **6**, 1201 (2000)
35.17 J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho: Science **264**, 553 (1994)
35.18 F. Capasso, C. Gmachl, R. Paiella, A. Tredicucci, A.L. Hutchinson, D.L. Sivco, J.N. Baillargeon, A.Y. Cho, H.C. Liu: IEEE J. Sel. Top. Quant. Electron. **6**, 931 (2000)
35.19 T. Kojima, M. Tamura, H. Nakaya, S. Tanaka, S. Tamura, S. Arai: Jpn. J. Appl. Phys. **37**, 4792 (1998)
35.20 P.J.A. Thijs, L.F. Tiemeijer, J.J. Binsma, T. van Dongen: IEEE J. Quant. Electron. **30**, 477 (1994)
35.21 G. Jones, A. Ghiti, M. Silver, E.P. O'Reilly, A.R. Adams: IEE Proc. J. **140**, 85 (1993)
35.22 A. Valster, A.T. Meney, J.R. Downes, D.A. Faux, A.R. Adams, A.A. Brouwer, A.J. Corbijn: IEEE J. Sel. Top. Quant. Electron. **3**, 180 (1997)
35.23 E.P. O'Reilly: Semicond. Sci. Technol. **4**, 121 (1989)
35.24 T. Yamamoto, H. Nobuhara, K. Tanaka, T. Odagawa, M. Sugawara, T. Fujii, K. Wakao: IEEE J. Quant. Electron. **29**, 1560 (1993)
35.25 M.P.C.M. Krijn, G.W. t'Hooft, M.J.B. Boermans, P.J.A. Thijs, T. van Dongen, J.J.M. Binsma, L.F. Tiemeijer: Appl. Phys. Lett. **61**, 1772 (1992)
35.26 G. Jones, A.D. Smith, E.P. O'Reilly, M. Silver, A.T.R. Briggs, M.J. Fice, A.R. Adams, P.D. Greene, K. Scarrott, A. Vranic: IEEE J. Quant. Electron. **34**, 822 (1998)
35.27 A. Ghiti, M. Silver, E.P. O'Reilly: J. Appl. Phys. **71**, 4626 (1992)
35.28 S.J. Sweeney, A.F. Phillips, A.R. Adams, E.P. O'Reilly, P.J.A. Thijs: IEEE Phot. Tech. Lett. **10**, 1076 (1998)
35.29 Y. Zou, J.S. Osinski, P. Godzinski, P.D. Dapkus, W. Rideout, W.F. Sharfin, R.A. Logan: IEEE J. Quant. Elec. **29**, 1565 (1993)
35.30 G.P. Agrawal, N.K. Dutta: *Long-Wavelength Semiconductor Lasers* (Van Nostrand, New York 1986)
35.31 A.R. Adams, M. Asada, Y. Suematsu, S. Arai: Jap. J. Appl. Phys. **19**, L621 (1980)
35.32 R. Fehse, S. Jin, S.J. Sweeney, A.R. Adams, E.P. O'Reilly, H. Riechert, S. Illek, A.Y. Egorov: Electron. Lett. **37**, 1518 (2001)
35.33 S.J. Sweeney, A.R. Adams, E.P. O'Reilly, M. Silver, P.J.A. Thijs: In: *Conf. Las. Elect. Opt* (IEEE, San Francisco 2000)
35.34 W.W. Lui, T. Yamanaka, Y. Yoshikuni, S. Seki, K. Yokoyama: Phys. Rev. B. **48**, 8814 (1993)
35.35 S.J. Sweeney, G. Knowles, T.E. Sale, A.R. Adams: Phys. Status Solidi B. **223**, 567 (2001)
35.36 J.I. Pankove: IEEE J. Quant. Electron. **QE-4**, 119 (1968)
35.37 E.P. O'Reilly, M. Silver: Appl. Phys. Lett. **63**, 3318 (1993)
35.38 N. Hatori, K. Otsubo, M. Ishida, T. Akiyama, Y. Nakata, H. Ebe, S. Okumura, T. Yamamoto, M. Sugawara, Y. Arakawa: In: *Proc. 30th Eur. Conf. Opt. Comm* (IEEE, Stockholm 2004)
35.39 B.W. Hakki, T.I. Paoli: J. Appl. Phys. Lett. **46**, 1299 (1975)
35.40 L.A. Coldren, S.W. Corzine, M.L. Mashanovitch: *Diode Lasers and Photonic Integrated Circuits* (Wiley, New York 2012)
35.41 K.J. Ebeling: *Integrated Optoelectronics* (Springer, Berlin, Heidelberg 1993)
35.42 L. Coldren, S. Corzine: Continuously-tunable single-frequency semiconductor lasers, IEEE J. Quant. Electron. **23**(6), 903–908 (1987)
35.43 M. Yamaguchi, T. Numai, Y. Koizumi, I. Mito, K. Kobayashi In: *Opt. Fib. Comm. Conf. Reno, Nevada, TUC4* (1987)
35.44 K. Kobayashi, I. Mito In: *Opt. Fib. Comm. Conf., Reno, Nevada, WC1* (1987)
35.45 H. Venghaus, N. Grote (Eds.): *Fibre Optic Communication: Key Devices* (Springer, Berlin, Heidelberg

2012)
35.46 F. Bachmann, P. Loosen, R. Popware: *High Power Diode Lasers: Technology and Applications* (Springer, New York 2007)
35.47 G. Sarlet, G. Morthier, R. Baets: J. Lightwave Tech. **18**(8), 1128 (2000)
35.48 K. Iga: Jpn. J. Appl. Phys. **47**, 1 (2008)
35.49 C.W. Wilmsen, H. Temkin, L. Coldren (Eds.): *Vertical-Cavity Surface-Emitting Lasers* (Cambridge Univ. Press, Cambridge 1999)
35.50 O. Svelto: *Principles of Lasers*, 5th edn. (Springer, New York 2009)
35.51 G. Knowles, S.J. Sweeney, T. Sale, A.R. Adams: IEE Proc. Optoelect. **148**, 256 (2001)
35.52 T.E. Sale, S.J. Sweeney, G. Knowles, A.R. Adams: Phys. Stat. Sol. (b) **223**, 587 (2001)
35.53 A.B. Ikyo, I.P. Marko, A.R. Adams, S.J. Sweeney, A. Bachmann, K. Kashani-Shirazi, M.C. Amann: IET Optoelectron. **3**, 305 (2009)
35.54 K. Hild, T.E. Sale, T.J. Hosea, C. Hirotani, M.Y. Mizuno, T. Kato: IEE Proc. Optoelectronics **148**, 220 (2001)
35.55 E. Kapon (Ed.): *Semiconductor Lasers I and II* (Academic, New York 1998)
35.56 J.J. Coleman, C.A. Bryce, C. Jagadish: *Advances in Semiconductor Lasers* (Academic, New York 2012)
35.57 T.E. Sale: *Vertical Cavity Surface Emitting Lasers* (Research Studies Press, London 1995)
35.58 Y. Ohiso, C. Amano, Y. Itoh, H. Takenouchi, T. Kurokawa: IEEE J. Quant. Electron. **34**, 1904 (1998)
35.59 M. Kondow, S. Kitatani, S. Nakatsuka, M.C. Larson, K. Nakahara, Y. Yazawa, M. Okai, K. Uomi: IEEE J. Sel. Top. Quant. Electron. **3**, 719 (1997)
35.60 T. Kitatani, K. Nakahara, M. Kondow, K. Uomi, T. Tanaka: Jpn. J. Appl. Phys. **39**, L86 (2000)
35.61 K.D. Choquette, J.F. Klem, A.J. Fischer, O. Blum, A.A. Allerman, I.J. Fritz, S.R. Kurtz, W.G. Breiland, R. Sieg, K.M. Geib, J.W. Scott, R.L. Naone: Electron. Lett. **36**, 1388 (2000)
35.62 G. Steinle, H. Riechert, A.Y. Egorov: Electron. Lett. **37**, 93 (2001)
35.63 R. Averbeck, G. Jaschke, L. Geelhaar, H. Riechert: In: *Proc. 19th Int. Semi. Las. Conf., Conference Digest* (IEEE, Piscataway 2004)
35.64 M.A. Wistey, S.R. Bank, H.B. Yuen, L.L. Goddard, J.S. Harris: J. Vac. Sci. Tech. B **22**, 1562 (2004)
35.65 M. Hempel, J.W. Tomm, M. Ziegler, T. Elsaesser, N. Michel, M. Krakowski: App. Phys. Lett. **97**, 231101 (2010)
35.66 S.J. Sweeney, L.J. Lyons, A.R. Adams, D. Lock: IEEE Sel. Top. Quant. Elect **9**, 1325 (2003)
35.67 H.G. Treusch, A. Ovtchinnikov, X. He, M. Kanskar, J. Mott, S. Yang: IEEE J. Selected Topics Quant. Electron. **6**, 601 (2000)
35.68 P. Crump, H. Wenzel, G. Erbert, P. Ressel, M. Zorn, F. Bugge, S. Einfeldt, R. Staske, U. Zeimer, A. Pietrzak, G. Tränkle: IEEE Photon. Technol. Lett. **20**, 1378 (2008)
35.69 D.A. Vinokurov, V.A. Kapitonov, A.V. Lyutetskii, D.N. Nikolaev, N.A. Pikhtin, A.V. Rozhkov, N.A. Rudova, S.O. Slipchenko, A.L. Stankevich, N.V. Fetisova, M.A. Khomylev, V.V. Shamakhov, K.A. Borshchev, I.S. Tarasov: Tech. Phys. Lett. **32**(8), 712 (2006)
35.70 P. Crump, C. Frevert, H. Hösler, F. Bugge, S. Knigge, W. Pittroff, G. Erbert, G. Tränkle: IEEE J. Sel. Top. Quant. Electron. **19**(4), 1501211 (2013)
35.71 P. Schreiber, B. Hoefer, P. Dannberg, U.D. Zeitner: Proc. SPIE **5876**, 1 (2005)
35.72 D. Botez, D.R. Scifres: *Diode Laser Arrays* (Cambridge Univ. Press, Cambridge 1994)
35.73 N.W. Carlson: *Monolithic Diode-Laser Arrays* (Springer, Berlin, Heidelberg 1994)
35.74 J. Mukherjee, M. Ziegler, J. LeClech, J.W. Tomm, B. Corbett, J.G. McInerney, J.P. Reithmaier, S. Deubert, A. Forchel: In: *Novel in-plane semiconductor lasers VIII*, SPIE Proc. Photonics West, Vol. 7230 (2009) p. 72300W
35.75 O. Okhotnikov: *Semiconductor Disk Lasers-Physics and Technology* (Wiley, Weinheim 2010)
35.76 R. Pagano, J. Mukherjee, P. Sajewicz, B. Corbett: IEEE J. Quant. Elect. **47**(4), 439 (2011)
35.77 J. Mukherjee, J.G. McInerney: IEEE Sel. Top. Quant. Elect. **13**(5), 1180 (2007)
35.78 M. Hempel, J.W. Tomm, M. Baeumler, H. Konstanzer, J. Mukherjee, T. Elsaesser: AIP Advances **1**, 042148 (2011)
35.79 J. LeClech, M. Ziegler, J. Mukherjee, J.W. Tomm, T. Elsaesser, J.P. Landesman, B. Corbett, J.G. McInerney, J.P. Reithmaier, S. Deubert, A. Forchel, W. Nakwaski, R.P. Sarzala: J. Appl. Phys. **105**, 014502 (2009)
35.80 J. Mukherjee, J.G. McInerney: Phys. Rev. A **79**(5), 053813 (2009)
35.81 K. O'Brien, S.J. Sweeney, A.R. Adams, B.N. Murdin, A. Salhi, Y. Rouillard, A. Joullie: Appl. Phys. Lett. **89**, 051104 (2006)
35.82 B.A. Ikyo, I.P. Marko, A.R. Adams, S.J. Sweeney, C.L. Canedy, I. Vurgaftman, C.S. Kim, M. Kim, W.W. Bewley, J.R. Meyer: Appl. Phys. Lett. **99**, 021102 (2001)
35.83 D.J. Paul: Laser Photonics Rev **4**(5), 610 (2010)
35.84 R.E. Kazarinov, R.A. Suris: Soviet Physics-Semiconductors **5**, 707 (1971)
35.85 J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A. Hutchinson, A.Y. Cho: Science **264**, 553 (1994)
35.86 J. Faist: *Quantum Cascade Lasers* (Oxford Univ. Press, Oxford 2013)
35.87 R. Kohler, A. Tredicucci, E. Beltram, H.E. Beere, E.H. Linfield, A.G. Davies, D.A. Ritchie, R.C. Lotti, E. Rossi: Nature **417**, 156 (2002)
35.88 S. Adachi: *Properties of Group-IV, III-V and II-VI Semiconductors* (Wiley, Chichester 2005)
35.89 I. Vurgaftman, J.R. Meyer, L.R. Ram-Mohan: J. Appl. Phys. **89**, 5815 (2001)
35.90 S.J. Sweeney, S. Jin: J. App. Phys. **113**, 043110 (2013)
35.91 H. Liu, T. Wang, Q. Jiang, R. Hogg, F. Tutu, F. Pozzi, A. Seeds: Nature Photon. **5**, 416 (2011)
35.92 S. Srinivasan, Y. Tang, G. Read, N. Hossain, D. Liang, S.J. Sweeney, J.E. Bowers: IEEE Microelectron. **33**, 22 (2013)
35.93 S.J. Sweeney: Light-emitting semiconductor device, WO Patent (Application), 149978 (2010)

35.94 D. Liang, J.E. Bowers: Nature Photon. **4**, 511 (2010)
35.95 G. Read, I.P. Marko, N. Hossain, S.J. Sweeney: J. Sel. Top. Quant. Electron. **21**(6), 1502208 (2015)
35.96 S. Jin, S.J. Sweeney: J. Appl. Phys. **114**, 21303 (2013)
35.97 P. Ludewig, N. Knaub, N. Hossain, S. Reinhard, L. Nattermann, I.P. Marko, S.R. Jin, K. Hild, S. Chatterjee, W. Stolz, S.J. Sweeney, K. Volz: Appl. Phys. Lett. **102**, 242115 (2013)
35.98 I.P. Marko, P. Ludewig, Z.L. Bushell, S.R. Jin, K. Hild, Z. Batool, S. Reinhard, L. Nattermann, W. Stolz, K. Volz, S.J. Sweeney: J. Phys. D **47**, 345103 (2014)
35.99 S.J. Sweeney, I.P. Marko, S.R. Jin, K. Hild, Z. Batool, P. Ludewig, L. Natterman, Z. Bushell, W. Stolz, K. Volz, C.A. Broderick, M. Usman, M. Harnedy, E.P. O'Reilly, R. Butkute, V. Pacebutas, A. Geiutis, A. Krotkus: In: *Proc. Semiconductor Laser Conf. (ISLC)* (2014) p. 80
35.100 K. Volz, A. Beyer, W. Witte, J. Ohlmann, I. Németh, B. Kunert, W. Stolz: J. Crystal Growth **315**(1), 37 (2011)
35.101 L.F. Tiemeijer, P.J.A. Thijs, T. van Dongen, R.W.M. Slootweg, J.M.M. van der Heijden, J.J.M. Binsma, M.P.C.M. Krijn: Appl. Phys. Lett. **62**, 826 (1993)
35.102 F. Devaux, S. Chelles, A. Ougazzaden, A. Mircea, J.C. Harmand: Semicond. Sci. Technol. **10**, 887 (1995)
35.103 J. Allam, A.R. Adams, M.A. Pate, J.S. Roberts: Inst. Phys. Conf. **112**, 375 (1990)
35.104 A.R. Adams, J. Allam, I.K. Czajkowski, A. Ghiti, E.P. O'Reilly, W.S. Ring: Strained-layer lasers and avalanche photodetectors. In: *Condensed Systems of Low Dimensionality*, ed. by J.L. Beeby (Plenum, New York 1991) p. 623
35.105 A.R. Adams: Electron. Lett. **40**, 1086 (2004)

36. Liquid Crystals

Geoffrey Luckhurst, David Dunmur

This chapter outlines the basic physics, chemical nature and properties of liquid crystals. These materials are important in the electronics industry as the electro-optic component of flat-panel liquid-crystal displays, which increasingly dominate the information display market.

Liquid crystals are intermediate states of matter which flow like liquids, but have anisotropic properties like solid crystals. The formation of a liquid-crystal phase and its properties are determined by the shape of the constituent molecules and the interactions between them. While many types of liquid-crystal phase have been identified, this Chapter focuses on those liquid crystals which are important for modern displays.

The electro-optical response of a liquid crystal display (LCD) depends on the alignment of a liquid-crystal film, its material properties and the cell configuration. Fundamentals of the physics of liquid crystals are explained and a number of different displays are described.

In the context of materials, the relationship between the physical properties of liquid crystals and their chemical composition is of vital importance. Materials for displays are mixtures of many liquid-crystal compounds carefully tailored to optimise the operational behaviour of the display. Our current understanding of how chemical structure determines the physical properties is outlined, and data for typical liquid-crystal compounds are tabulated. Some key references are given, but reference is also made to more extensive reviews where additional data are available.

36.1 **Introduction to Liquid Crystals** 933
36.1.1 Calamitic Liquid Crystals 935
36.1.2 Chiral Liquid Crystals 937
36.1.3 Discotic Liquid Crystals 938

36.2 **The Basic Physics of Liquid Crystals** 940
36.2.1 Orientational Order 940
36.2.2 Director Alignment 940
36.2.3 Elasticity 941
36.2.4 Flexoelectricity 944
36.2.5 Viscosity 945

36.3 **Liquid-Crystal Devices** 947
36.3.1 A Model Liquid-Crystal Display: Electrically Controlled Birefringence (ECB) Mode 948
36.3.2 High-Volume Commercial Displays: The Twisted Nematic (TN) and Super-Twisted Nematic (STN) Displays 951
36.3.3 Complex LC Displays and Other Cell Configurations 952

36.4 **Materials for Displays** 956
36.4.1 Chemical Structure and Liquid-Crystal Phase Behaviour 956
36.4.2 The Formulation of Liquid-Crystal Display Mixtures 957
36.4.3 Relationships Between Physical Properties and Chemical Structures of Mesogens 958

References 964

36.1 Introduction to Liquid Crystals

Liquid crystals have been known for almost 130 years but it is only in the last 40 years or so that their unique application in display devices has been recognised. Now they are seen as extremely important materials having made possible the development of thin screens for use with personal computers (PCs) and in televisions. In fact a wide range of different liquid-crystal (LC) display devices has been developed. The common feature for each of these is that the optical characteristics of the display are changed on application of an electric field across a thin liquid-crystal film. The process causing this change is associated with a variation

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_36

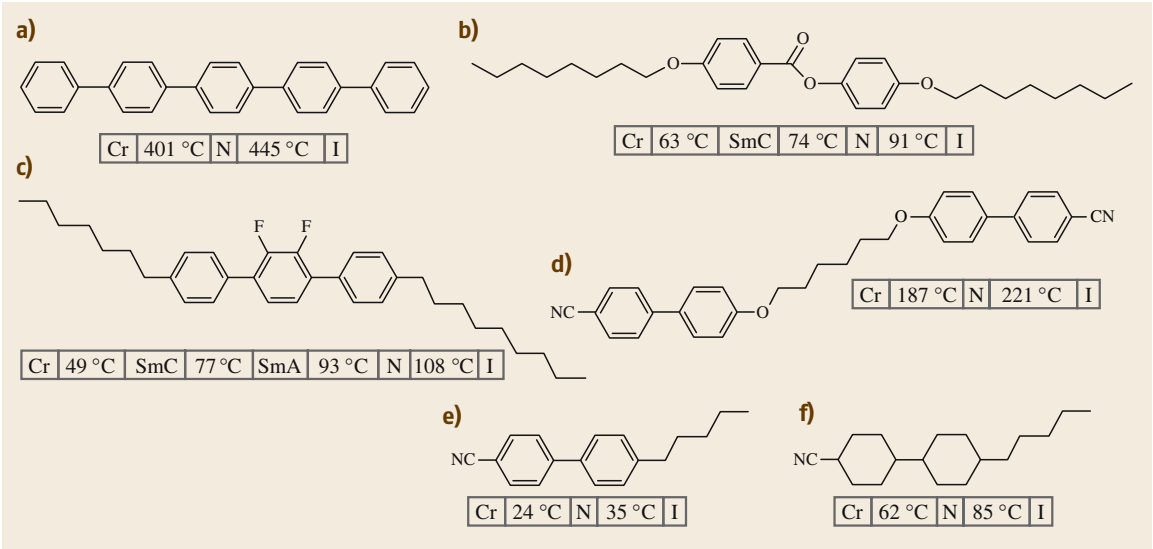


Fig. 36.1 The molecular structures for a selection of compounds which form calamitic liquid crystals

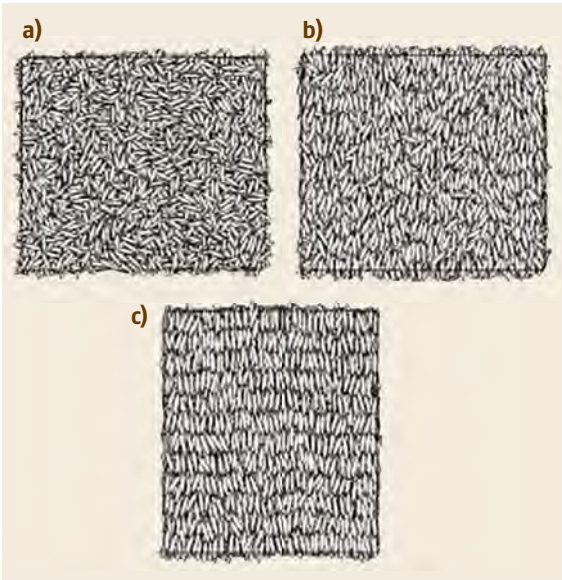


Fig. 36.2a–c The molecular organisation in **(a)** the isotropic phase **(b)** the nematic phase and **(c)** the smectic A phase obtained from the simulation of a Gay–Berne calamitic mesogen

in the macroscopic organisation of the liquid crystal within the cell. The liquid crystal is, therefore, strictly behaving as a molecular material and not an electronic one. Nonetheless the display itself is closely integrated with electronic components. Since liquid crystals may be unfamiliar to those concerned with conventional electronic materials, this section begins with an introduction to liquid crystals and the compounds that form them. The following section describes the basic physics for liquid crystals which are needed to understand their use in display devices. The functioning of the most important displays is described in Sect. 36.3, which makes contact with the basic physics outlined in Sect. 36.2. The liquid-crystal materials used in display devices are discussed in the final section, where the necessary optimisation of a wide range of properties is addressed.

The majority of chemical compounds can exist in three states of matter, namely crystal, liquid or gas, each with its defining characteristics. There is a fourth state known as a liquid crystal and, as the name suggests, this state has characteristics of both crystals and liquids. Thus a liquid crystal flows when subject to a stress, like a liquid, but certain of its properties are anisotropic, like a crystal. This macroscopic behaviour, often used to identify the phase, implies that at the microscopic or molecular level the material has an element of long-range orientational order together with some translational disorder at long range. It is this combination of order and disorder that makes liquid crystals so fascinating and gives them their potential for applications, especially in the field of electro-optic displays.

A variety of different classes of materials are known to form liquid crystals at some point on their phase diagram [36.1]. These include organic materials where the liquid crystal is formed, on heating, between the crystal and isotropic liquid phases. Such materials are known as thermotropic liquid crystals and are the subject of this Chapter. Another class is formed by amphiphilic organic materials in which part of the constituent molecules favours one solvent, normally water, while the other part does not. When the amphiphile is

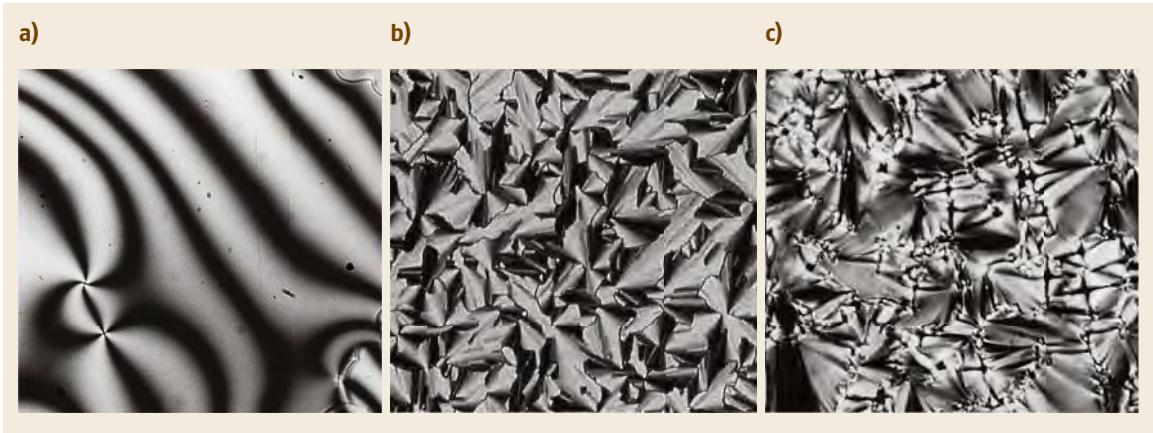


Fig. 36.3a–c Typical optical textures observed with a polarising microscope for **(a)** nematic, **(b)** smectic A and **(c)** columnar liquid-crystal phases

dissolved in the water, the molecules form aggregates which then interact to give the liquid-crystal phase, the formation of which is largely controlled by the concentration of amphiphile. These are known as lyotropic liquid crystals; they underpin much of the surfactant industry, although they are not used in displays and so will not be considered further. Colloidal dispersions of inorganic materials such as clays can also form liquid-crystal phases depending on the concentration of the colloidal particles. Solutions of certain organic polymers also exhibit liquid-crystal phases and, like the colloidal systems, the solvent acts to increase the separation between polymer chains but does not significantly affect their state of aggregation. A prime example of such a system is the structural polymer Kevlar, which for the same weight is stronger than steel; it is formed by processing a nematic solution of the polymer.

The following sections return to thermotropic liquid crystals and describe the molecular organisation within the phases, mention some of their properties and briefly indicate the relationship between the phase and molecular structures.

36.1.1 Calamitic Liquid Crystals

In view of the anisotropic properties of liquid crystals, it seems reasonable that a key requirement for their formation is that the molecules are also anisotropic. This is certainly the case, with the majority of liquid crystals having rod-like molecules, such as those shown in Fig. 36.1. One of the simplest nematogenic rod-like molecules is *p*-quinquephenyl (Fig. 36.1a), which is essentially rigid. However, flexible alkyl chains can also be attached at one or both ends of the molecule (Fig. 36.1b,c) or indeed in the centre of the molecule (Fig. 36.1d), and rigid polar groups such as a cyano (Fig. 36.1d–f) may be attached at the end of the molecule. The rigid part is usually constructed from planar phenyl rings (Fig. 36.1a–e) but they can be replaced by alicyclic rings (Fig. 36.1f) which enhance the liquid crystallinity of the compound. The term calamitic, meaning rod-like, is applied to the phases that they form. There are, in fact, many different calamitic liquid-crystal phases but we shall only describe those which are of particular relevance to display applications.

At an organisational level the simplest liquid crystal is called the nematic and in this phase the rod-like molecules are orientationally ordered, but there is no long-range translational order. A picture showing this molecular organisation, obtained from a computer simulation of a Gay–Berne mesogen [36.2] is given in Fig. 36.2b. The molecular shape is ellipsoidal and the symmetry axes of the ellipsoids tend to be parallel to each other and to a particular direction known as the director.

In contrast there is no ordering of the molecular centres of mass, except at short range. The essential difference between the nematic and isotropic phases (Fig. 36.2a) is the orientational order, which is only short range in the isotropic liquid. At a macroscopic level the nematic phase is characterised by its high fluidity and by anisotropy in properties such as the refractive index. The anisotropic properties for a nematic have cylindrical symmetry about the director, which provides an operational definition of this unique axis and is the optic axis for the phase. The anisotropy in the refractive index combined with the random director distribution results in the turbidity of the phase, which contrasts with the transparency of the isotropic liquid. This on its own would not be sufficient to identify the liquid crys-

tal as a nematic phase but identification is possible from the optical texture observed under a polarising microscope. These textures act as fingerprints for the different liquid-crystal phases. An example of such a texture for a nematic phase is shown in Fig. 36.3a; it is created by the anisotropy or birefringence in the refractive index combined with a characteristic distribution of the director in the sample.

The next level of order within liquid crystal phases is found for the smectic A phase. Now, in addition to the long-range orientational order, there is translational order in one dimension, giving the layered structure shown in Fig. 36.2c [36.2]. The director associated with the orientational order is normal to the layers. Within a layer there is only short-range translational order as in a conventional liquid. In this structure the layer spacing is seen to be comparable but slightly less than the molecular length, as found experimentally for many smectic phases.

At a macroscopic level the layer structure means that the fluidity of a smectic A phase is considerably less than for a nematic phase. The properties are anisotropic and the birefringence is responsible for the turbidity of the phase, as found for a nematic. However, under a polarising microscope the optical texture is quite different to that of a nematic, as is apparent from the focal conic fan texture shown in Fig. 36.3b.

A variant on the smectic A is the smectic C phase. The essential difference to the smectic A phase is that the director in the smectic C phase is tilted with respect to the layer normal. The defining characteristic of the smectic C phase is then the tilt angle, which is taken as the angle between the director and the layer normal. This tilt in the structure reduces the symmetry of the phase to the point group C_{2h} in contrast to $D_{\infty h}$ for nematic and smectic A phases. This lowering in symmetry naturally influences the symmetry of the properties. The fluidity of the smectic C phase is comparable to that of a smectic A phase. However, the optical texture can be quite different and it has elements similar to a nematic phase and to a smectic A; the focal conic fan structure is less well defined and is said to be broken. The nematic-like features result from the fact that the tilt direction of the director is not correlated between the smectic layers and so it adopts a distribution analogous to that of the director in a nematic phase.

The molecular factors which influence the ability of a compound to form a liquid-crystal phase have been well studied both experimentally [36.3] and theoretically [36.4]. Consider the simple nematic as formed, for example, by *p*-quinquephenyl (Fig. 36.1a); this melts at 401 °C to form the nematic phase, which then undergoes a transition to the isotropic phase at 445 °C; the transition temperatures are denoted by T_{CrN} and T_{NI}, respectively. The very high value of T_{NI}, which is a measure of the stability of the nematic phase, is attributed to the large length-to-breadth ratio of *p*-quinquephenyl. In contrast *p*-quaterphenyl, formed by the removal of just one of the five phenyl rings, does not exhibit a liquid-crystal phase at atmospheric pressure, even though its shape anisotropy is still relatively large. This occurs because, on cooling, the isotropic liquid freezes before the transition to the nematic phase can occur. Indeed many compounds with anisotropic molecules might be expected to form liquid-crystal phases, but do not because of their high melting points. As a consequence the molecular design of liquid crystals needs to focus not only on increasing the temperature at which the liquid crystal–isotropic transition occurs but also on lowering the melting point. One way by which this can be achieved is to attach flexible alkyl chains to the end of the rigid core (Fig. 36.1). In the crystal phase the chain adopts a single conformation but in a liquid phase there is considerable conformational disorder, and it is the release of conformational entropy on melting that lowers the melting point. The addition of the chain also affects the nematic–isotropic transition temperature, which alternates as the number of atoms in the chain passes from odd to even. This odd–even effect is especially dramatic when the flexible chain links two mesogenic groups (Fig. 36.1d) to give what is known as a liquid-crystal dimer [36.5]. The odd–even effect is particularly marked because the molecular shapes for the dimers with odd and even spacers differ significantly on average, being bent and linear, respectively.

The attachment of alkyl chains to the rigid core of a mesogenic molecule has another important consequence, as it tends to promote the formation of smectic phases (Fig. 36.1b,c). The reason that the chains lead to the formation of such layered structures is that, both energetically and entropically, the flexible chains prefer not to mix with the rigid core, and so by forming a layer structure they are able to keep apart. Indeed it is known that biphenyl (a rigid rod-like structure) is not very soluble in octane (a flexible chain). The lack of compatibility of the core and the chain increases with the chain length and so along a homologous series it is those members with long alkyl chains that form smectic phases. For example, 4-pentyl-4′-cyanobiphenyl (Fig. 36.1e) forms only a nematic phase whereas the longer-chain homologue, 4-decyl-4′-cyanobiphenyl only exhibits a smectic A phase. To obtain a tilted smectic phase such as a smectic C there clearly needs to be a molecular interaction which favours an arrangement for a pair of parallel molecules that is tilted with respect to the intermolecular vector. Such a tilted structure can be stabilised by electrostatic

interactions; for example by off-axis electric dipoles (Fig. 36.1b,c) [36.6] or with a quadrupolar charge distribution.

The ability of a compound to form a liquid crystal is not restricted to just one phase. The delicate balance of the intermolecular interactions responsible for the various liquid-crystal phases means that transitions between them can result from modest variations in temperature. This is apparent for the 4,4′-dialkyl-2′,3′-difluoroterphenyl shown in Fig. 36.1c, which on cooling the isotropic liquid forms nematic, smectic A and smectic C phases; such a compound is said to be polymorphic. Materials which form even more liquid-crystal phases are known [36.1]. The occurrence of several liquid-crystal phases in a single system can be of value in processing the material for display applications.

36.1.2 Chiral Liquid Crystals

The mesogenic molecules that have been considered so far are achiral in the sense that the molecule is superimposable on its mirror image. Molecules may also be chiral in that they are not superimposable on their mirror images; this chirality can result from the tetrahedral arrangement of four different groups around a single carbon atom. This is illustrated in Fig. 36.4a which shows such an arrangement together with its mirror image; these are known as enantiomers. The presence of a chiral centre will certainly change the nature of the interactions between the molecules and it is relevant to see whether the chiral interactions might not influence the structure of the liquid-crystal phases exhibited by the material. From a formal point of view it might be expected that the molecular chirality of a mesogen should be expressed through the symmetry of the liquid-crystal structure. This proves to be the case, provided no other interactions oppose the chiral deformation of the liquid-crystal phase. In fact the first liquid crystal to be discovered [36.7] was chiral; this was cholesteryl benzoate where the cholesteryl moiety contains many chiral centres. The structure of the liquid crystal is nematic-like in that there is no long-range translational order but there is long-range orientational order. However, the difference between this phase and a nematic formed from achiral materials is that the director is twisted into a helix. The helix may twist in a left-handed or a right-handed sense and these structures, shown in Fig. 36.4b, are mirror images of each other. The phase structure is certainly chiral and so is known as a chiral nematic, although originally it was called a cholesteric phase. The symbol for the phase is N*, where the asterisk indicates that the phase has a chiral structure.

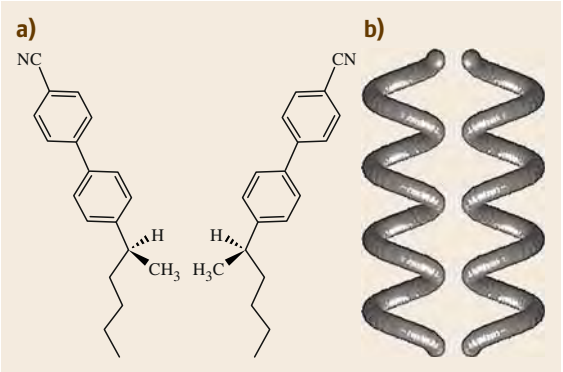


Fig. 36.4 **(a)** A chiral mesogenic molecule, (R) 2-[4′-cyano-4-biphenyl]-hexane, together with its mirror image, (S) 2-[4′-cyano-4-biphenyl]-hexane. **(b)** The left- and right-handed helical organisation of the director for a chiral nematic

The helical structure is characterised by the pitch of the helix p which is the distance along the helix axis needed for the director to rotate by 2π. Since the directions parallel and antiparallel to the director are equivalent the periodicity of the chiral nematic is $p/2$. For many chiral nematics the helical pitch is comparable to the wavelength of visible light. This, together with the periodic structure of the phase, means that Bragg reflection from a chiral nematic will be in the visible region of the spectrum and so this phase will appear coloured with the wavelength of the reflected light being related to the pitch of the helix. This pitch is sensitive to temperature, especially when the chiral nematic phase is followed by a smectic A. This sensitivity has been exploited in the thermochromic application of chiral nematics, where the reflected colour of the phase changes with temperature [36.1].

The chirality of cholesteryl benzoate clearly results from the chiral centres present in the mesogenic molecule. However, the chirality can also be introduced indirectly to a mesogen by simply adding a chiral dopant, which does not need to be mesogenic. The mixture will be chiral and this is sufficient to lead to a chiral nematic. The pitch of this mixture depends on the amount of the dopant and the inverse pitch, p^{-1}, proves to be proportional to its concentration. The handedness of the helix induced by the dopant will depend on its stereochemical conformation and will be opposite for the two enantiomers. Accordingly, if both enantiomers are present in equal amounts, i. e., as a racemic mixture, then doping a nematic with this will not convert it to a chiral nematic.

Chiral smectic phases are also known in which the director adopts a helical structure as a result of introducing molecular chirality into the material either

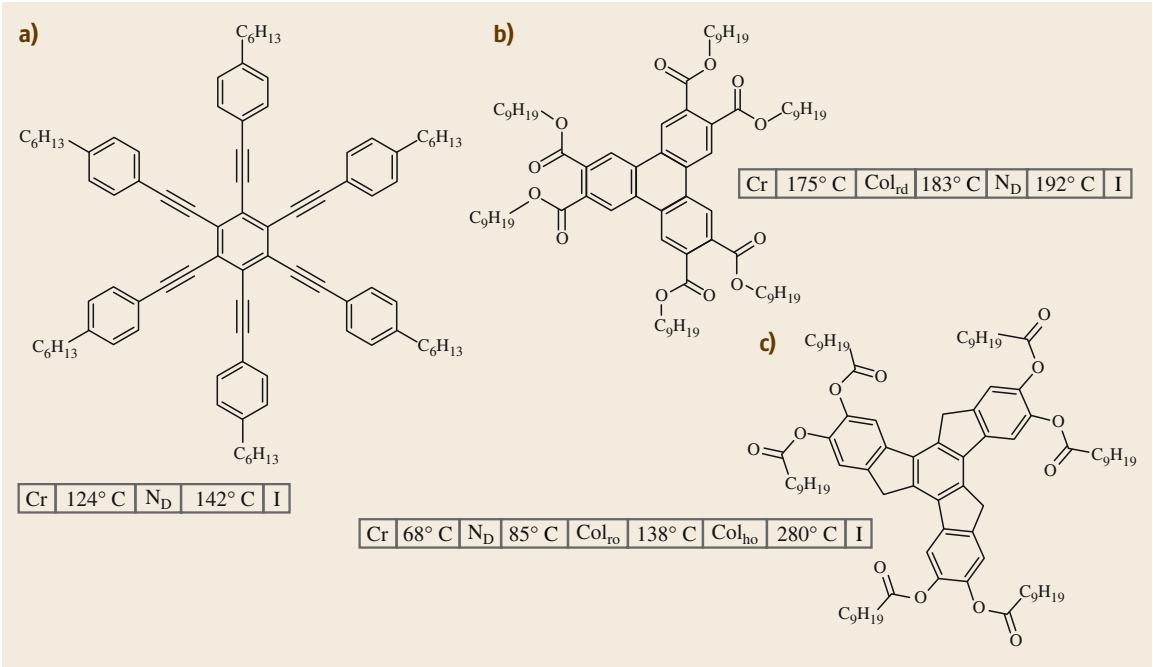


Fig. 36.5 A selection of molecular structures for compounds that form discotic liquid crystals

as a dopant or as an intrinsic part of the mesogenic molecule. The chiral smectic C phase, denoted SmC*, provides an appropriate example with which to illustrate the structure of such phases. In an achiral smectic C phase the tilt direction of the director changes randomly from layer to layer, analogous to the random director orientation in an achiral nematic. For the chiral smectic C phase, as might be anticipated, the tilt direction of the director rotates in a given sense, left-handed or right-handed, and by a small, fixed amount from layer to layer. Other structural features of the smectic C phase remain unchanged. Thus, the director of the chiral smectic C phase has a helical structure with the helix axis parallel to the layer normal. The pitch of the helix is somewhat smaller than that of the associated chiral nematic phase. The magnitude of the pitch is inversely related to the tilt angle of the smectic C, and since this grows with decreasing temperature so the pitch decreases. The reduced symmetry C_2 of the SmC* phase leads to the introduction of a macroscopic electrical polarisation [36.8]. This is of potential importance for the creation of fast-switching displays.

The ability of the SmC* phase to adopt a helical structure results from the fact that the tilt direction for the director acts in an analogous manner to the director in a nematic, and importantly that the layer spacing is preserved in the helical structure. In marked contrast there are strong forces inhibiting the creation of a twisted structure for a smectic A composed of chiral molecules. The director is normal to the layers and so the creation of a twisted structure would require a variation in the layer thickness but this has a high energy penalty associated with it. Accordingly many of the smectic A phases formed from chiral molecules have the same structure as those composed of achiral molecules. There are, however, exceptions and these occur when the chiral interactions are especially strong and, presumably, the translational order of the layers is small. Under such conditions the smectic A structure is partially destroyed, creating small SmA-like blocks about 1000 Å wide, separated by screw dislocations [36.9, 10]. These defects in the organisation allow the directions for the blocks to rotate coherently to give a chiral helical structure; the pitch of the helix is found to be larger than that in an analogous chiral nematic phase. This chiral phase is just one example of a liquid-crystal structure stabilised by defects; it is known as a twist grain-boundary phase and denoted by TGBA*. The letter A indicates that the director is normal to the layers in the small smectic blocks; there is a comparable phase in which the director is tilted, denoted by TGBC*.

36.1.3 Discotic Liquid Crystals

The key requirement for the formation of a liquid crystal is an anisotropic molecule, as exemplified by the rod-like molecules described in the previous sections.

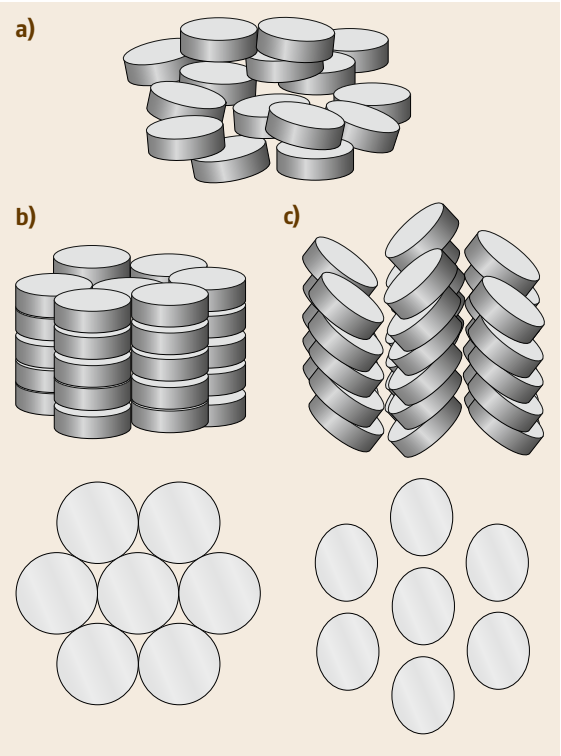


Fig. 36.6a–c The molecular organisation in discotic liquid-crystal phases, **(a)** nematic, **(b)** hexagonal columnar and **(c)** rectangular columnar

However, there is no reason why disc-like molecules should not also exhibit liquid-crystal phases. Nonetheless, it was not until 1977 that the first example of a thermotropic liquid crystal formed from disc-like molecules was reported [36.11]. Since that time several liquid-crystal phases have been identified and these phases are known collectively as discotic liquid crystals. The range of compounds that exhibits these phases is now extensive and continues to grow [36.12], although it does not match the number that form calamitic liquid crystals. The molecular structures of three compounds which form discotic liquid crystals are shown in Fig. 36.5.

As for rod-like molecules, the simplest liquid-crystal phase formed by disc-like molecules is the nematic, usually denoted N_D, where the D indicates the disc-like nature of the molecules. Within the nematic phase, shown in Fig. 36.6a, the molecular centres of mass are randomly distributed and the molecular symmetry axes are orientationally correlated. The nematic structure is the same as for that formed from rod-like molecules, the only difference being that the symmetry axes which are orientationally ordered are the short axes for the discs and the long axes for the rods. The point symmetry of the discotic and calamitic nematic is the same, namely $D_{\infty h}$. The discotic nematic is recognised in the same way as the calamitic nematic; that is it flows like a normal fluid and its anisotropy is revealed by a characteristic optical texture analogous to that shown in Fig. 36.3a. In fact, the refractive index of a discotic nematic along the director is smaller than that perpendicular to it, which is the opposite to that for a calamitic nematic. It has been suggested that this difference may be of value in display devices [36.13] but this concept has not as yet been commercialised.

The other class of discotic liquid crystals possesses some element of long-range translational order and these are known as columnar phases, two examples of which are sketched in Fig. 36.6b,c. The disc-like molecules are stacked face-to-face into columns. A single column has a one-dimensional structure, and as such is not expected to exhibit long-range translational order, although this can result from interactions between neighbouring columns in the liquid-crystal phase. The column–column interactions will result in the columns being aligned parallel to each other; these interactions will also determine how the columns are packed. When the discs are orthogonal to the column axis the cross section is essentially circular and so the columns pack hexagonally, as shown in Fig. 36.6b. The symbol given to this phase is Col_{hd}, where h denotes hexagonal packing of the columns and d indicates that the arrangement along the column is disordered. The point group symmetry of this phase is D_{6h}. The disc-like molecules may also be tilted with respect to the column axis, giving an elliptical cross section to the columns. As a result the columns are packed on a rectangular lattice; there are four possible arrangements and just one of these is indicated in Fig. 36.6c. In general the mnemonic used to indicate a rectangular columnar phase is Col_{rd}. The point group symmetry of the rectangular columnar phase is D_{2h} and the extent to which the structure deviates from that of the Col_{hd} phase will depend on the magnitude of the tilt angle within the column. The columnar phases can be identified from their optical textures and an example of one is shown in Fig. 36.3c.

The columnar phases have potential electronic applications because of the inhomogeneity of the molecules that constitute them, i. e., the central part is aromatic while the outer part is aliphatic. As a result of the overlap between the π-orbitals on the centres of neighbouring discs it should be possible for electrical conduction to take place along the core of the column. This should occur without leakage into adjacent columns because of the insulation provided by the alkyl chains. It should also be possible to anneal these molecular wires because of their liquid-crystal prop-

erties [36.14]. This and the ability to avoid defects in the columns which can prevent electronic conduction in crystals mean that the columnar phase has many potential advantages over non-mesogenic materials. In addition, discotic systems are also used as compensating films to improve the optical characteristics for some liquid-crystal displays.

At a molecular level the factors that are responsible for the formation of the discotic liquid-crystal phases are similar to those for calamitic systems. Thus, the molecular design should aim to increase the liquid crystal–isotropic transition temperature while decreasing the melting point. The latter is certainly achieved by attaching flexible alkyl chains to the perimeter of the rigid disc. The creation of the columnar phases should be relatively straightforward provided the central core is both planar and large. Then, because of the strong attractive forces between the many atoms in the rigid core the molecules will wish to stack face-to-face in a column. The formation of the columns will also be facilitated by the flexible alkyl chains attached to the core. Clearly then it may prove to be difficult to create the nematic phase before the columnar phase is formed unless the disc–disc interactions can be weakened. One way in which this can be achieved is by destroying the planarity of the core, for example, by using phenyl rings attached to the molecular centre so that they can rotate out of the plane (Fig. 36.5a). It is to be expected that the columnar phases should occur below the nematic phase, corresponding to an increase in order with decreasing temperature. This is usually observed, for example, for the hexasubstituted triphenylenes (Fig. 36.5b). However, the truxene derivatives, with long alkyl chains on the perimeter, (Fig. 36.5c) exhibit quite unusual behaviour. For these compounds the crystal melts to form a discotic nematic and then at a higher temperature a columnar phase appears. This deviates from the expected sequence, and because the nematic phase appears at a lower temperature than the columnar phase it is usually referred to as a re-entrant nematic. The occurrence of a re-entrant phase is often attributed to a conformational change which strengthens the molecular interactions with increasing temperature thus making the more ordered phase appear at higher temperatures.

36.2 The Basic Physics of Liquid Crystals

36.2.1 Orientational Order

The defining characteristic of a liquid crystal is the long-range orientational order of its constituent molecules. That is, for rod-like molecules, the molecular long axes tend to align parallel to each other even when separated by large distances. The molecules tend to be aligned parallel to a particular direction known as the director and denoted by $\boldsymbol{n}$. This is an apolar vector, that is $\boldsymbol{n} = -\boldsymbol{n}$, because the nematic does not possess long-range ferroelectric order. The properties of the nematic phase are cylindrically symmetric about the director, which provides a macroscopic definition of this. The anisotropy of the properties results from the orientational order and the extent of this is commonly defined [36.15] by

$$S = \left\langle \frac{1}{2}\left(3\cos^2\beta - 1\right)\right\rangle , \qquad (36.1)$$

although other definitions are possible [36.16]. Here β is the angle made by a molecule with the director and the angular brackets indicate the ensemble average. In the limit of perfect order S is unity while in the isotropic phase S vanishes. The temperature dependence of S is shown in Fig. 36.7 for the nematogen, 4,4′-dimethoxyazoxybenzene; this behaviour is typical of most nematic liquid crystals. At low temperatures S is about 0.6 and then decreases with increasing temperature, reaching about 0.3 before it vanishes discontinuously at the nematic–isotropic transition, in keeping with the first-order nature of this transition. It is also found, both experimentally and theoretically, that the orientational order of different nematic liquid crystals is approximately the same provided they are compared at corresponding temperatures, either the reduced, T/T_{NI}, or shifted, $T_{\mathrm{NI}} - T$, temperatures [36.16]. Since many properties of liquid crystals are related to the long-range orientational order these also vary with temperature especially in the vicinity of the transition to the isotropic phase.

36.2.2 Director Alignment

The director in a bulk liquid crystal is distributed randomly unless some constraint is applied to the system; a variety of constraints can be employed and two of these are of special significance for display applications. One of them is an electric field and because of the inherent anisotropy in the dielectric permittivity of the liquid crystal the director will be aligned. The electric energy density controlling the alignment is given by [36.17]

$$U_{\mathrm{elec}} = -\frac{\varepsilon_0 \Delta\varepsilon}{2}(\boldsymbol{n}\cdot\boldsymbol{E})^2 . \qquad (36.2)$$

Here ε_0 is the permittivity of a vacuum and the scalar product $\boldsymbol{n}\cdot\boldsymbol{E}$ is $E\cos\theta$, where E is the magnitude of the field $\boldsymbol{E}$ and θ is the angle between the director and the field, $\Delta\varepsilon$ is the anisotropy in the dielectric tensor

$$\Delta\varepsilon = \varepsilon_{||} - \varepsilon_{\perp} \,, \tag{36.3}$$

where the subscripts denote the values parallel (||) and perpendicular ($\perp$) to the director. If the dielectric anisotropy is positive then the director will be aligned parallel to the electric field and, conversely, if $\Delta\varepsilon$ is negative, then the director is aligned orthogonal to the field. The molecular factors that control the sign of $\Delta\varepsilon$ will be discussed in Sect. 36.4. Of course, intense electric fields are also able to align the molecules in an isotropic phase but what is remarkable about a nematic liquid crystal is the very low value of the field needed to achieve complete alignment of the director. Thus for a bulk nematic free of other constraints the electric field necessary to align the director is typically about 30 kV/m, although the value does clearly depend on the magnitude of the dielectric anisotropy. This relatively small value results because of the long-range orientational correlations which mean that the field acts, in effect, on the entire ensemble of molecules and not just single molecules.

The other constraint, essential for display devices, is the interaction between the director and the surface of the container [36.18]. At the surface there are two extreme arrangements for the director. One is with the director orthogonal to the surface, the so-called homeotropic alignment. In the other the director is parallel to a particular direction in the surface; this is known as uniform planar alignment. The type of alignment depends on the way in which the surface has been treated. For example, for a glass surface coated with silanol groups a polar liquid crystal will be aligned homeotropically, while to achieve this alignment for a non-polar nematic the surface should be covered with long alkyl chains. In these examples the direct interaction of a mesogenic molecule with the surface produces an orthogonal alignment which is then propagated by the long-range order into the bulk. To achieve uniform planar alignment of the director the surface is coated with a polymer, such as a polyimide, which on its own would result in planar alignment. To force the director to be parallel to a particular direction in the surface the polymer is rubbed which aligns the director parallel to the direction of rubbing. There is still some uncertainty about the mechanism responsible for uniform planar alignment. It might result from alignment of the polymer combined with anisotropic intermolecular attractions with the mesogenic molecules, although it had been thought [36.19] to have its origins in surface groves and the elastic interactions which are described later.

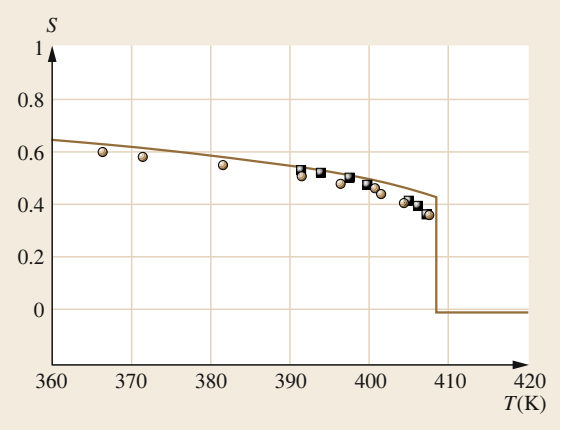


Fig. 36.7 The temperature variation of the orientational order parameter S for 4,4′-dimethoxyazoxybenzene; the different *symbols* indicate results determined with different techniques

The energy of interaction between the surface and the director clearly depends on the nature of the surface treatment and the particular nematic. *Rapini* and *Papoular* [36.20] have suggested the following simple form for the surface energy density

$$U_S = -\frac{A}{2}(\boldsymbol{n}\cdot\boldsymbol{e})^2 \,, \tag{36.4}$$

where A is the anchoring energy and $\boldsymbol{e}$ is the easy axis or direction along which the director is aligned. Clearly it has an analogous form to that for the anisotropic interaction between the nematic and an electric field (36.2) but is essentially phenomenological. The anchoring energy is determined to be in the range 10^{-7}–10^{-5} J/m^2 [36.21]. The upper value corresponds to strong anchoring in that typical values of the electric field would not change the director orientation at the surface. In contrast the lower value is associated with weak anchoring and here the director orientation at the surface can be changed by the field.

36.2.3 Elasticity

As the name suggests, a liquid crystal has some properties typical of crystals and others of liquids. The elastic properties of crystals should, therefore, be reflected in the behaviour of liquid crystals. Here it is the director orientation which is the analogue of the atomic positions in a crystal. In the ground state of a nematic liquid crystal the director is uniformly aligned. However, the elastic torques, responsible for this uniform ground

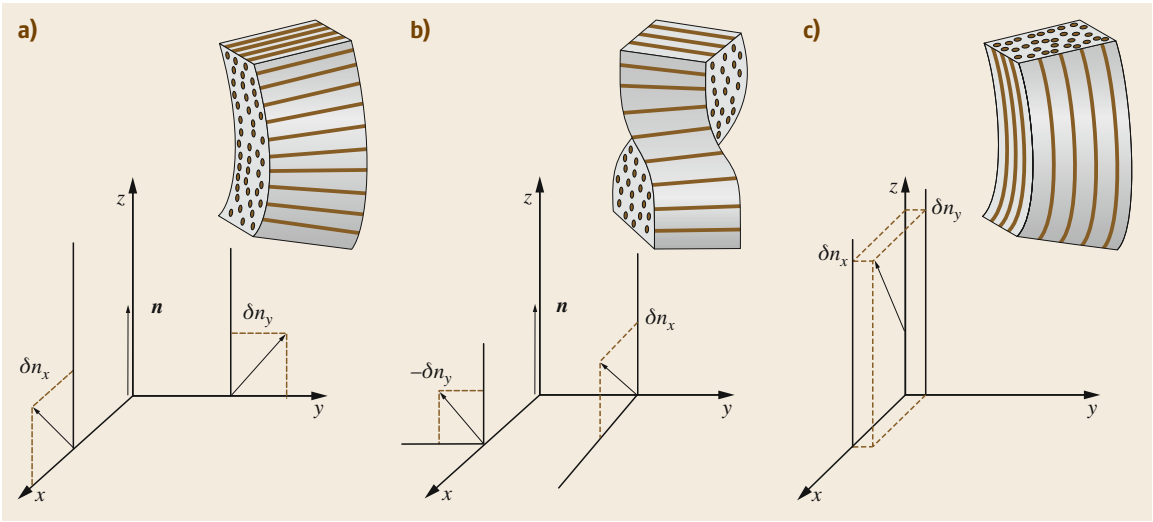


Fig. 36.8a–c The three fundamental deformations for the nematic director (**a**) splay, (**b**) twist and (**c**) bend

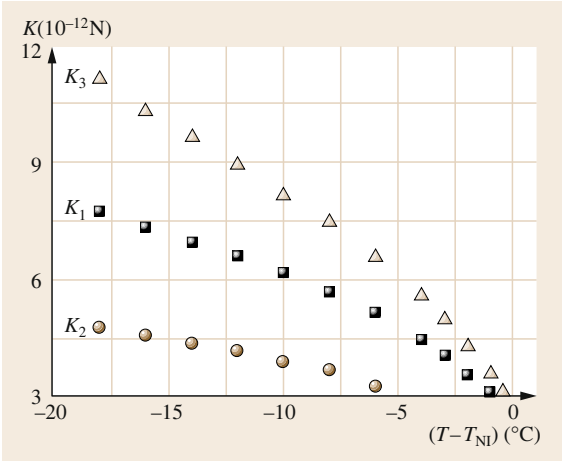


Fig. 36.9 The temperature dependence of the three elastic constants, K_1, K_2 and K_3, for the nematogen 5CB

state, are weak and, at temperatures within the nematic range, the thermal energy is sufficient to perturb the director configuration in the bulk. This perturbed state can take various forms depending on a combination of factors but, whatever the form, it can be represented as a sum of just three fundamental distortion modes. These are illustrated in Fig. 36.8 and are the splay, twist and bend deformations [36.22]. At a more formal level these modes are also shown in terms of the small deviations of the director from its aligned state at the origin as the location from this is varied. For example, for the twist deformation away from the origin along the x-axis there is a change in the y-component of the director, n_y, and the displacement along the y-axis causes a change in n_x. The magnitude of the twist deformation is measured by the gradients $\partial n_y/\partial x$ and $\partial n_x/\partial y$.

The energy needed to stabilise a given distortion of the director field is clearly related to the extent of the deformation via the gradients. This distortion energy density for a bulk nematic liquid crystal is given by continuum theory [36.22] as

$$f = \frac{1}{2}\left[K_1(\nabla\cdot\boldsymbol{n})^2 + K_2(\boldsymbol{n}\cdot\nabla\times\boldsymbol{n})^2 + K_3(\boldsymbol{n}\times\nabla\times\boldsymbol{n})^2\right], \quad (36.5)$$

where the terms $\nabla\cdot\boldsymbol{n}$, $\boldsymbol{n}\cdot\nabla\times\boldsymbol{n}$ and $\boldsymbol{n}\times\nabla\times\boldsymbol{n}$ correspond to the splay, twist and bend deformations, respectively. The contribution each makes to the free energy is determined by the proportionality constants K_1, K_2 and K_3, which are usually known as the Frank elastic constants for splay, twist and bend, respectively. They are small, typically 5×10^{-12} N, and their small magnitude explains why the thermal energy is able to distort the uniform director arrangement so readily. The elastic constants are not in fact constant but vary with temperature and nor are they equal as the results for 4-pentyl-4′-cyanobiphenyl (5CB) shown in Fig. 36.9 demonstrate. The twist elastic constant is seen to be the smallest while the largest is the bend elastic constant. This means that it is easiest to induce a twist deformation in a nematic while a bend deformation is the most difficult to create. All three elastic constants decrease with increasing temperature in keeping with the decreasing order as the transition to the isotropic phase is approached. Like the orientational order the elastic constants vanish discontinuously at the first-order nematic–isotropic transi-

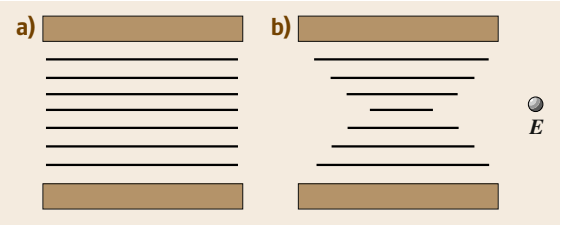


Fig. 36.10a,b The geometry for the Fréedericksz experiment used to determine the twist elastic constant, K_2; **(a)** in zero field and **(b)** above the threshold value

tion.

The continuum theory is especially valuable in predicting the behaviour of display devices and this is illustrated by considering one of the ingenious experiments devised by *Fréedericksz* to determine the elastic constants [36.23]. In these a thin slab of nematic is confined between two glass plates with a particular director configuration, either uniform planar or homeotropic produced by surface forces. These forces control the director orientation just at the two surfaces and the alignment across the slab is propagated by the elastic interactions. A field is then applied which will move the director away from its original orientation and the variation of the director orientation with the field strength provides the elastic constant. To employ the continuum theory in order to describe the experiment it is necessary to add the field energy density (36.2) to the elastic free-energy density in (36.5). The director configuration is then obtained by integrating the free-energy density over the volume of the sample, and minimising this, subject to the surface constraints. The geometry of the experiment is shown in Fig. 36.10 and provided the dielectric anisotropy $\Delta\varepsilon$ is positive the director will move from being orthogonal to the field to being parallel to it. However, the extent of this twist deformation will vary across the cell, being greatest at the centre and zero at the surfaces, in the limit of strong anchoring. The dependence of the director orientation, at the centre of the slab, with respect to the electric field is shown in Fig. 36.11. As the field strength is increased from zero the director orientation remains unchanged until a threshold value E_{th} is reached when the angle between the director and the field starts to decrease continuously. At very high values of the field the director at the centre of the slab tends to be parallel to the field. This behaviour can be understood in the following simple terms. Below the threshold field the elastic energy exceeds the electrical energy and so the director retains its uniform planar alignment. Above the threshold field strength the elastic energy is less than the electrical energy and so

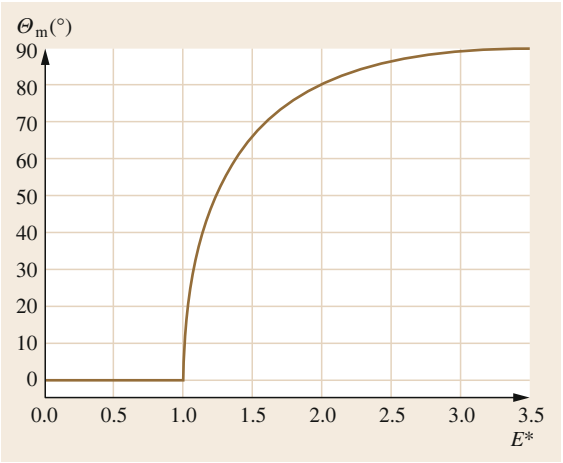


Fig. 36.11 The dependence of the director orientation θ_m at the centre of a nematic slab on the scaled field strength, $E^* = E/E_{th}$

the director begins to move to be parallel to the electric field. The threshold electric field is predicted to be

$$E_{th} = \frac{\pi}{l}\sqrt{\frac{K_2}{\varepsilon_0 \Delta\varepsilon}}\,, \tag{36.6}$$

where l is the slab thickness. Once the threshold field has been measured and provided the dielectric anisotropy is known, the twist elastic constant would be available. Other Fréedericksz experiments with analogous expressions for the threshold field lead to the determination of the splay and bend elastic constants.

For real display applications (Sect. 36.3) the director alignment at the surface deviates from either uniform planar or homeotropic alignment. This deviation is known as a surface pre-tilt and is illustrated in Fig. 36.12a for near-uniform planar alignment with the tilt direction on the two surfaces differing by 180°. The cell with this arrangement has what is known as antiparallel alignment, so named because of the difference in the tilt direction caused by the direction of rubbing on the surfaces being antiparallel. In zero field, therefore, the director is uniformly aligned across the cell but tilted with respect to the x-axis set in the surface. A continuum theory calculation analogous to the case when the pre-tilt angle θ_0 is zero allows the dependence of the director orientation θ_m in the centre of the cell to be determined as a function of the strength of the field applied across the cell. The results of these calculations, for a nematic with positive $\Delta\varepsilon$, are shown in Fig. 36.12b as a function of the scaled

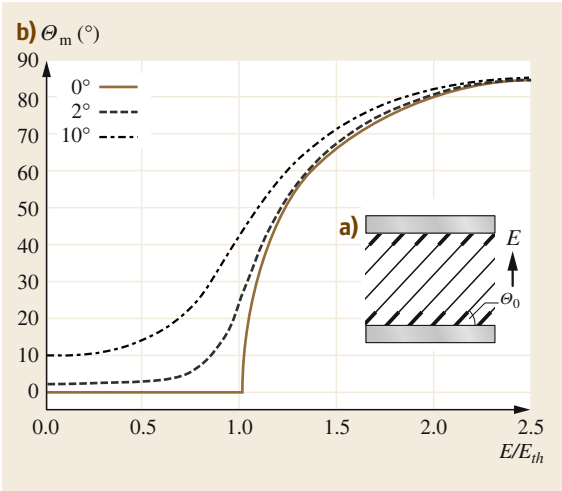


Fig. 36.12 **(a)** The director alignment in a cell with pre-tilt at the two surfaces assembled so that the rubbing directions are antiparallel. **(b)** The electric field dependence of the director orientation θ_m at the centre of the cell for values of the pre-tilt angle of 0° (*solid line*), 2° (*dashed line*) and 10° (*dash-dotted line*)

field strength E/E_{th}, where E_{th} is the threshold field for zero pre-tilt. The theoretical dependence for a tilt angle of zero is analogous to that considered for the twist deformation. In other words, below the threshold field the director is parallel to the easy axis and then above this threshold the director moves to become increasingly parallel to the field. When there is a surface pre-tilt the behaviour is quite different and this is especially apparent for a pre-tilt angle of 10° (Fig. 36.12b). In zero field the angle θ_m made by the director with the x-axis is 10° and, as the field increases, so does θ_m. The rate of increase grows as the threshold field is approached and then is reduced as E/E_{th} increases beyond unity. Comparison of this behaviour with the conventional Fréedericksz experiment (Fig. 36.12b) shows that a pre-tilt of 10° has a significant effect on the way in which the director orientation changes with the field strength. Indeed, even for a pre-tilt of just 2°, there is a pronounced difference in behaviour in the vicinity of the threshold field. This clearly has important implications for the accurate determination of the elastic constants [36.20]. It also shows how unique the behaviour is when the surface pre-tilt angle is zero.

Chiral nematics are often employed in liquid-crystal display devices. Locally, their structure is analogous to a nematic but the director is twisted into a helical structure. The continuum theory for the chiral nematic must, therefore, be consistent with the helical ground state structure of the phase. To achieve this, a constant is

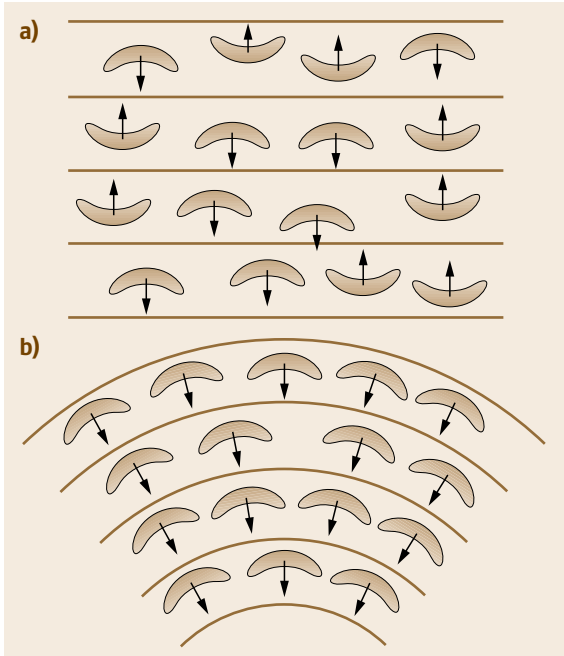


Fig. 36.13a,b The orientational distribution of bent molecules in **(a)** a uniformly aligned nematic and **(b)** one subject to a bend deformation; the x-axes are denoted by *arrows*

added to the twist term in the elastic free-energy density. Thus (36.5) for a nematic becomes

$$f = \frac{1}{2}\left[K_1(\nabla\cdot\boldsymbol{n})^2 + K_2\left(\boldsymbol{n}\cdot\nabla\times\boldsymbol{n} - \frac{2\pi}{p}\right)^2 + K_3(\boldsymbol{n}\times\nabla\times\boldsymbol{n})^2\right] \quad (36.7)$$

for a chiral nematic, where p is the pitch of the helix.

36.2.4 Flexoelectricity

Another property of solids which is mimicked by liquid crystals is piezoelectricity. For solids this is the generation of a macroscopic electrical polarisation as a result of the deformation of certain ionic materials. It is to be expected, therefore, that deformation of the director distribution for a liquid crystal will create a macroscopic polarisation; this proves to be the case [36.24] and the phenomenon is known as flexoelectricity. The origin of flexoelectricity can be understood in the following way. It is generally assumed that mesogenic molecules are cylindrically symmetric but examination of the molecular structure of real mesogens shows that this is not the case. The molecules are asymmetric and at one extreme

can be thought of as bent, as shown in Fig. 36.13; here z is the molecular long axis and x is the axis bisecting the bond angle. When the director is uniformly aligned the z-axis will be parallel to the director and the x-axis will be randomly arranged orthogonal to it (Fig. 36.13). If the director is now subject to a bend deformation the molecular long axis will still tend to be parallel to the local director, however, the x-axis will tend to align parallel to a direction in the plane formed by the bent director. This change in the distribution function for the x-axis will introduce polar order into the system and if there is an electrical dipole moment along the x-axis then the deformed nematic will exhibit a macroscopic polarisation $\boldsymbol{P}$. A similar argument shows that, if the molecule is wedge-shaped, a splay deformation of the director distribution will also induce a polarisation in the nematic [36.24].

The magnitude of $\boldsymbol{P}$, the induced dipole per unit volume, clearly depends on the extent of the deformation in the director distribution. In the linear response regime the induced polarisation for mesogenic molecules of arbitrary asymmetry is given by [36.25]

$$\boldsymbol{P} = e_1 \boldsymbol{n} \nabla \cdot \boldsymbol{n} + e_3 \boldsymbol{n} \times \nabla \times \boldsymbol{n} \,, \tag{36.8}$$

where the vectors $\boldsymbol{n}\nabla\cdot\boldsymbol{n}$ and $\boldsymbol{n}\times\nabla\times\boldsymbol{n}$ represent the splay and bend deformations, respectively. Since the polarisation per unit volume is also a vector then the proportionality constants, e_1 and e_3, are scalars. These are known as the splay and bend flexoelectric coefficients, respectively, and their dimensions are $\mathrm{C\,m^{-1}}$. The determination of individual flexoelectric coefficients is challenging [36.26], however, it seems that $(e_1 + e_3)$ is of the order of 10^{-12}–$10^{-11}\,\mathrm{C\,m^{-1}}$, although from the previous discussion their magnitude should vary amongst the mesogens because of their dependence on the molecular asymmetry and the size and location of the dipole moment. The molecular model proposed to understand the flexoelectricity of nematics suggests that, for rod-like molecules, devoid of asymmetry, the flexoelectric coefficients should vanish, but that does not seem to be the case. Indeed, it has been proposed that polarisation can result even for rod-like molecules if they possess an electrostatic quadrupole moment [36.27]. This is not inconsistent with the polarisation predicted by (36.8) which follows from the reduced symmetry of a nematic with splay and bend deformations of the director [36.24]. Although the existence of flexoelectricity is of considerable fundamental interest the inverse effect in which the director is deformed from a uniform state by the application of an electric field is of relevance for liquid-crystal displays (Sect. 36.3). This deformation occurs because the polarisation induced by the director deformation can couple with the applied electric field and so stabilise the deformation [36.26]. Since the coupling is linear in the electric field then reversal of the field will simply reverse the deformation; a novel bistable device based solely on this reversal is described in Sect. 36.3.

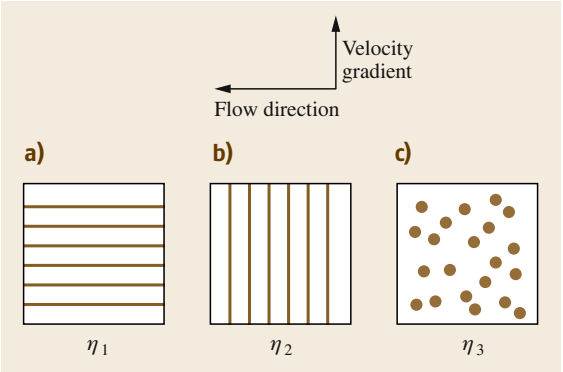


Fig. 36.14a–c The principal flow geometries of the Miesowicz experiments with the director pinned **(a)** parallel to the flow direction, **(b)** parallel to the velocity gradient and **(c)** orthogonal to both the flow direction and velocity gradient

36.2.5 Viscosity

For a nematogen the fluidity of the nematic phase is comparable to that of the isotropic phase appearing at a higher temperature. It is this fluidity, similar to that of a liquid, which is responsible, in part, for the display applications of nematics. An indication of the fluidity of conventional liquids is provided by a single viscosity coefficient η, measured from the flow of the liquid subject to an applied stress in a viscometer. The flow behaviour of a nematic is made more complex by its defining long-range orientational order and the resultant anisotropy of the phase.

This complexity can be appreciated at a practical level in terms of the Miesowicz experiments to determine the viscosity coefficients of a nematic [36.28]. In these experiments it is helpful to consider flow in a viscometer with a square cross section such that there is a velocity gradient orthogonal to the direction of flow. For a nematic, flow through the viscometer will now depend on the orientation of the director with respect to these two axes. A magnetic field is employed to align the director along a particular axis and it must be sufficiently strong that flow does not perturb the director alignment. There are three relatively simple flow geometries and these are shown in Fig. 36.14: with the director parallel to the flow direction (Fig. 36.14a), with the director parallel to the velocity gradient (Fig. 36.14b), and with the direc-

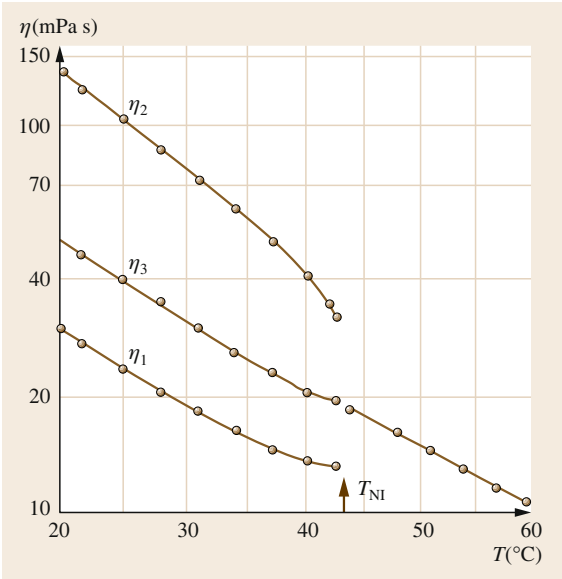


Fig. 36.15 The temperature dependence of the three Miesowicz viscosity coefficients, η_1, η_2, and η_3, for the nematic phase of MBBA; the viscosity coefficient for the isotropic phase is also shown

tor orthogonal to both the flow direction and velocity gradient (Fig. 36.14c). The viscosity coefficients were denoted by Miesowicz as a) η_1, b) η_2, and c) η_3, although other notation has been proposed in which η_1 and η_2 are interchanged [36.29]. The three viscosity coefficients are clearly expected to differ given the anisotropy of the nematic phase and these differences are found to be large. They are illustrated for the room-temperature nematogen 4-methoxybenzylidene-4′-butylaniline (MBBA) [36.30] in Fig. 36.15 where the viscosities are plotted against temperature.

It is immediately apparent that flow is easiest when the director is parallel to the direction of flow, as might have been anticipated. Conversely flow of the nematic is most difficult when the director is orthogonal to the flow direction but parallel to the velocity gradient. The intermediate viscosity η_3 occurs when the director is orthogonal to both the flow direction and the velocity gradient. Here it is seen that this third viscosity coefficient for the nematic is very similar to the viscosity coefficient for the isotropic phase extrapolated to lower temperatures.

The viscosity coefficient $\eta(\theta,\phi)$ for an arbitrary orientation (θ,ϕ) of the director with respect to the flow direction and velocity gradient can be related to the director orientation. It might be expected that this orientation dependence would involve just the three Miesowicz viscosity coefficients, rather like the transformation of a second-rank tensor from its principal axis system. In fact this is not quite the case and a fourth viscosity coefficient η_{12} needs to be introduced. The orientation dependence is then [36.28]

$$\eta(\theta,\phi) = \eta_1 \cos^2\theta + \left(\eta_2 + \eta_{12}\cos^2\theta\right) \times \sin^2\theta\cos^2\phi + \eta_3 \sin^2\theta \sin^2\phi \,. \quad (36.9)$$

We see that for the principal orientations $(0, \pi/2)$, $(\pi/2, 0)$ and $(\pi/2, \pi/2)$ the expression gives η_1, η_2, and η_3, as required. The optimum director orientation with which to determine η_{12} is with the director at 45° to both the flow direction and the velocity gradient; the viscosity coefficient for this is

$$\eta\left(\frac{\pi}{4},0\right) = \frac{\eta_{12}}{4} + \frac{\eta_1+\eta_2}{2} \,. \quad (36.10)$$

The value of η_{12} determined for MBBA proves to be significantly smaller than the other three viscosity coefficients.

The four viscosity coefficients have been defined at a practical level in terms of flow in which the director orientation is held fixed. The converse of these experiments, in which the director orientation is changed in the absence of flow, allows the definition of the fifth and final viscosity coefficient. This is known as the rotational viscosity coefficient; it is denoted by the symbol γ_1 and plays a major role in determining the response times of display devices (Sect. 36.4). To appreciate the significance of γ_1, it is helpful, as for the other four viscosity coefficients, to consider an experiment with which to measure it. In this an electric field is suddenly applied an angle θ to a uniformly aligned director, then providing the dielectric anisotropy is positive the director orientation will be changed and rotates towards the field direction. The electric torque responsible for the alignment is given by

$$\Gamma_{\text{elec}} = -\frac{\varepsilon_0 \Delta\varepsilon}{2} \sin 2\theta \,, \quad (36.11)$$

which is the derivative of the electric energy in (36.2). The rotation of the director is opposed by the viscous torque

$$\Gamma_{\text{visc}} = \gamma_1 \frac{\mathrm{d}\theta}{\mathrm{d}t} \,, \quad (36.12)$$

which is linear in the rate at which the director orientation changes, with the proportionality constant being γ_1. Provided the only constraint on the director is the electric field and provided, $\theta_0 \leq 45°$, the director will move as a monodomain so that the elastic terms vanish. The inertial term for a nematic is small and so the movement of the director is governed by the equation

in which the two torques are balanced. That is, the electric torque causing rotation is balanced by the viscous torque opposing it; this gives

$$\gamma_1 \frac{\mathrm{d}\theta}{\mathrm{d}t} = -\frac{\varepsilon_0 \Delta\varepsilon}{2} E^2 \sin^2\theta \,. \tag{36.13}$$

The solution to this differential equation is

$$\tan\theta = \tan\theta_0 \exp\left(-\frac{t}{\tau}\right), \tag{36.14}$$

where θ_0 is the initial orientation of the director with respect to the electric field and τ is the relaxation time

$$\tau = \frac{\gamma_1}{\varepsilon_0 \Delta\varepsilon E^2} \,. \tag{36.15}$$

Measurement of the time-dependent director orientation allows τ to be determined and from this γ_1, given values of $\Delta\varepsilon$. For MBBA the rotational viscosity coefficient is found to be slightly less than η_2 and to parallel its temperature dependence [36.31].

The five independent viscosity coefficients necessary to describe the viscous behaviour of a nematic have been introduced in a pragmatic manner by appealing to experiments employed to measure these coefficients. However, the viscosity coefficients can be introduced in a more formal way as has been shown by Ericksen and then by *Leslie* in their development of the theory for nematodynamics [36.32]. The Leslie–Ericksen theory in its original form contained six viscosity coefficients, but subsequently Parodi has shown, using the Onsager relations, that there is a further equation linking the viscosity coefficients, thus reducing the number of independent coefficients to five [36.33]. These five coefficients are linearly related to those introduced by reference to specific experiments.

36.3 Liquid-Crystal Devices

The idea to use liquid crystals as electro-optic devices goes back to the early days of liquid-crystal research. In 1918 *Björnståhl*, a Swedish physicist, demonstrated that the intensity of light transmitted by a liquid crystal could be varied by application of an electric field [36.34]. As optical devices of various types became established in the first decades of the 20th century, mainly in the entertainment industry, ways of controlling light intensity became important to the developing technologies. One device that soon found commercial application was the Kerr cell shutter, in which an electric field caused the contained fluid (usually nitrobenzene) to become birefringent. Placing such a cell between crossed polarisers enabled a beam of light to be switched on and off very rapidly. The first report of liquid crystals being of interest for electro-optic devices was in 1936, when the Marconi Company filed a patent [36.35] which exploited the high birefringence of nematic liquid crystals in an electro-optic shutter. However, it was another 35 years before commercial devices became available which used the electro-optic properties of liquid crystals. The long interruption to the development of liquid-crystal devices can be attributed to the lack of suitable materials. We shall see in this Section how the physical properties of liquid crystals determine the performance of devices.

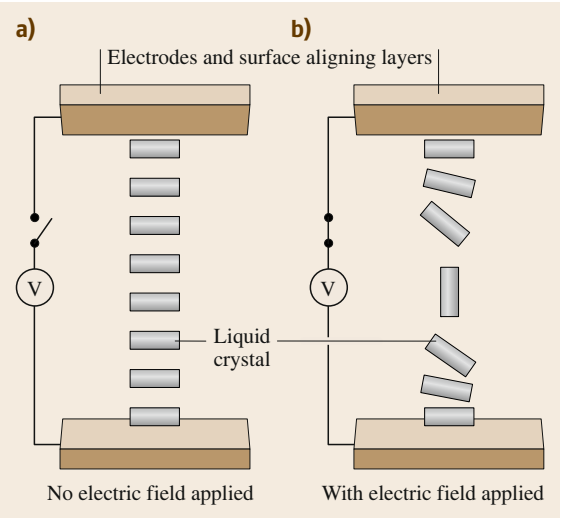


Fig. 36.16 Electrically controlled birefringence (ECB) cell

The optical properties of liquid crystals are exploited in displays, although the operational characteristics of such devices also depend crucially on many other physical properties (Sect. 36.2). Since the devices to be described all depend on the application of an electric field, their operation will be influenced by electrical properties such as dielectric permittivity and electrical conductivity. There is a range of electro-optic effects that can be used in devices, and the precise manner in which the properties of the liquid-crystal materials affect the device behaviour depends on the effect and the configuration of the cell. Thus there is not a single set of ideal properties that can define the best liquid-crystal material, rather the material properties have to

be optimised for a particular application. In this section, different devices will be described, and their dependence on different properties will be outlined. The way in which the material properties can be adjusted for any application will be discussed in the final section of this chapter.

36.3.1 A Model Liquid-Crystal Display: Electrically Controlled Birefringence (ECB) Mode

If an electric field is applied across a film of planar-aligned liquid crystal having a positive dielectric anisotropy, then the director of the liquid crystal will tend to align along the electric field. Thus the director will rotate into the field direction, and the optical retardation or birefringence of the film will change. If the film is placed between crossed polarisers, then the change in optical retardation will be observed as a change in the intensity of transmitted light. The configuration for an ECB-mode display is schematically illustrated in Fig. 36.16. This illustrates the principal components of a liquid-crystal display. The liquid crystal is contained between two glass plates that have been coated with a transparent conducting layer, usually an indium–tin oxide alloy. Such treatment allows the application of an electric field across the liquid-crystal film, which can then be viewed along the field direction. For most applications, the surfaces of the electrodes are treated so that a particular director orientation is defined at the surface. This can be achieved in a variety of ways, depending on the desired surface director orientation. For the ECB-mode display under consideration, the director alignment should be parallel to the glass substrates and along a defined direction, which is the same on both surfaces. The standard technique to produce this alignment is to coat the glass plates with a thin (0.5 nm) layer of polyimide, which is mechanically rubbed in a particular direction. The cell is then placed between crossed polarisers, the extinction directions of which make angles of $\pm 45^\circ$ with the surface-defined director orientation.

The unperturbed state of the liquid-crystal film will be determined by the surfaces which contain it. If these have been treated in such a way that the liquid-crystal director is parallel to the surface along a particular direction, then the film will act as an optical retardation plate. Thus incident polarised light will, in general, emerge elliptically polarised, and there will be a phase retardation between components of the light wave parallel to the fast and slow axes of the retardation plate. For electromagnetic waves polarised at $\pm 45^\circ$ to the director the phase retardation ϕ, in radians, will be determined by the intrinsic birefringence of the liquid crystal ($\Delta n = n_e - n_o$), the film thickness ℓ and the wavelength of the light λ

$$\phi = \frac{2\pi\ell}{\lambda}(n_e - n_o)\,; \tag{36.16}$$

n_e and n_o are respectively the extraordinary (slow) and ordinary (fast) refractive indices of the liquid crystal (assuming that Δn is positive). If the emergent elliptically polarised light passes through a second polariser, crossed with respect to the incident polarisation direction, only a proportion of the incident intensity will be transmitted. The normalised intensity of light transmitted by a pair of crossed polarisers having a birefringent element between them, the axis of which is at $\pm 45^\circ$ to the extinction directions of the polarisers is given by

$$T = \left(\frac{1}{2}\right)\sin^2\frac{\phi}{2} \tag{36.17}$$

(the factor of one half appears for incident unpolarised light – if the light is polarised, as from a laser, then the factor is one).

Thus the initial appearance of the cell will be brightest if the cell thickness, birefringence and wavelength are chosen to give ϕ equal to π, 3π, 5π etc. It is normal to select the cell thickness to give $\phi = \pi$, and under these conditions the display is known as *normally white*. It is possible, though less satisfactory, to configure the display so that it operates in a *normally black* state, corresponding to a phase retardation of a multiple of 2π.

Application of an electric field causes the director orientation to change such that the optical retardation of the cell decreases to zero, and hence the cell becomes non-transmitting, at least in the normally white configuration. Under these circumstances the optical retardation across the cell becomes

$$\phi = \frac{2\pi}{\lambda}\int_0^\ell \left(n_e\left[\theta\,(z)\right] - n_0\right)\mathrm{d}z\,, \tag{36.18}$$

where the effective extraordinary refractive index $n_e[\theta(z)]$ depends on the angle $[90^\circ - \theta(z)]$ between the director and the field and is a function of position z in the cell. This effective index is given by

$$\frac{1}{n_e^2[\theta(z)]} = \frac{\sin^2[\theta(z)]}{n_o^2} + \frac{\cos^2[\theta(z)]}{n_e^2}\,; \tag{36.19}$$

when the director is along the field direction $\theta = 90^\circ$, so $n_e(90^\circ) = n_o$, and $\phi = 0$.

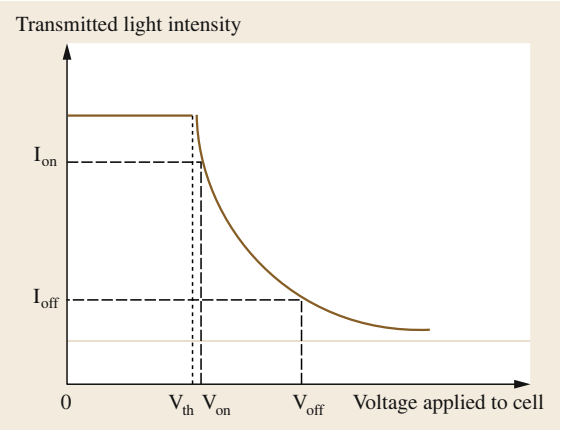


Fig. 36.17 Variation of optical transmission with voltage for an ECB cell between crossed polarisers

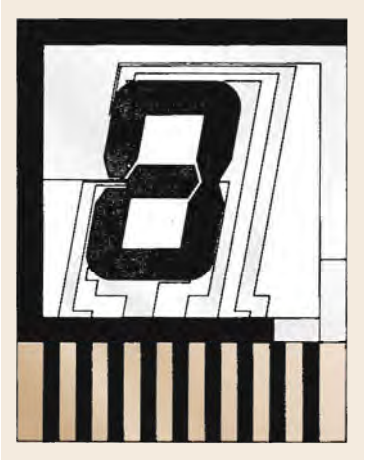


Fig. 36.18 Schematic of a directly addressed seven-segment liquid-crystal display

The orientational distribution of the director in the cell in the presence of an applied electric field is determined by the strength of the field, the electric permittivity and elastic constants of the liquid crystal, and most importantly by the properties at the interface between the liquid crystal and the aligning surfaces. If the director at the surface satisfies the *strong-anchoring* condition, i. e., it is unaffected by the applied electric field, and the surface director is strictly perpendicular to the electric field, then the reorientation of the director exhibits a threshold response, known as a Fréedericksz transition. The change in transmitted intensity as a function of voltage can be calculated using continuum theory and simple optics [36.36], and a typical transmission curve for an ECB cell is illustrated schematically in Fig. 36.17.

The threshold voltage for director reorientation is independent of cell thickness, and is given by

$$V_{\mathrm{th}} = \pi \sqrt{\frac{K_1}{\varepsilon_0 \Delta\varepsilon}} \,, \tag{36.20}$$

where $\Delta\varepsilon$ is the anisotropy in the dielectric permittivity and K_1 is the splay elastic constant. However, it is clear from (36.17) and (36.18) that the intensity of light transmitted by the ECB cell depends on the cell thickness and the wavelength of light. This undermines the usefulness of displays based on the ECB mode, since they require cells of uniform thickness and also they will tend to show colouration in white light. Another important operating characteristic of displays is the *angle of view*, i. e., how the image contrast changes as the angle of incidence moves away from 90° with respect to the plane of the cell. This is clear from Fig. 36.16, where the optical paths in the distorted state for observation to the left or right of the perpendicular to the electrodes are clearly different.

In order to construct a useful display from a simple on/off shutter, it is necessary to consider how image data will be transferred to the display. This is known as *addressing*, and to a large extent it is determined by the circuitry that drives the display. However, we shall see that certain properties of liquid crystals also contribute to the effectiveness of different types of addressing. The simplest method of displaying images on a liquid-crystal display is to form an array of separate cells of the type illustrated by the ECB cell, each having a separate connection for the application of an electric field. Images can then be created by switching on, or off, those cells required to form the image. This technique is known as *direct addressing*, and can be illustrated by the seven-segment displays used in watches and numerical instrument displays (Fig. 36.18).

To create complex displays, a large number of separate cells, known as picture elements or pixels, must be fabricated, usually in the form of a matrix. Providing separate electrical connections for these pixel arrays (e.g., 640×480) for a standard visual graphics array (VGA) computer screen, is impossible, and so other methods of addressing have had to be developed. Historically, the first was the technique known as *passive matrix addressing*, in which the array of cells are identified in rows and columns, and connections are only made to the rows and columns: for an array of $n \times m$ pixels, only $n + m$ connections are made instead of $n \times m$, as required for direct addressing. Each row is activated in turn (sequentially), and appropriate voltage pulses applied to the columns. Only those pixel elements for which the sum of column and row voltages exceeds a *threshold* are switched to an on-state. However, the problem with this method is that many unwanted pixels in the off-state still have a voltage applied, and may

be partially activated in the display: so-called *crosstalk*. The time-sharing of activating signals is known as multiplexing, and it relies on the addressed pixels holding their signal while other pixels which make up the image are activated. Provided that the multiplexing is on a time scale of milliseconds, any fluctuation in the image goes unnoticed. However, if the multiplexing becomes too slow, the image starts to flicker. In fact there is a limit to the number of rows of the matrix which can be addressed ($n_{\max}$), which is related to the ratio of the on-voltage to off-voltage by a result due to *Alt* and *Pleshko* [36.37]:

$$\frac{V_{\text{on}}}{V_{\text{off}}} = \sqrt{\frac{n_{\max}^{1/2}+1}{n_{\max}^{1/2}-1}} \,. \tag{36.21}$$

The cell characteristic which determines the values of V_{on} and V_{off} is the optical transmittance curve, as illustrated in Fig. 36.17. Depending on the desired contrast ratio $I_{\text{on}}/I_{\text{off}}$ for a pixel, then the voltages $V_{\text{on}}/V_{\text{off}}$ are determined. Thus, in the operation of a passively addressed matrix display, there is a tradeoff between the contrast ratio ($I_{\text{on}}/I_{\text{off}}$) and the number of rows, i. e., the complexity of the display. The shape of the transmittance curve is determined by the liquid-crystal material properties, and so these affect the resolution of the image displayed. For a desired contrast ratio of 4, the corresponding on/off ratio for an ECB cell might be 1.82. This gives a maximum number of rows as four, which corresponds to a very-low-resolution display, which would be unusable except for a very basic device. This brief description of a simple electro-optic display operating in the ECB mode illustrates that its performance depends on the dielectric, optical and elastic properties of the liquid-crystal material used. Additionally, the properties at the liquid-crystal-substrate interface and the geometry of the cell will influence the electro-optic response of the cell.

One performance characteristic that is of very great importance is the speed with which the display information can be changed, since this determines the quality and resolution of moving images. The dynamics of fluids are related to their viscosity, and it has already been shown that the viscous properties of liquid crystals are complicated to describe, and are correspondingly difficult to measure. The complication arises because liquid crystals are elastic fluids, and so there is a coupling between the flow of the fluid and the orientation of the director within the fluid. We have seen that the motion of the director can be described in terms of a rotational viscosity, and the optical properties exploited in displays are related to changes in the director orientation. Thus it is the rotational viscosity that is of primary importance in determining the time response of displays. However, the fact that changes of director orientation cause fluid flow in liquid crystals complicates the process.

The time response of a liquid-crystal display pixel can be illustrated by reference to the ECB display, although other cell configurations modify the behaviour to some extent. In what follows, we shall assume that the reorientation of the director within a display pixel does not cause the nematic liquid crystal to flow. For a uniform parallel-aligned nematic-liquid-crystal film, the time for the director to respond depends on the magnitude of the electric field (or voltage) applied to the cell. If the voltage applied is only just greater than the threshold voltage, then the time is very long, while if a large voltage is applied, then the director responds quickly. It is found that the time response can usually be represented as an exponential behaviour, although effects of flow will change this. Neglecting these, a response time τ_{on} can be defined in terms of the change in transmitted-light intensity as

$$\frac{I(t)-I(0)}{I_{\text{on}}-I(0)} = 1-\exp\left[-\left(\frac{t}{\tau_{\text{on}}}\right)\right]. \tag{36.22}$$

and the relaxation time for switching on the display is given approximately by [36.38, 39]

$$\tau_{\text{on}} = \frac{\gamma_1 \ell^2}{\pi^2 K_{\text{eff}}}\left[\left(\frac{V_{\text{on}}}{V_{\text{th}}}\right)^2 - 1\right]^{-1} . \tag{36.23}$$

On removing the applied voltage, the display element returns to its off-state, but with a different relaxation time which is independent of the applied voltage, such that

$$\tau_{\text{off}} = \frac{\gamma_1 \ell^2}{\pi^2 K_{\text{eff}}} ; \tag{36.24}$$

here γ_1 is the rotational viscosity coefficient, and the effective elastic constant K_{eff} that appears in these expressions is the splay elastic constant K_1 for the ECB-mode display. These equations can be modified for other display configurations by changing K_{eff}.

Although the ECB-mode display is the simplest that can be envisaged, based on the Fréedericksz effect, there are many disadvantages, in particular with respect to its viewing characteristics, and it has not been used commercially to any significant extent. However, the apparently simple modification of twisting the upper plate by 90° has resulted in the phenomenally successful twisted nematic display, which represents a large part of today's multi-billion-dollar market.

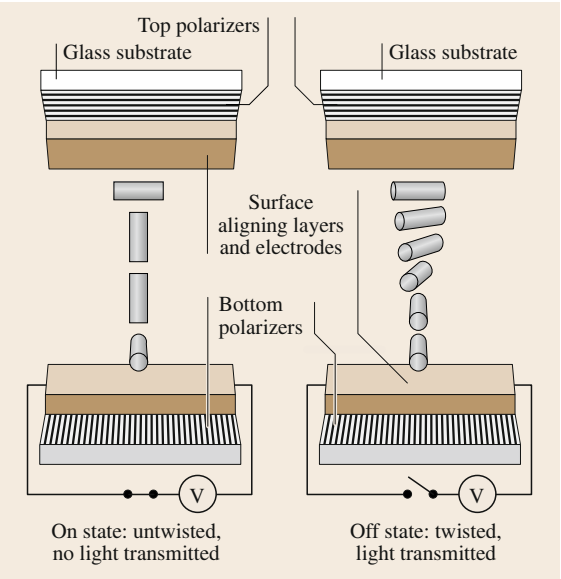


Fig. 36.19 Schematic of a twisted nematic display

36.3.2 High-Volume Commercial Displays: The Twisted Nematic (TN) and Super-Twisted Nematic (STN) Displays

The simple twisted-nematic display is essentially the same as the ECB display depicted in Fig. 36.16, except that the orientations of the surface director at the containing glass plates are rotated by 90°. However, the TN cell, invented by *Schadt* and *Helfrich* in 1970 [36.40], represented a considerable improvement over earlier devices, and rapidly became the preferred configuration for commercial displays. A schematic representation of the TN cell is given in Fig. 36.19.

This twisted configuration for a liquid crystal film was discovered by Mauguin, who found that instead of producing elliptically polarised light, such a twisted film could rotate the plane of polarisation by an angle equal to the twist angle between the surface directors of the glass plates [36.41]. In fact by working through the optics of twisted films, Mauguin showed that perfect rotation of the plane of polarisation only resulted if the film satisfied the following condition

$$2\Delta n\ell \gg \lambda\,. \tag{36.25}$$

Thus for the cell illustrated in Fig. 36.19, the off-state would be perfectly transmitting. This is known as the *normally white* mode. In reality the cells used for TN displays do not meet the Mauguin condition, and the transmission for a 90° twisted cell between crossed polarisers and incident unpolarised light is given by [36.42]

$$T = \frac{1}{2}\left(1 - \frac{\sin^2\left(\frac{\pi}{2}\sqrt{1+u^2}\right)}{1+u^2}\right), \tag{36.26}$$

where $u = \frac{2\Delta n\ell}{\lambda}$. Equation (36.26) shows that, for sufficiently large u, the transmission T is indeed a maximum of 0.5, however it is also a maximum for $u = \sqrt{3}$, $\sqrt{15}$, $\sqrt{35}$, etc. These points on the transmission curve correspond to the *Gooch–Tarry minima*; they are labelled as minima, since they were found for a TN cell operating in the *normally black* state. Most commercial cells operate under conditions of the first or second minima so that thin cells can be used, which give faster responses. It is, therefore, important that the birefringence of the liquid-crystal material can be adjusted to match the desired cell thickness, so that the display can have the best optical characteristics in the off-state.

Application of a sufficiently strong electric field across the twisted film of a nematic liquid crystal having a positive dielectric anisotropy causes the director to align along the field direction. Under these circumstances the film no longer rotates the plane of polarised light, and so appears dark. The transmission as a function of voltage for a twisted cell is similar to that shown in Fig. 36.17, except that the transmission varies more strongly with change in voltage above the threshold, and drops to zero much more rapidly. In contrast to the ECB cell discussed above, the threshold voltage for a TN cell depends on all three elastic constants

$$V_{\mathrm{th}}^{\mathrm{TN}} = \pi\sqrt{\frac{\left(K_1 + \frac{\zeta}{2}(K_3 - 2K_2)\right)}{\varepsilon_0\Delta\varepsilon}}\,, \tag{36.27}$$

where ζ is the twist angle (usually $\pi/2$). The relative change of the transmission intensity with voltage of the TN cell is greater than for the ECB cell, and it can be shown that the steepness of the transmission curve increases as the property ratios K_3/K_1 and $\Delta\varepsilon/\varepsilon_\perp$ decrease. The on/off voltage ratio for a TN cell is closer to unity, than for an ECB cell, and so for similar contrast ratios more lines can be addressed: up to about 20 for typical cells and materials. This is significantly larger than for the ECB cell, and so the TN cell allows more complex images to be displayed. There is still a wavelength dependence for the transmission, although this is less marked than for the ECB mode. However, even with the improved multiplexing capabilities of the TN display over the ECB cell, it is still not good enough to use for computer screens. One very successful approach to solve this problem was to modify the TN cell geometry so that instead of a 90° twist, the directors

on opposite sides of the cell are rotated by about 270°. This is known as a super-twisted nematic cell.

The concept of the 270° super-twisted nematic (STN) display seems at first sight to be irrational [36.43, 44]. The director is an apolar vector and so there should be no difference between a 270° and a 90° twisted cell. However, it is possible to maintain a director twist greater than 90° if the liquid crystal is chiral. The use of chiral additives in 90° TN cells was already established, since a very small quantity of chiral dopant added to a TN mixture would break the left/right twist degeneracy in the cell and so remove patches of reversed twist, giving a much improved appearance to the display. If the amount of chiral dopant was increased, then the chiral liquid-crystal mixture would develop a significant intrinsic pitch. By adjusting the concentration of the chiral dopant, the pitch of the mixture could be matched to the 270° twist across the cell thickness of ℓ, i. e., $p \approx 4\ell/3$. An STN cell operates in the same way as a TN cell, so that an applied electric field causes the director to rotate towards the field direction, thereby changing the optical retardation through the film and the transmission between external crossed polarisers. However, the additional twist in the STN cell has a significant effect on the optical properties of the nematic film.

The optical characteristics of the STN off-state are usually outside the Mauguin condition, which means that polarised light passing through the cell is not guided, and emerges elliptically polarised. The degree of ellipticity is wavelength dependent, and so in white light the off-state appears coloured, as does the on-state. An ingenious solution to this problem is to have two identical STN cells, one behind the other, but where the second compensating cell has a twist of the opposite sense. In operation, only the first of the cells has a voltage applied to it. The compensation cell acts to subtract the residual birefringence of the liquid crystal layer, and the display now switches between white and black. Despite this additional complexity, the huge advantage of the STN display is the rapid change in optical transmission with increasing voltage, and a full optical analysis shows that under optimum conditions the rate of change of transmission with voltage can become infinite. In modern implementations of the STN display, the residual birefringence can be compensated by an optical film, avoiding the need for double cells. It is easy to operate an STN display with a V_{on}/V_{off} ratio of 1.1, which corresponds to an n_{max} of 100. The shapes of the transmission/voltage curve of ECB, TN and STN cells are a direct consequence of the dielectric and elastic properties of the liquid-crystal material, but also depend strongly on the configuration of the cells and the surface alignment

In the description given of displays and their performance, some important aspects have been ignored. The surface orientation of the directors has been assumed to be pinned in the surface plane, which is the requirement for a threshold response. However, it has been found that the performance of displays can be greatly improved if this condition is relaxed, and a pre-tilt is introduced to the cell, such that the surface director may make an angle of up to 60° to the plane of the containing glass plates. This pre-tilt can be introduced by different surface treatments, and it depends on the interfacial properties of the liquid crystal. An important performance characteristic of displays is the angle dependence of the contrast ratio, or more simply how the appearance of the displayed image changes with angle of view. This is largely determined by the display design, and can be accurately calculated from the optical properties of the cell. The refractive indices of the liquid crystal will affect the contrast ratio and angle of view, but precise control of these performance parameters is difficult. Improvements to the appearance in terms of the angle of view or brightness of displays have been achieved by placing precisely manufactured birefringent polymer films behind or in front of liquid-crystal cells.

36.3.3 Complex LC Displays and Other Cell Configurations

The STN configuration described above used passive matrix addressing, and this opened up the possibility of relatively large-area, high-resolution displays, which could be used in laptop computers and other hand-held displays. The next step was to introduce colour by dividing each picture element into three sub-pixels with red, green and blue filters. However, the demand in some market sectors for larger displays with improved appearance having higher resolution (extended graphics array (XGA) displays have 1024×768 pixels) overwhelmed the capability of passive matrix addressing, and the alternative method of active matrix addressing is now used for more complex displays. This technique requires each pixel to have its own switch, as in the simple seven-segment display already described. For high-resolution displays hundreds of thousands of switches are provided by thin-film transistors deposited onto the glass substrate, which forms the screen. These are known as thin-film-transistor twisted-nematic (TFT-TN) displays [36.45]. The sophistication of these displays relies on the capabilities of integrated circuit technology, but the properties of the liquid-crystal materials must still be optimised for the device configuration. A representation of a TFT-TN display is given in Fig. 36.20.

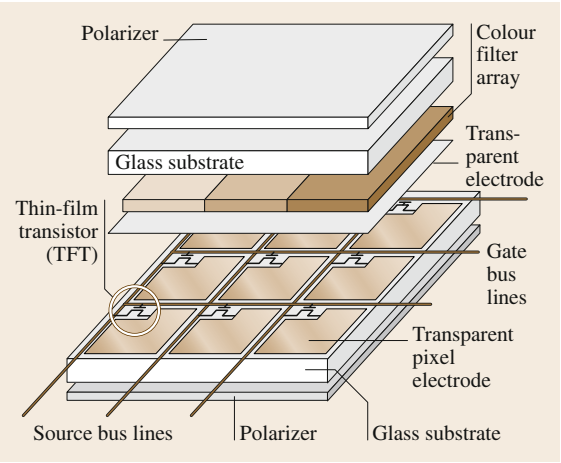


Fig. 36.20 Schematic of a complex colour TFT-TN display

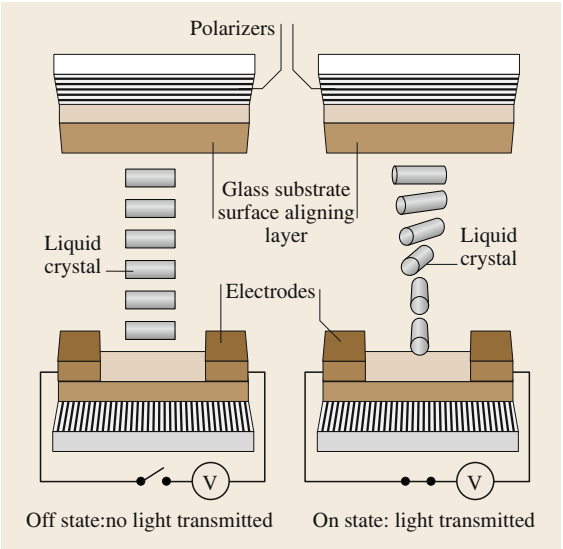


Fig. 36.21 Schematic of an in-plane switching mode display

Although being described as an active matrix, the TFT-TN display still uses sequential addressing of pixel rows, and so activated pixels must remain switched on while the rest of the display image is created. The TFTs provide a source of voltage to each liquid-crystal pixel, which must then hold its charge, as a capacitor, until the image is changed. So another property of liquid crystals becomes important, that of low electrical conductivity, since the charge on a pixel will be lost by conduction through the liquid crystal. This determines the choice of materials for TFT-TN displays. Generally, high dielectric anisotropy is a desirable property for liquid-crystal display mixtures, since it reduces the operating voltage. However, materials with a high dielectric constant tend to have a high electrical conductivity, since charges either from impurities or leached from surfaces will be stabilised in high-dielectric-constant fluids. Thus the selection of suitable materials requires a compromise between its dielectric and conductance properties, and, of course, the all important refractive indices.

Many different cell configurations, which exploit the optical properties of liquid crystals in different ways, have been tried, and some of these have been commercialised to meet particular market requirements. One rather successful approach to the problem of restricted viewing angle has been the development of the in-plane switching (IPS) mode twisted-nematic display [36.46, 47]; this is illustrated in Fig. 36.21. The two optical states of the cells are (i) twisted, and (ii) planar (parallel aligned film), and the director is switched between these states by application of an electric field across electrodes on a single plate of the cell. The state that is stabilised by the electric field depends on the dielectric anisotropy of the liquid-crystal material. The preferred configuration uses materials having a negative dielectric anisotropy, so that the off-state is a planar-aligned liquid-crystal film. Application of an electric field to the in-plane electrodes will cause the director at the bottom surface to align perpendicularly to its initial direction, and so induce a twist through the cell (Fig. 36.21).

From an optical point of view, the director is always in the plane of the cell, and this means there is less distortion of an image when viewed at angles other than 90°. Another advantage of the IPS device is that the electric field is confined to the lower plate, and the lines of force do not extend across the cell to the grounded upper plate. This means that a very low electrical conductivity of the liquid-crystal material is less important than for conventional TFT-TN displays. The threshold voltage for the IPS-mode device is given by

$$V_{\mathrm{th}}^{\mathrm{IPS}} = \frac{\pi d}{\ell}\sqrt{\frac{K_2}{\varepsilon_0 \Delta\varepsilon}}\,, \tag{36.28}$$

where ℓ is the thickness of the liquid-crystal film, and d is the separation of the in-plane electrodes. Not surprisingly the threshold depends only on the twist elastic constant, which is usually smaller by about a factor of two than the splay and bend elastic constants. While this helps to reduce the operating voltage, the smaller elastic energy associated with the pure twist deformation results in longer switching times. A further disadvantage of the IPS display is that the optical transmission of the cell is reduced by the requirement to have both electrodes deposited on one plate, thereby

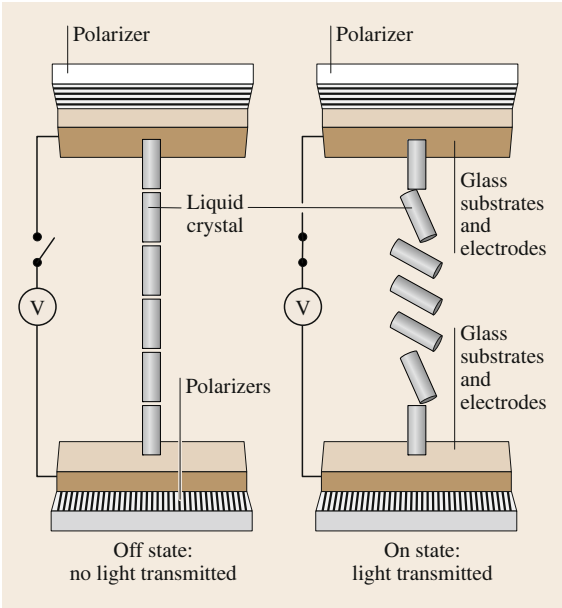


Fig. 36.22 Schematic of a twisted vertically aligned display

making smaller the active area available to display the image.

A configuration which shares some of the characteristics of the IPS cell and has been successfully commercialised is the twisted vertically aligned nematic (TVAN) cell. The two optical states for this configuration are uniform vertical (homeotropic) alignment of the director for the off-state, and a twisted geometry for the on-state (Fig. 36.22). The liquid-crystal material used has a negative dielectric anisotropy, so application of an electric field between the plates causes the director to align perpendicularly with respect to the field direction. A small quantity of optically active material (chiral dopant) is added to the liquid-crystal mixture to ensure that the switched director adopts a twisted configuration through the cell. The advantages of this cell are good viewing-angle characteristics and high optical contrast. Improvements in display technology continue to be made, often simplifying earlier devices. For example, high-quality displays described as vertically aligned nematic (VAN) devices are now available based on the TVAN configuration, but without the twist. The material requirement here is for a liquid crystal of negative dielectric anisotropy, that will align perpendicularly to an applied electric field.

In the devices described above, one state is defined by the surface conditions of the cell, while the other is defined by the action of the applied electric field. A bistable device is one in which two stable field-free states exist, both of which are accessible

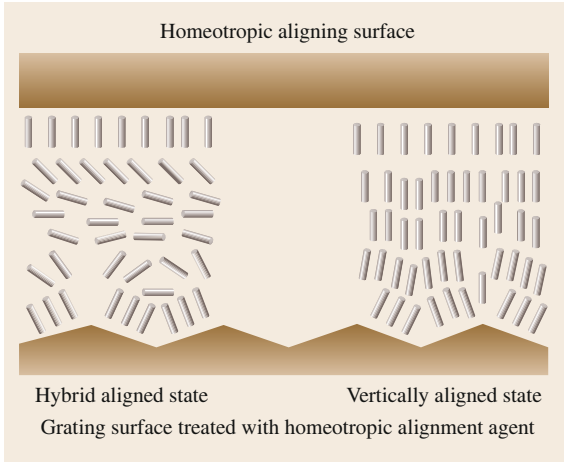


Fig. 36.23 Schematic of a zenithal bistable device

by switching with an external field. The first bistable liquid-crystal display to be developed was based on a ferroelectric effect observed in chiral tilted smectic C liquid crystals [36.48]. This ferroelectric smectic display has achieved some limited commercial success in specialist markets, but relies on a surface stabilisation of smectic layers, which is very sensitive to mechanical shock. Recently [36.49], bistable nematic displays have been developed in which two alignment states within a liquid-crystal cell, having different optical transmission, can be stabilised. If one of the substrates of a normal cell is replaced by a surface which has potentially two states of minimum energy corresponding to two surface alignments of the director, then it becomes possible to switch these states selectively using an electric field. A suitable bistable surface is provided by a grooved surface (grating) which has been treated with a surfactant to favour homeotropic alignment of the director at its surface [36.50]. Thus the two surface states correspond to (i) that determined by the grating, and (ii) that determined by the surfactant where the director is perpendicular to the substrate. Combining this intrinsically bistable substrate with a second substrate having a director alignment direction perpendicular to the grating direction gives a cell configuration capable of supporting two optically distinct stable states, which can be switched between using an applied voltage. Various director configurations are possible with this type of cell, and one example is illustrated in Fig. 36.23.

In the absence of any perturbation, the director orientation within the cell will be determined by the homeotropic alignment at one substrate and the alignment at the grating substrate. This hybrid (uniform planar and homeotropic) alignment causes a spatially varying director tilt through the sample. Application of

an electric field to a positive-dielectric-anisotropy material will cause the director to align parallel to the applied field, and eventually a fully homeotropic configuration for the director is stabilised. The switch back from the homeotropic state to the hybrid state is thought to be due to a flexoelectric interaction. Other alignment configurations are also possible for the so-called zenithal bistable nematic (ZBD) cell. Displays based on these configurations share the good viewing characteristics of both the IPS and TVAN configurations, but they have the great advantage that the image is retained when the voltage is removed.

All the displays described so far rely on the coupling between an applied electric field and the dielectric properties of the liquid-crystal material, but, as we have shown, other material properties are just as important to the operating characteristics of the display. The appearance of a display depends on the optical properties of the liquid crystal and the cell configuration, but the operating voltage and switching times of a display are crucial in determining the types of application. Changing the nature of the interaction between the switching electric field and the liquid crystal gives rise to another range of possibilities for liquid-crystal devices. Under certain circumstances, a liquid crystal can be made to exhibit permanent ferroelectric (or spontaneous) polarisation, and this couples linearly with an external electric field, in contrast to dielectric properties which couple with the square of the electric field strength. Not surprisingly this makes a big difference to the switching behaviour of liquid crystals.

The final display to be considered is based on flexoelectric coupling between the electric field and the liquid crystal. Flexoelectricity occurs, in principle, with all liquid crystals, chiral or not, and shows itself as a bulk electric polarisation induced by an elastic strain. Conversely application of an electric field can cause an elastic strain. In general, flexoelectricity is rather small and difficult to detect, however it is thought to be responsible for an electro-optic effect observed in chiral nematic liquid crystals, which is being investigated for display applications. The effect, sometimes known as the deformed helix mode [36.51], is similar in some respects to the ferroelectric switching observed in chiral smectic C phases, but there is no longer a requirement for a layered structure. Chiral nematic liquid crystals spontaneously form helical structures in which the director rotates with a pitch determined by the molecular structure. If an electric field is applied perpendicularly to a chiral-nematic helix, then there is a tendency for the helix to unwind, depending on the sign of the dielectric anisotropy. Even if the dielectric anisotropy of the material is zero, there is an elastic strain which can generate a polarisation (flexoelectric polarisation), which

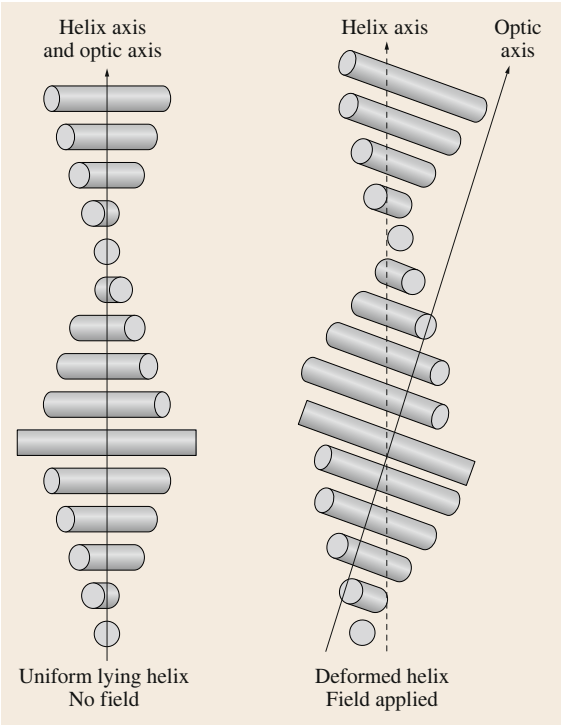


Fig. 36.24 Schematic of a deformed helix mode flexoelectric display

will interact with an applied electric field. This may be exploited in a device configuration, where a thin film of a chiral nematic liquid crystal, having a small or zero dielectric anisotropy, is aligned such that its helix axis is parallel to the containing glass plates (Fig. 36.24).

Application of an electric field across the plates will cause a distortion of the helix through the splay and bend flexoelectric coefficients, which appears as a rotation of the optic axis in the plane of the film [36.25]. Reversal of the electric field direction will reverse the rotation of the optic axis, with an intrinsic switching time about one hundred times faster than conventional nematic displays. Optically, the effect observed is very similar to that exhibited by smectic ferroelectric displays.

There are many cell configurations that can be used with liquid crystals to produce optical switches, displays or light modulators, and some of the more important have been described. The precise operation and performance of these liquid-crystal devices depends on both the cell design and the material properties of the liquid crystal. To a large extent the configuration of the liquid crystal within the cell is determined by such factors as the surface treatment of the plates enclosing the liquid crystal and the interactions between the surfaces and the liquid crystal. Our under-

standing of these interactions is very limited at the present time, and much more research is necessary before a quantitative theory can be formulated. However, given the cell configuration, the performance of the liquid-crystal device depends critically on the physical properties of the liquid-crystal material. Thus the electrical switching characteristics will depend on the dielectric properties, while the optical appearance of the device will be determined by the refractive indices. Elastic properties contribute to both the electric field response and the optical appearance, since any deformation of the director will be determined by the elastic properties of the liquid crystal. Finally, the all-important dynamical behaviour will be controlled by the viscous properties of the liquid crystal. All these material properties will be discussed in the next section.

36.4 Materials for Displays

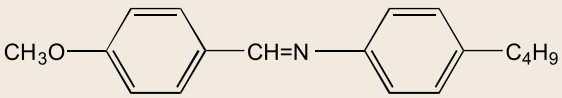
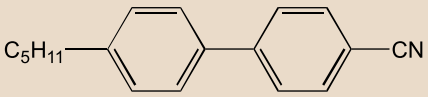
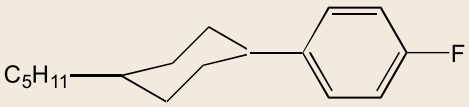
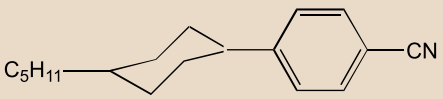
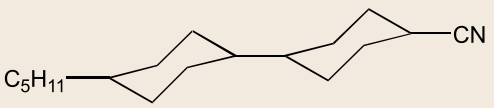
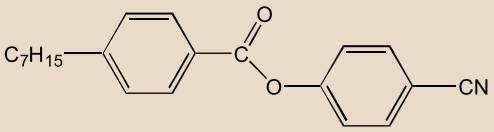
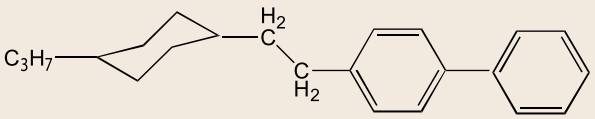
The most important requirement for a liquid-crystal display material is that it should be liquid crystalline over the temperature range of operation of the device. Despite this, some of the first experimental display devices incorporated heaters in order to maintain the material in the liquid-crystal phase: for example, the first liquid-crystal shutter patented by Marconi and early prototype displays developed by the Radio Corporation of America (RCA). It was not until the late 1960s that room-temperature nematic liquid crystals suitable for display applications were discovered. The first of these were based on Schiff's bases, which although easy to prepare, were difficult to purify and were susceptible to chemical decomposition in a device. One material which attracted particular attention from experimental physicists was MBBA, which has a nematic range of 22–47 °C (Table 36.1). The basic two-ring core linked by an imine group was the structural unit of many compounds having different terminal groups which were prepared for display mixtures in the early 1970s. It was found that mixtures of Schiff's bases often had lower crystal-to-nematic transition temperatures than any of the components, and furthermore would often remain liquid crystalline even below the thermodynamic crystallisation temperature. These two phenomena of eutectic behaviour and super-cooling have been exploited in the development of materials for devices. The early experiments on liquid-crystal displays were primarily focused on nematic or chiral compounds, but we have seen that other displays have been developed which use different liquid-crystal phases, most importantly the chiral smectic C phase.

Early studies established guiding principles for the development of display materials. First, the phase behaviour must be acceptable, i. e., the right phase stable over a suitable temperature range. Secondly, the material must have the correct electrical and optical properties for the particular display application envisaged, and above all must be of sufficient chemical purity to prevent any deterioration in performance over time. Again, guided by the early experiments, suitable display materials require the synthesis of compounds of appropriate chemical structure, and then the formulation of mixtures to optimise the properties. There have been a number of reviews of liquid-crystal materials for displays [36.53, 54, 56–59] and these contain many tables of data on a wide range of compounds. In this Section, we will give a brief account of the basic chemical structures used for materials in modern liquid-crystal displays, and then show how mixtures are devised to give the best possible performance characteristics for different displays. It has to be recognised that many of the details of display materials are matters of commercial confidentiality, and so it is not possible to give precise accounts of materials currently used or under investigation. However, the generic chemical structures and principles used in developing suitable mixtures are generally applicable.

36.4.1 Chemical Structure and Liquid-Crystal Phase Behaviour

There is a huge literature on the relationship between the structure of mesogens and the nature and stability of the liquid-crystal phases they form [36.60]. The studies have embraced empirical correlations of chemical structure and phase behaviour, theoretical calculations for simple particles (hard rods, spherocylinders etc.) representing mesogens, and computer simulations of collections of particles of varying complexities. For the display applications considered in this Chapter, the desired phases are nematic, and occasionally chiral nematic or chiral smectic C. Such phases are formed by molecules having extended structures, which usually require the presence of terminal alkyl chains to reduce the crystallisation temperatures. Components in nematic display mixtures typically have two, three or four carbocyclic rings joined directly or through a variety of linking groups.

Table 36.1 Some typical liquid-crystal materials, including selected physical properties ((ex) – extrapolated from measurements on a nematic solution)

Compound Transition temperatures (°C) [reference]	$\Delta\varepsilon$ (T °C)	Δn (T °C)	Rotational viscosity (γ_1/m Pa s)
 T_{CrN} 22 °C T_{NI} 47 °C [36.53]	≈ 1	0.2	$\gamma_1 = 109$ [36.52] (37 °C)
 T_{CrN} 22 °C T_{NI} 35 °C [36.53]	8.5 (29 °C)	0.18 (25 °C)	$\gamma_1 = 102$ [36.52] (25 °C)
 T_{CrI} 34 °C [36.54]	3.2 (ex)	0.05 (ex)	
 T_{CrN} 31 °C T_{NI} 55 °C [36.53]	9.9 (48 °C)	0.12 (40 °C)	$\gamma_1 = 128$ [36.52]
 T_{CrN} 62 °C T_{NI} 85 °C [36.53]	3.5 (78 °C)	0.05	
 T_{CrN} 44 °C T_{NI} 57 °C [36.53]	19.9 (40 °C)	0.15	
 T_{CrN} 67 °C T_{NI} 82 °C [36.55]	≈ 0	0.15	

36.4.2 The Formulation of Liquid-Crystal Display Mixtures

The two requirements for a liquid crystal to be used in a display are a suitable temperature range of phase stability and appropriate physical properties. These requirements cannot be satisfied for complex displays by a single compound, and commercial display materials may contain up to twenty different components. The formulation of these mixtures is essentially an empirical process, but guided by the results of thermodynamics and experience. The principles behind the preparation of multicomponent mixtures can be illustrated initially by consideration of a binary mixture.

It is well-known that the melting point of a binary mixture of miscible compounds is depressed, sometimes below the melting points of both components. Furthermore, the melting point of the binary mixture may exhibit a minimum at a particular composition, known as the eutectic. This occurs with liquid-crystalline compounds, and provides a method of reducing the lower temperature limit for liquid-crystal phase stability in mixtures. The upper temperature limit of the liquid-crystal range is fixed by the transition to an isotropic liquid. The phase rule of Willard Gibbs

predicts that in binary mixtures there will always be a region of two-phase coexistence in the vicinity of a phase transition; that is, the transition from liquid crystal to isotropic occurs over a range of temperatures for which both the isotropic liquid and liquid crystal are stable in the mixture. Because of the weak first-order nature of most liquid crystal to isotropic phase transitions, the two-phase region is small. The character of phase transitions is determined by the corresponding entropy change, and a weak first-order transition has a small $\left(\approx 2\,\mathrm{J\,K^{-1}\,mol^{-1}}\right)$ associated entropy. If the latter were zero, then the transition would be second order, and there would no longer be a region of two-phase coexistence. The phase diagrams of multicomponent nematic mixtures can be calculated by thermodynamic methods [36.61, 62] and the transition temperatures of the mixtures can vary with composition in a variety of ways. For mixtures of two liquid-crystalline compounds of similar chemical constitution, the variation of the nematic to isotropic transition temperature is approximately linear with composition [36.63].

It is possible to calculate the variation of the melting point with composition using an equation attributed to Schroeder and van Laar. For each component i, the mixture composition (mole fraction x_i) and the melting point of the mixture T are related by

$$\ln x_i = -\frac{\Delta H_i}{R}\left(\frac{1}{T} - \frac{1}{T_i}\right), \tag{36.29}$$

where ΔH_i and T_i are, respectively, the latent heat of fusion and melting point of the pure component i. For a binary mixture there are two such equations which can be solved to give the eutectic temperature and composition. In a multicomponent mixture, the set of equations (36.29) can be solved subject to the condition,

$$\sum_i x_i = 1 \tag{36.30}$$

to predict the eutectic of the mixture.

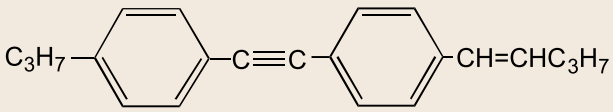
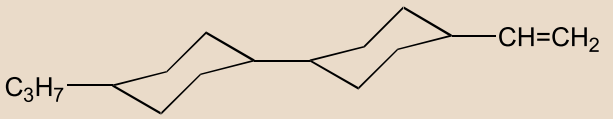
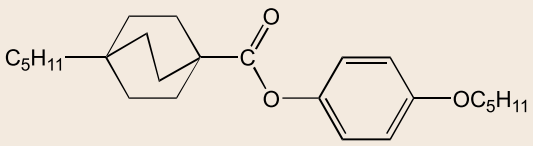
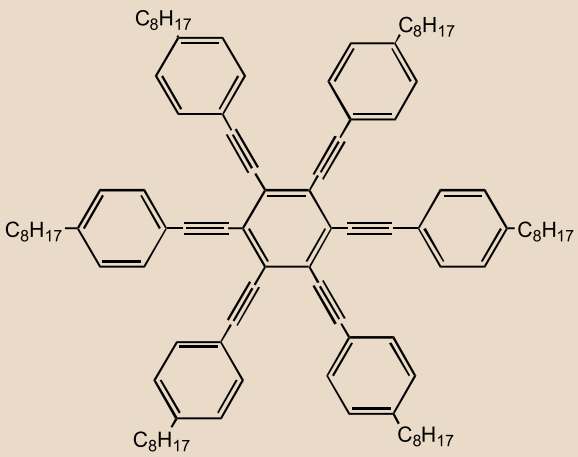
While there is a reasonable thermodynamic basis to the prediction of the phase diagrams of mixtures, the determination of the physical properties of mixtures from the properties of individual components is much more difficult. Given the absence of any better theories, it is common to assume that in mixtures, physical properties such as dielectric anisotropy, birefringence and even viscosity vary linearly with the amount of any component, at least for small concentrations. While this may give an indication of the effect of different components on the properties of a display mixture, it can also be very misleading. One theoretical problem is that, for a mixture at a particular temperature, the orientational order parameters of the different components are not equal. The more anisometric components (e.g., three-ring mesogens) are likely to have a larger orientational order parameter than smaller (two-ring) mesogens. Since the various physical properties of interest in displays depend on the order parameter in different ways, it is difficult to predict the contribution of different components to the overall mixture properties. Despite this, many tables of data for liquid-crystal compounds of interest for display mixtures are prepared [36.60] on the basis of extrapolated measurements on mixtures at low composition, normally $< 20\,\mathrm{wt\%}$. There is always a problem concerning the temperatures at which to compare the physical properties of liquid crystals and their components. Many measurements are made at room temperature, so that this becomes the temperature for comparison. However, a more useful approach is to compare properties at equal reduced temperatures (or at the same shifted temperatures, $T_{\mathrm{NI}} - T$), since under these conditions the orientational order parameters are likely to be similar.

36.4.3 Relationships Between Physical Properties and Chemical Structures of Mesogens

Electrical and Optical Properties

These properties include the dielectric permittivity, electrical conductivity and refractive indices. The magnitude of the dielectric anisotropy determines the threshold voltage necessary to switch a display, and influences the transmission/voltage characteristics of the cell. Depending on the particular display configuration, a positive or negative dielectric anisotropy may be required. Refractive indices strongly affect the appearance of a display. Usually the refractive indices or birefringence must be adjusted for a particular cell configuration to give the optimum on/off contrast ratio. Coloration in displays can sometimes occur in materials of high refractive index, and so it is desirable to keep the birefringence as low as possible, compatible with an acceptable contrast ratio. For twisted structures, the magnitude of the birefringence also determines the efficiency of light guiding, and so close control of the values of the principal refractive indices of a display mixture is important. For non-conducting materials, the refractive indices are measures of the dielectric response of a material at very high, i. e., optical frequencies, and it is possible to formulate a single theory which relates the dielectric and optical properties of a liquid crystal to its molecular properties. Unfortunately this is

Table 36.2 Materials of high, low and negative birefringence

Compound	Transition temperatures	Δn [reference] (T °C)
	T_{NI} 112 °C	0.31 [36.64] (92 °C)
	T_{CrSmB} 23 °C T_{SmBN} 35 °C T_{NI} 49 °C	0.052 [36.59]
	T_{NI} 93.5 °C	0.074 [36.64] (58.5 °C)
	T_{CrN} 80 °C T_{NI} 96 °C	−0.193 [36.64] (61 °C)

not possible for the electrical conductivity. The latter is largely determined by the purity of the liquid crystal, but it is found that the higher the value of the permittivity, the larger the electrical conductivity. Materials of high electrical conductivity tend to leak charge, and so an image may deteriorate during a multiplexing cycle. In general it is desirable to minimise the conductivity of a display mixture, although this was not the case for the first liquid-crystal displays reported [36.65]. These utilised the strong light scattering which results when an electric field is applied to certain nematic materials. The scattering is due to electrohydrodynamic instabilities in liquid-crystal materials which have a significant electrical conductivity. Such materials are not suitable for use in modern, fast-multiplexed displays.

The dielectric anisotropy $\Delta\varepsilon$ and birefringence Δn of a nematic can be related to molecular properties of polarisability and dipole moment using a theory originally developed by Maier and Meier [36.66]. The birefringence is given by

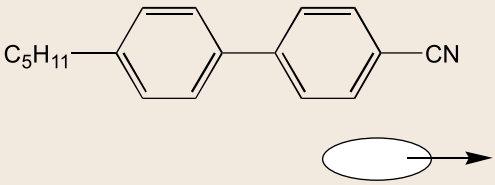
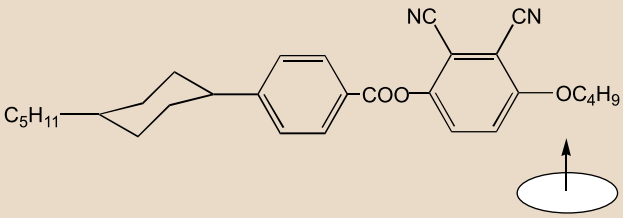
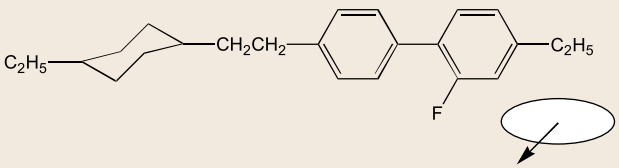
$$\Delta n \approx \frac{NS}{\varepsilon_0}(\alpha_l - \alpha_t)\,, \tag{36.31}$$

where N is the density in molecules per m^3 and $\Delta\alpha = (\alpha_l - \alpha_t)$ is the anisotropy of the molecular polarisability. S is the order parameter, defined in Sect. 36.2.1, and small corrections due to the local field anisotropy have been neglected. Such corrections cannot be ignored in the corresponding expression

$$\Delta\varepsilon = \frac{NhFS}{\varepsilon_0}\left[\Delta\alpha + \frac{\mu^2}{2k_B T}(3\cos^2\beta - 1)\right] \tag{36.32}$$

for the dielectric anisotropy, especially for materials of high permittivity. In (36.32) h and F are local-field correction factors, while μ is the molecular dipole moment, and β is the angle between the dipole direction and the long axis of the molecule. For molecules containing a number of dipolar groups, μ is the root mean

Table 36.3 Materials of high, low and negative dielectric anisotropy; the *inset* figure indicates the direction of the total dipole moment with respect to the core of the molecule ($1\,\mathrm{D} = 3.33564 \times 10^{-30}\,\mathrm{C\,m}$)

Compound	Transition temperatures	$\Delta\varepsilon$ [reference] (T °C)	Total dipole moment μ (D)
	T_{CrN} 22.5 °C T_{NI} 35 °C	11.5 [36.67] (25 °C)	4.8
	T_{CrN} 143 °C T_{NI} 150 °C	−10.0 [36.68] (145 °C)	6.4
	T_{CrN} 13 °C T_{NI} 64 °C	0.0 [36.69] (58 °C)	1.4

square of the vector sum of all contributing groups. Both the birefringence and the dielectric anisotropy increase with decreasing temperature, and the detailed variation with temperature is largely determined by the temperature dependence of the order parameter S.

The manipulation of birefringence is achieved by changing the chemical constitution of the mesogen. Thus extending the electronic conjugation along the axis of a mesogen will result in an increase in longitudinal polarisability, and hence an increase in birefringence. Saturated carbocyclic rings, such as cyclohexane, and aliphatic chains generally have small polarisabilities and mesogens containing a predominance of such moieties form low-birefringence liquid crystals. Mixtures for displays require a positive birefringence, which is associated with calamitic or rod-like mesogens. In order to improve the viewing characteristics of displays, optically retarding films are placed in front of the display, and depending on their function, these may be of negative or positive birefringence. The latter can be fabricated by encapsulation or polymerisation of suitable molecules of an extended structure. On the other hand, films of negative birefringence have been made using discotic materials, i. e., mesogenic molecules formed from disc-like structures which have a negative polarisability anisotropy (Sect. 36.1.3). Some examples of liquid crystals having different birefringence are shown in Table 36.2.

The introduction of chirality into liquid crystals has important consequences for their optical properties. The selective reflection of coloured light from the helical structure of a chiral nematic has already been mentioned in Sect. 36.1.2. All chiral materials will rotate the plane of incident polarised light, and the particular optical properties associated with chiral thin films are exploited in many liquid-crystal device applications.

The dipole moment of a molecule is increased if strongly electronegative or electropositive groups are substituted into the structure, with the result that the dielectric permittivity increases. For mesogenic molecules, the locations of the electropositive or electronegative groups are important, since not only the magnitude but also the orientation of the molecular dipole strongly influences the dielectric anisotropy. From (36.32) it can be seen that the dipolar contribution to the dielectric anisotropy may be positive or negative depending on the value of the angle β, since for values of β greater than 54.7° the dipolar contribution to the permittivity anisotropy becomes negative. This is illustrated by the mesogens shown in Table 36.3, where different structures can be designed to give large positive, negative or zero dielectric anisotropy.

The most effective substituent for producing materials of high dielectric anisotropy is the cyano group, and mixtures containing cyano-mesogens were the basis for the rapid development of complex displays in the 1980s

and early 1990s. However, these mixtures tended to have relatively high viscosities, which gave rise to slow switching times. Another disadvantage was the high electrical conductivity associated with the high dielectric anisotropies which caused charge leakage during multiplexing, and hence degradation of the image.

As the demands placed on liquid-crystal materials by more sophisticated display technologies have increased, new families of molecules have been synthesised and screened for their physical properties. However, it is no longer the properties of the pure mesogens that are of interest, rather how they behave in mixtures. For this reason, the physical properties of most components of display mixtures are measured in mixtures, and values for the pure mesogens are obtained by extrapolation. Data derived in this way are useful in designing display mixtures and for comparison purposes, but cannot be relied upon to give quantitative information about the relationship between molecular structure and physical properties.

The major display technologies using TN, STN or TFT-TN configurations require display mixtures having a positive dielectric anisotropy. Many materials have been developed to improve the electro-optical behaviour of these displays, particularly using fluorine-substituted mesogens to provide the required dielectric and optical properties (for examples see [36.54, 58, 70]). Some of these fluorinated mesogens are shown in Table 36.4. However, within the past seven years, new display configurations have emerged, such as the in-plane switching (IPS) and vertically aligned (VAN and TVAN) nematic modes, which require mixtures with negative dielectric anisotropy. Using the design strategy illustrated above for simple mesogens, it has been possible to prepare a large number of materials with the desired negative dielectric anisotropy. These are again mostly based on fluorine-substituted compounds, and as before their properties have mostly been determined by extrapolation of measurements on mixtures.

Elastic Properties

The property known as elasticity is characteristic of liquid crystals, and distinguishes them from isotropic liquids. It has been shown in Sect. 36.2.3 that the macroscopic orientational disorder of the director in liquid crystals can be represented in terms of three normal modes, designated as splay, twist and bend, and associated with each of these deformations is an elastic constant. Since the elastic properties of display materials contribute to the electro-optic response, their optimisation for particular display configurations is important to maximise the performance of commercial devices. However, despite their importance, the relationships between the magnitudes of elastic constants and the chemical structure of mesogens are poorly understood. There is a good reason for this; the optical and dielectric properties are to a first approximation single particle properties. That is they are roughly proportional to the molecular number density and are also linearly dependent on the order parameter. Because elastic properties are a measure of the change in energy due to displacements of the director, they are related to the orientation-dependent intermolecular forces. Thus, at a molecular level, elastic properties are two-particle properties, and are no longer linearly proportional to the number density. A further consequence is that the elastic properties depend to lowest order on the square of the order parameters. Molecular theories of elasticity in nematic liquid crystals have been developed [36.71] and the simplest results suggest that the different elastic constants can be related to molecular shape

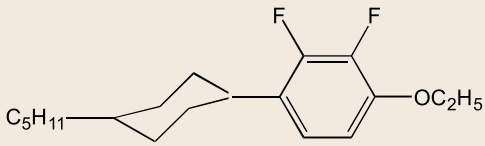
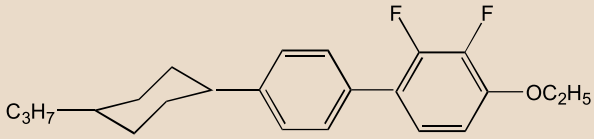
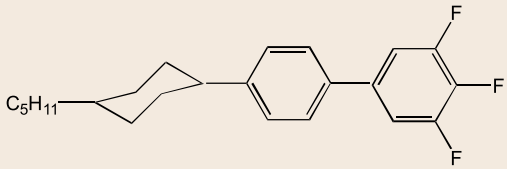
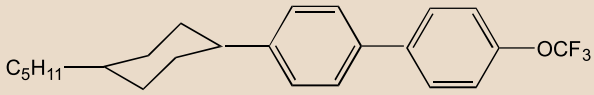
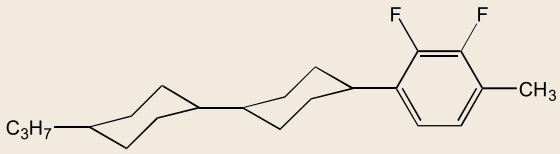
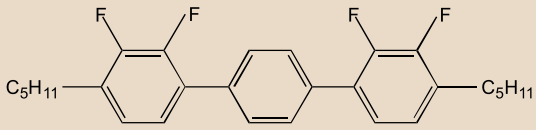
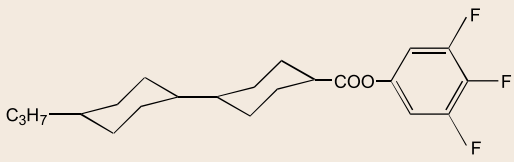
$$K_1 = K_2 \propto \langle x^2 \rangle \quad \text{and} \quad K_3 \propto \langle z^2 \rangle \,, \tag{36.33}$$

where $\langle z^2 \rangle$ and $\langle x^2 \rangle$ are the average intermolecular distances parallel and perpendicular to the molecular alignment direction, respectively. Thus theory predicts that for rod-like molecules the splay elastic constant should be smaller than the bend elastic constant, and increasing the molecular length should increase K_3, while increasing the molecular width should increase K_1. This is roughly in accord with experimental results, except that the prediction of equal splay and twist elastic constants is not confirmed (Fig. 36.9). In general, the twist elastic constant is about one half of the splay elastic constant. Hard particle theories [36.72] evaluated for spherocylinders provide further guidance on the relationship of elastic constants to molecular shape. These theoretical results can be presented in a simplified way as follows

$$\begin{aligned} \frac{K_1 - K}{K} &= \Delta(1 - 3\sigma) \,; \\ \frac{K_2 - K}{K} &= -\Delta(2 + \sigma) \,; \\ \frac{K_3 - K}{K} &= \Delta(1 + 4\sigma) \,, \end{aligned} \tag{36.34}$$

where $K = \frac{1}{3}(K_1 + K_2 + K_3)$. The quantities Δ and σ are parameters of the theory, where Δ is approximately equal to the square of the length:width ratio of the spherocylinder, and σ depends on the degree of orientational order. Despite the fact that details of internal chemical structure are ignored, these theoretical results for nematics are in approximate agreement with experimental measurements on simple nematics. If the nematic material has an underlying smectic phase, or if there is a tendency for local smectic-like ordering,

Table 36.4 Fluorinated mesogens of positive and negative dielectric anisotropy used in liquid-crystal mixtures for modern displays. All measurements listed have been obtained by extrapolation from measurements on nematic solutions

Compound Transition temperatures (°C) [reference]	$\Delta\varepsilon$	Δn	Rotational viscosity (γ_1/m Pa s)
 T_{CrN} 49 °C T_{NI} 13 °C [36.59]	−6.2	0.1	$\gamma_1 = 110$
 T_{CrN} 80 °C T_{NI} 173 °C [36.59]	−5.9	0.156	$\gamma_1 = 233$
 T_{CrN} 30 °C T_{NI} 58 °C [36.54]	11.3	0.134	$\gamma_1 = 191$ [36.73]
 T_{CrSmB} 43 °C T_{SmBN} 128 °C T_{NI} 147 °C [36.54]	9	0.14	$\gamma_1 = 180$ [36.73]
 T_{CrN} 67 °C T_{NI} 145 °C [36.59]	−2.7	0.095	$\gamma_1 = 218$
 T_{CrN} 88 °C T_{NI} 89 °C [36.59]	−4.3	0.2	$\gamma_1 = 210$
 T_{CrN} 56 °C T_{NI} 117 °C [36.59]	11.1	0.067	$\gamma_1 = 175$

this can strongly affect the elastic constants. Both the twist and bend elastic constants are infinite in a smectic phase, and in a nematic phase they diverge as the transition to a smectic phase is approached.

The elastic constants contribute directly to the threshold voltage and the response times of displays. Threshold voltages increase with increasing elastic constants, and the elastic constants responsible depend on the configuration of the display. Thus for the planar-aligned electrically controlled birefringence display (ECB), the threshold voltage depends on K_1, while the switching voltage for TN displays depends on a combination of all three elastic constants (36.27). The IPS display voltage depends only on K_2, and for VAN and

TVAN devices, the threshold voltage is determined by K_3. Different combinations of elastic constants determine the transmission/voltage curves, which are important in the multiplexing of complex displays. For example, decreasing the ratio K_3/K_1 increases the steepness of the curve for TN displays, and so increases the number of lines that may be addressed. On the other hand, for the STN display, if the ratio K_3/K_1 is decreased, the number of lines that may be addressed also decreases. Identification of the important elastic constants necessary to optimise the operation of these displays is relatively straightforward; however, manipulation of the components of displays mixtures to give the best results is much more difficult.

Ferroelectric and Flexoelectric Properties

The electro-optic properties considered so far result from interaction of an electric field with the anisotropic permittivity of a material. This might be termed a quadratic response since the dielectric term in the free energy (36.2) is quadratic in the electric field, and as a consequence the electro-optic response does not depend on the sign of the electric field. For materials having a centre of symmetry, such as achiral nematic and smectic liquid crystals, this response is the only one possible. However, if the centro-symmetry of the liquid crystal is broken in some way, then a linear electric polarisation becomes possible, which results in a linear response to an applied electric field. One example of this, in the context of displays, is the chiral smectic C phase, which in the surface-stabilised state exhibits ferroelectricity, i. e., a spontaneous electric polarisation. The origin of the symmetry breaking in this case is the chirality of the material, and the polarisation is directed along an axis perpendicular to the tilt plane of the smectic C. Another way in which the symmetry can be broken is through elastic strain. This effect was first described by *Meyer* [36.24], and it can be represented as a polarisation resulting from a splay or bend deformation (Sect. 36.2.4). Since at a molecular level, strain is related to molecular interactions, the flexoelectric response depends on a coupling of intermolecular forces and the molecular charge distribution. Two molecular mechanisms have been identified which contribute to the strain-induced polarisation. If the molecules have a net dipole moment, then the longitudinal component can couple with the molecular shape to give a splay polarisation along the director axis, while the transverse component couples with the shape to give a bend polarisation perpendicular to the director axis. Even in the absence of a net dipole, a quadrupolar charge distribution in a molecule can result in strain-induced polarisation [36.26]. Both contribute to the splay and bend flexoelectric coefficients, but only the dipolar part

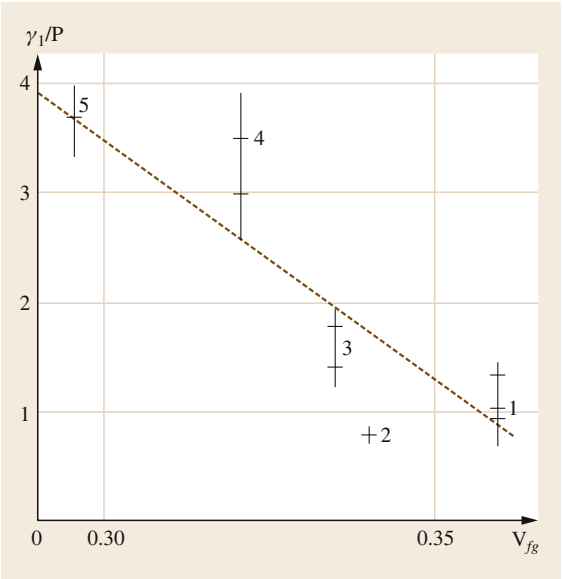


Fig. 36.25 Relationship between the rotational viscosity coefficient γ_1 ($P = 10^{-1}$ Pa s) and the geometric free volume (V_{fg}) at 25 °C for bicyclic polar mesogens. Compound numbers represent alkyl series as follows: 1-cyanophenylalkylcyclohexanes; 2-alkylcyanobiphenyls; 3-cyanophenylalkylpyridines; 4-cyanophenylalkylbicyclooctanes; 5-alkoxycyanobiphenyls. (After [36.52])

persists in the sum $e_1 + e_3$. Thus it is common to quote flexoelectric coefficients as a sum and difference rather than as separate coefficients.

The measurement of flexoelectric coefficients has been a challenge to experimentalists, and there is a wide range of values in the literature for standard materials (Sect. 36.2.4). It is, therefore, premature to draw any conclusions about structure/property relationships for flexoelectricity from the limited experimental data available. There have been attempts [36.74, 75] to model flexoelectricity for collections of Gay–Berne particles simulating wedge-shaped molecules. Application of the surface interaction model to flexoelectric behaviour [36.76] has allowed the calculation of flexoelectric coefficients for a number of molecules; these calculations include the quadrupolar contribution. The importance of molecular shape is clearly demonstrated, and in particular changes of shape, either through conformational changes or *cis–trans* isomerisation, have large effects on the magnitude and sign of the flexoelectric coefficients.

Flexoelectric effects contribute to the electro-optic response of nematic displays, especially those with hybrid alignment, i. e., planar on one electrode and homeotropic on the other electrode, but they are not usually considered in the optimisation of mixture prop-

erties. However, flexoelectric properties are of direct importance to the operation of displays based on the switching of the direction of the optic axis in chiral nematics: the so-called deformed helix mode [36.25].

Viscous Properties

As explained in Sect. 36.2.5, the flow properties of liquid crystals are complicated. Since the materials are anisotropic, the viscosities in different directions are different, furthermore because of the torsional elasticity, viscous stress can couple with the director orientation to produce complex flow patterns. Thus there are five viscosity coefficients necessary for nematics, in addition to the elastic constants, and as many as 20 viscosities for smectic C liquid crystals [36.77]. To relate all or indeed any of these to molecular structure is a formidable challenge. However, for most liquid-crystal displays, the only viscosity of interest is that which relates to the reorientation of the director: the so-called rotational viscosity. This depends on the temperature and order parameter, and on the forces experienced by the rotating director. The rotational viscosities for all liquid-crystalline materials can be represented by one or other of the following parameterised relations

$$\gamma_1 = aS^x \exp\left(\frac{A(T)}{k_B T}\right)$$

or

$$\gamma_1 = bS^y \exp\left(\frac{B}{T - T_0}\right), \qquad (36.35)$$

where a, b, A and B are material parameters, and S is the order parameter raised to a power of x or y between 0 and 2. The first of these expressions emphasises the diffusional nature of rotational relaxation in a liquid crystal, that is molecules rotating in an external potential. The second expression taken from polymer physics describes rotation in terms of free volume, where T_0 is the temperature at which the free volume becomes zero, and the rotational viscosity infinite, i. e., a glass transition.

At the simplest level, the rotational viscosity depends on the molecular shape and size. As the length of the molecule increases, from two rings to three rings etc. γ_1 increases, similarly it increases with the length of the alkyl chain. Varying the nature of the rings in the mesogenic core can have a dramatic effect on the rotational viscosity, which correlates well with free volume, as shown in Fig. 36.25, and the glass temperature of the material.

There is a correlation between the increasing dielectric anisotropy and increasing rotational viscosity which can be attributed to local dipolar intermolecular forces which impede end-over-end rotation of molecules. Thus mesogens having cyano-groups in a terminal position tend to have relatively high viscosities. Other dipolar groups such as fluorine do not have such a deleterious effect on viscosity as the cyano-group, and so fluorine-containing mesogens are increasingly preferred in the formulation of display mixtures.

As with the determination of other properties, the rotational viscosities of mesogenic components are often determined by extrapolation from measurements on mixtures doped with the component under investigation. Such a method only provides approximate values to compare different components, but in the formulation of mixtures for display applications it is only the rotational viscosity of the final mixture that is important. Rotational viscosities of some mesogens of interest for display mixtures are given in the Tables 36.1 and 36.4 accompanying this section. In many instances, the viscosities, as with other properties, have been determined by extrapolation from measurements on mixtures. The measurement of rotational viscosities is experimentally difficult, and some authors prefer to quote the results of bulk-viscosity measurements in terms of a kinematic viscosity. In fact there is a good correlation between kinematic viscosity and rotational viscosity, and where possible both values have been included in the tables. From the various tables, the effect of increasing the molecular length on the viscosity is clearly seen, as is the effect of replacing an F atom with a CN group. Lateral substitution, which produces materials of negative dielectric anisotropy, tends to increase the rotational viscosity. Despite this the fastest nematic displays now use vertical alignment and materials of negative dielectric anisotropy. The operating characteristics of display mixtures depend on the physical properties of individual components in a very complex manner, and optimisation of mixture properties has to be carried out in a concerted way.

References

36.1 P.J. Collings, M. Hird: *Introduction to Liquid Crystals* (Taylor & Francis, London 1997)

36.2 M.A. Bates, G.R. Luckhurst: Computer simulation studies of anisotropic systems. XXX. The phase behaviour and structure of a Gay–Berne mesogen, J. Chem. Phys. **110**, 7087 (1999)

36.3 M. Hird: In: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda,

G.R. Luckhurst (INSPEC, London 2001) Chap. 1.1
36.4 P. Pasini, C. Zannoni: *Advances in the Simulation of Liquid Crystals* (Kluwer, Dordrecht 1998)
36.5 J.W. Emsley, G.R. Luckhurst, G.H. Shilstone, I. Sage: The preparation and properties of the α,ω-bis(4,4′-cyanobiphenyloxy)alkanes: nematogenic molecules with a flexible core, Mol. Cryst. Liq. Cryst. Lett. **110**, 223 (1984)
36.6 G.W. Gray: In: *The Molecular Physics of Liquid Crystals*, ed. by G.R. Luckhurst, G.W. Gray (Academic, New York 1979) Chap. 12
36.7 T.J. Sluckin, D.A. Dunmur, H. Stegemeyer: *Crystals that Flow* (Taylor Francis, New York 2004)
36.8 R.B. Meyer, L. Liebert, L. Strzelecki, P.J. Keller: Ferroelectric liquid crystals, J. Phys. (Paris) **L36**, 69 (1975)
36.9 S.R. Renn, T.C. Lubensky: Abrikosov dislocation lattice in a model of the cholesteric-to-smectic *A* transition, Phys. Rev. A **38**, 2132 (1988)
36.10 J.W. Goodby, M.A. Waugh, S.M. Stein, E. Chin, R. Pindak, J.S. Patel: Characterization of a new helical smectic liquid crystal, Nature **337**, 449 (1989)
36.11 S. Chandrasekhar, B.K. Sadashiva, K.A. Suresh: Liquid crystals of disc-like molecules, Pramana **9**, 471 (1977)
36.12 K. Praefcke: In: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda, G.R. Luckhurst (INSPEC, London 2001) Chap. 1.2
36.13 G.G. Nair, D.S.S. Rao, K.S. Prasad, S. Chandrasekhar, S. Kumar: Electrooptic and viewing angle characterisitics of a display device employing a discotic nematic liquid crystal, Mol. Cryst. Liq. Cryst. **397**, 545 (2003)
36.14 N. Boden, R. Bissel, J. Clements, B. Movaghar: Discotic liquid crystals: Self-organising molecular wires, Liq. Cryst. Today **6**(1), 1 (1996)
36.15 V. Tsvetkov: Über die Molekülanordnung in der anisotrop-flüssigen Phase, Acta Physicochim. **15**, 132 (1942)
36.16 G.R. Luckhurst: In: *Dynamics and Defects in Liquid Crystals*, ed. by P.E. Cladis, P. Palffy-Muhoray (Gordon Breach, Philadelphia 1998) p. 141
36.17 P.G. de Gennes: *The Physics of Liquid Crystals* (Oxford Univ. Press, Oxford 1974) p. 96
36.18 J. Cognard: Alignment of nematic liquid crystals and their mixtures, Mol. Cryst. Liq. Cryst. **1**, 1 (1982)
36.19 D.W. Berreman: Solid surface shape and the alignment of an adjacent nematic liquid crystal, Phys. Rev. Lett. **28**, 1683 (1972)
36.20 A. Rapini, M. Papoular: Distortion d'une lamelle nématique sous champ magnétique. Conditions d'anchorage aux parois, J. Phys. Colloq. (France) **30**, C4-54 (1969)
36.21 A. Sugimura: In: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda, G.R. Luckhurst (INSPEC, London 2001) Chap. 10.2
36.22 F.C. Frank: On the theory of liquid crystals, Trans. Faraday Soc. **25**, 19 (1958)
36.23 V. Fréedericksz, V. Zolina: Forces causing the orientation of an anisotropic liquid, Trans. Faraday Soc. **29**, 919 (1933)
36.24 R.B. Meyer: Piezoelectric effects in liquid crystals, Phys. Rev. Lett. **22**, 918 (1969)
36.25 P. Rudquist, S.T. Lagerwall: On the flexoelectric effect in nematics, Liq. Cryst. **23**, 503 (1997)
36.26 A.E. Petrov: In: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda, G.R. Luckhurst (INSPEC, London 2001) Chap. 5.5
36.27 J. Prost, J.P. Marcerou: On the microscopic interpretation of flexoelectricity, J. Phys. (Paris) **38**, 315 (1977)
36.28 J.K. Moscicki: In: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda, G.R. Luckhurst (INSPEC, London 2001) Chap. 8.2
36.29 W. Helfrich: Molecular theory of flow alignment of nematic liquid crystals, J. Chem. Phys. **50**, 100 (1969)
36.30 C. Gähwiller: The viscosity coefficients of a room-temperature liquid crystal (MBBA), Mol. Cryst. Liq. Cryst. **20**, 301 (1973)
36.31 H. Kneppe, F. Schneider, N.K. Sharma: Rotational viscosity γ_1 of nematic liquid crystals, J. Chem. Phys. **77**, 3203 (1982)
36.32 F.M. Leslie: In: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda, G.R. Luckhurst (INSPEC, London 2001) Chap. 8.1
36.33 O. Parodi: Stress tensor for a nematic liquid crystal, J. Phys. (Paris) **31**, 581 (1970)
36.34 Y. Björnståhl: Untersuchungen über die anisotropen Flüssigkeiten, Ann. Phys. **56**, 161 (1918)
36.35 B. Levin, N. Levin: Improvements in or relating to light valves, British Patent, 441 274 (1936)
36.36 P. Yeh, C. Gu: *Optics of Liquid Crystal Displays* (Wiley, New York 1999)
36.37 P.M. Alt, P. Pleshko: Scanning limitations of liquid crystal displays, IEEE Trans. Electron Devices **21**, 146 (1974)
36.38 E. Jakeman, E.P. Raynes: Electro-optic response times in liquid crystals, Phys. Lett. A **39**, 69 (1972)
36.39 K. Tarumi, U. Finkenzeller, B. Schuler: Dynamic behaviour of twisted nematic cells, Jpn. J. Appl. Phys. **31**, 2829 (1992)
36.40 M. Schadt, W. Helfrich: Voltage-dependent optical activity of a twisted nematic liquid crystal, Appl. Phys. Lett. **18**, 127 (1971)
36.41 C. Mauguin: Sur les cristaux liquides de Lehmann, Bull. Soc. Fr. Mineral. **34**, 71 (1911)
36.42 C.H. Gooch, H.A. Tarry: The optical properties of twisted nematic liquid crystal structures with twist angles $\leq 90^\circ$, J. Phys. D **8**, 1575 (1975)
36.43 T.J. Scheffer, J. Nehring: A new highly multiplexed liquid crystal display, Appl. Phys. Lett. **45**, 1021 (1984)
36.44 C.M. Waters, E.P. Raynes, V. Brimmell: Design of highly multiplexed liquid crystal dye displays, Mol. Cryst. Liq. Cryst. **123**, 303 (1985)
36.45 A.J. Snell, K.D. Mackenzie, W.E. Spear, P.G. LeComber, A.J. Hughes: Application of amorphous silicon field effect transistors in addressable liquid crystal display panels, Appl. Phys. **24**, 357 (1981)

36.46 G. Baur, R. Kiefer, H. Klausmann, F. Windscheid: In-plane switching: A novel electro-optic effect, Liq. Cryst. Today **5**(3), 13 (1995)

36.47 M. Oh-e, K. Kondo: Electro-optical characteristics and switching behaviour of the in-plane switching mode, Appl. Phys. Lett. **67**, 3895 (1995)

36.48 N.A. Clark, S.T. Lagerwall: Submicrosecond bistable electro-optic switching in liquid crystals, Appl. Phys. Lett. **36**, 899 (1980)

36.49 G.P. Bryan-Brown, C.V. Brown, I.C. Sage, V.C. Hui: Voltage-dependent anchoring of a liquid crystal on a grating surface, Nature **392**, 365 (1998)

36.50 C.V. Brown, L. Parry-Jones, S.J. Elston, S.J. Wilkins: Comparison of theoretical and experimental switching curves for a zenithally bistable nematic liquid crystal device, Mol. Cryst. Liq. Cryst. **410**, 417 (2004)

36.51 J.S. Patel, R.B. Meyer: Flexoelectric electro-optics of a cholesteric liquid crystal, Phys. Rev. Lett. **58**, 1538 (1987)

36.52 V.V. Belyaev: In: *Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda, G.R. Luckhurst (INSPEC, London 2001) Chap. 8.4

36.53 G.W. Gray, S.M. Kelly: Liquid crystals for twisted nematic displays, J. Mater. Chem. **9**, 2037 (1999)

36.54 V.F. Petrov: Liquid crystals for AMLCD and TFT-PDLCD applications, Liq. Cryst. **19**, 729 (1995)

36.55 H. Takatsu, K. Takeuchi, H. Sato: Synthesis and some physical properties of 1-cyclohexyl-2-(4″-halobiphenyl-4′-yl) ethanes, Mol. Cryst. Liq. Cryst. **100**, 345 (1983)

36.56 D. Coates: In: *Liquid Crystals, Applications and Uses*, Vol. 1, ed. by B. Bahadur (World Scientific, Singapore 1990) p. 91

36.57 L. Pohl, U. Finkenzeller: In: *Liquid Crystals, Applications and Uses*, Vol. 1, ed. by B. Bahadur (World Scientific, Singapore 1990) p. 1139

36.58 K. Tarumi, M. Bremer, T. Geelhaar: Recent liquid crystal material development for active matrix displays, Ann. Rev. Mater. Sci. **27**, 423 (1997)

36.59 D. Pauluth, K. Tarumi: Advanced liquid crystals for television, J. Mater. Chem. **14**, 1219 (2004)

36.60 D.A. Dunmur, A. Fukuda, G.R. Luckhurst (Eds.): *Physical Properties of Liquid Crystals: Nematics* (INSPEC, London 2001)

36.61 D.S. Hulme, E.P. Raynes, K.J. Harrison: Eutectic mixtures of nematic 4′-substituted 4-cyanobiphenyls, J. Chem. Soc. Chem. Commun. (1974), doi:10.1039/C39740000098

36.62 D. Demus, C. Fietkau, R. Schubert, H. Kehlen: Calculation and experimental verification of eutectic systems with nematic phases, Mol. Cryst. Liq. Cryst. **25**, 215 (1974)

36.63 R.L. Humphries, P.G. James, G.R. Luckhurst: A molecular field treatment of liquid crystalline mixtures, Symp. Faraday Trans. **5**, 107 (1971)

36.64 D.A. Dunmur: In: *Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda, G.R. Luckhurst (INSPEC, London 2001) Chap. 7.1

36.65 G.H. Heilmeier, L.A. Zanoni, L.A. Barton: Dynamic scattering: A new electro-optic effect in certain classes of nematic liquid crystals, Proc. IEEE **56**, 1162 (1968)

36.66 D.A. Dunmur, K. Toriyama: In: *Physical Properties of Liquid Crystals*, ed. by D. Demus, J. Goodby, G.W. Gray, H.-W. Spiess, V. Vill (Wiley-VCH, Weinheim 1999) p. 129

36.67 D.A. Dunmur, M.R. Manterfield, W.H. Miller, J.K. Dunleavy: The dielectric and optical properties of the homologous series of cyano-alkylbiphenyl liquid crystals, Mol. Cryst. Liq. Cryst. **45**, 127 (1978)

36.68 K. Toriyama, D.A. Dunmur, S.E. Hunt: Transverse dipole association and negative dielectric anisotropy of nematic liquid crystals, Liq. Cryst. **5**, 1001 (1989)

36.69 D.A. Dunmur, D.A. Hitchen, X.-J. Hong: The physical and molecular properties of some nematic fluorobiphenylalkanes, Mol. Cryst. Liq. Cryst. **140**, 303 (1986)

36.70 S. Naemura: In: *Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda, G.R. Luckhurst (INSPEC, London 2001) Chap. 11.2

36.71 H. Gruler: The elastic constants of a nematic liquid crystal, Z. Naturforsch. **30a**, 230 (1975)

36.72 R.G. Priest: Theory of the Frank elastic constants of nematic liquid crystals, Phys. Rev. A **7**, 720 (1973)

36.73 K. Tarumi, M. Heckmeier: In: *Physical Properties of Liquid Crystals: Nematics*, ed. by D.A. Dunmur, A. Fukuda, G.R. Luckhurst (INSPEC, London 2001) Chap. 11.4

36.74 J. Stelzer, R. Beradi, C. Zannoni: Flexoelectric effects in liquid crystals formed by pear-shaped molecules. A computer simulation study, Chem. Phys. Lett. **299**, 9 (1999)

36.75 J.L. Billeter, R.A. Pelcovits: Molecular shape and flexoelectricity, Liq. Cryst. **27**, 1151 (2000)

36.76 A. Ferrarini: Shape model for the molecular interpretation of the flexoelectric effect, Phys. Rev. E **64**, 021710 (2001)

36.77 I.W. Stewart: *The Static and Dynamic Continuum Theory of Liquid Crystals* (Taylor Francis, London 2004)

37. Organic Photoconductors

David S. Weiss, Martin Abkowitz†

This Chapter surveys organic photoreceptor devices used in electrophotography. Included in the discussion are the materials (polymers, pigments, charge-transport molecules, etc.), device architecture, fabrication methods, and device electrical characteristics that are critical to the successful functioning of an electrophotographic device (printer).

The Chapter is organized as follows. A brief discussion of the history of xerography and the contributions of Chester Carlson is followed by operational considerations and critical materials properties. The latter includes dark conductivity, photodischarge–charge transport, and photogeneration. Organic photoreceptor characterizations of dark decay, photosensitivity, and electrical-only cycling are discussed in detail. This is followed by discussions of photoreceptor architecture, coating technologies, substrate, conductive layer, and coated layers which carry out specific functions such as smoothing, charge blocking, charge transport, backing, and surface protection.

37.1 **Chester Carlson and Xerography** 968
37.2 **Operational Considerations and Critical Materials Properties** 970
37.2.1 Dark Conductivity 970
37.2.2 Photodischarge–Charge Transport 971
37.2.3 Photogeneration 977
37.3 **OPC Characterization** 978
37.3.1 Dark Decay 979
37.3.2 Photosensitivity 979
37.3.3 Electrical-Only Cycling 980
37.4 **OPC Architecture and Composition** 980
37.4.1 OPC Architecture 980
37.4.2 Coating Technologies 981
37.4.3 Substrate and Conductive Layer 982
37.4.4 Smoothing Layer and Charge-Blocking Layer 983
37.4.5 Charge-Generation Layer (CGL) 983
37.4.6 Charge-Transport Layer (CTL) 988
37.4.7 Overcoat Layer 989
37.5 **Photoreceptor Fabrication** 990
37.6 **Summary** 990
References 992

Organic photoconductors, devices fabricated from organic photoconductors, and the applications of these devices, are the topics which will be covered in this chapter. The term *organic* is used in the chemical sense to encompass materials with carbon as a major constituent. This includes molecular as well as polymeric materials that are not naturally occurring but are purposefully designed and synthesized in the laboratory for specific physical, chemical, dark-electrical, and photoelectrical characteristics. Organic photoconductors are single, or more commonly, a formulated blend of materials which have photoconductive characteristics. Put simply, these are materials in which the electrical conductivity increases on exposure to light [37.1, 2]. Organic photoreceptors are thin-film devices made from organic photoconductive materials with physical, chemical, dark-electrical, and photoelectrical characteristics designed for optimum performance in specific applications. The most important application is in electrophotography, where organic photoreceptors are utilized in machines for digital printing and copying. In this chapter we will discuss details of organic photoreceptor architecture and their composition and characteristics that have enabled their utilization in electrophotographic printing. In this space it will only be possible to touch briefly on many important aspects of organic photoreceptors but references will be provided for those wishing to delve deeper [37.3–7].

Organic photoconductors are materials and organic photoreceptors are devices made from organic photoconductors. The common acronym for an organic photoreceptor is OPC, which of course stands for organic photoconductor, and simply indicates that acronyms are often inscrutable. The application for which or-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_37

ganic photoreceptors have received the most attention and development effort is as the photosensitive image-creating element in electrophotographic-based printers and copiers. In these machines the photoreceptor is a highly uniform, defect-free, multilayer thin film ($\approx 20-50\,\mu m$ photoconductor thickness) coated on a metal drum or a flexible belt, with an active area from 160 to over $12\,000\,cm^2$. Organic photoreceptors for electrophotography were first introduced in the 1960s and, after years of development for improved physical and electrical characteristics, they form the basis of a multi-billion-dollar industry today.

37.1 Chester Carlson and Xerography

Electrophotography has an interesting history [37.8–11] and we will briefly tell this story with emphasis on the enabling aspects of organic photoreceptors in this technology. The story begins with Chester F. Carlson, a patent lawyer who saw the need for a method of reproducing documents. In his kitchen laboratory in Queens, New York, he demonstrated that a charged photoconductor could be used to produce an electrostatic latent image which could subsequently be visualized by contact with oppositely charged particles. The first electrophotographic image was *10.-22.-38 Astoria*. The patent application was filed on 4th April 1939 and on 6th October 1942, US patent 2 297 691 was issued [37.12]. Figure 37.1 shows the title page of that patent. In concept, electrophotography is a relatively simple process with six (or seven) steps but it took over 20 years to bring it to fruition and another decade to make it truly practical. The essential steps are all illustrated in the US patent cover page: charging, exposure, developing, transfer and fixing. Not pictured are the final steps, blanket exposure and cleaning, to prepare the photoreceptor for the next electrophotographic cycle.

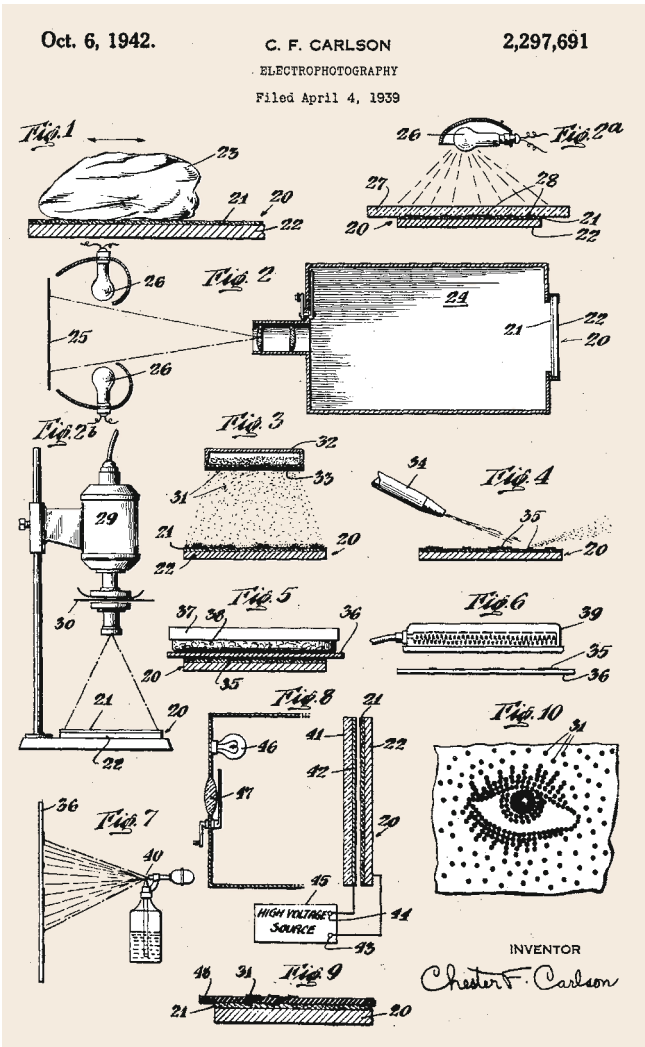


Fig. 37.1 Title page of the first U.S. Patent on electrophotography by C.F. Carlson

Carlson had great difficulty finding commercial backing for his invention. He was unable to convince the management at a long list of influential and affluent companies that making copies of documents was an attractive business proposition. In the end the Battelle Development Corporation took on the project and made significant advances towards making it more practical. In Carlson's invention a coating of photoconductive material such as sulfur or anthracene is prepared on a conductive, grounded, metal substrate. The photoconductor is tribocharged by contact with an appropriate material (silk rabbit's fur, etc.). The photoconductor is then image-wise exposed with light of a wavelength that causes the photoconductive response. The areas struck by light become conductive and the surface charge is neutralized. Thus, there is now a difference in surface charge density, and hence surface potential, between the exposed and unexposed areas. This difference in surface potential, the so-called electrostatic latent image, was visualized by contacting the surface with small marking particles, such as spores from the lycopodium club moss, with surface charge opposite to the original surface charge. The final copy is produced by transferring the *toned* image on the photoreceptor to the receiver (Carlson accomplished this with a heated sheet

of waxed paper). Carlson later dubbed his process *xerography*, meaning dry writing, based on the use of dry powder as the marking particles. Electrophotography is a more general term encompassing, for example, the use of liquid toners in which the charged marking particles are suspended in an insulating fluid.

Two significant early advances were the use of amorphous selenium as the photoconductive material and the development of a practical method (the corotron) for the generation of an air corona for surface charging the photoconductive material. The corotron is simply a fine wire held at high potential (above the air ionization threshold of $\approx 7\,\mathrm{kV}$) and shielded on the back with a grounded metal plate. Xerography was eventually purchased by the Haloid Corporation of Rochester, New York, (later to become the Xerox Corporation) with Carlson as a consultant. The first xerographic product was the XeroX Copier model A introduced in 1949. This initial product had totally manual operation which required several minutes to produce a copy. The photoreceptor was a selenium plate that measured 9×14 inches. From these modest beginnings, advances in engineering the various processes and materials progressed in an evolutionary manner leading to the introduction of the automated copy machine, the Xerox 914 in 1959. This utilized a selenium-coated drum photoreceptor. The evolution continues today and new product introductions have rapidly shifted from copiers (reflection of light from a document to expose the photoreceptor) to digital printers (pixel-wise exposure of the photoreceptor based on data from a previously compiled image file) using drum- or flexible-belt-based organic photoreceptors. Early examples of copiers utilizing organic photoreceptors are the IBM copier 1 (1970) and the Kodak Ektaprint 100 (1975). Initially the photoreceptors were *single-layer* in that the photoconductive materials were all contained in a single layer. For high-speed high-quality applications more complex photoreceptors were introduced. These had multilayer structures with each layer being optimized for a specific function. The single-layer OPC architecture is, on the other hand, what one finds today in desktop laser printers. The choice of single versus multilayer OPC hinges on the quality and printing speed requirements of the end use printer, and on the OPC manufacturing cost. The details of photoreceptor architecture and composition will be discussed later.

There are some subtle differences in the electrophotographic process between copiers and digital printers. Since the electrostatic image in a copier is produced by reflection of the exposing light from a document (black text for example), areas of the document which are dark (text) reflect no light and the corresponding area of the photoreceptor retains its surface charge. To produce a final print in which the dark areas correspond to dark areas in the original document one must *develop* the image with a marking particle (toner) with charge that is opposite to that on the photoreceptor surface. This process is called *charged area development*. In a digital printer the exposure is controlled by an image file and the image is composed of a microscopic halftone pattern of *pixels*. Pixel size is determined by the desired *resolution* such that 600 dpi corresponds to a pixel size of about $40\,\mu\mathrm{m}$. In a digital process it is possible to expose either those areas which will be developed or undeveloped in the final print. The approach sometimes chosen is to expose only those areas which will be toned since most documents are text-based and text occupies only a small fraction of the total document area. This approach decreases the on-time of the exposing system and lengthens its lifetime. Thus, in this case those areas to be developed have been exposed and have decreased surface potential. This process, *discharged area development*, is carried out by using toner with a charge polarity that is the same as that of the photoreceptor surface. Factors in OPC design relating to their optimization for digital imaging have been investigated but in practice the OPCs used in today's copiers and printers differ little [37.13–16]. Discussion of the chemistry, physics, and engineering involved in toners, developers, and development systems is beyond the scope of this chapter [37.17, 18].

In addition to the photoreceptor some of the major subsystems of a modern digital printer are:

- Exposure: Scanning laser or light-emitting diode (LED) arrays.
- Development: Magnetic brush or other technology with black and/or colored marking particles plus magnetic carrier.
- Transfer: Heat, pressure, etc. to remove toner from the photoreceptor and place it on the receiver. In some systems the toner is transferred first to an intermediate drum or web and then to the final receiver.
- Erase: Blanket exposure to return the photoreceptor to a uniform surface-charge state.
- Clean: Blade or brush to remove untransferred toner from the photoreceptor surface.
- Fixing: Melting the toner particles onto the receiver.
- Computers and software: Image capture, rendering, and storage, process control, receiver handling, etc.
- Process control: Software and hardware (sensors, voltmeters, etc.) to maintain image quality.
- Sensors and controls: For receiver handling.

A full listing would be much longer and the modern electrophotographic printer is a highly complex sys-

tem where the hardware, software, and materials have been successfully co-optimized to meet the product aims [37.19–21].

Since the photoreceptor surface is either in physical contact with (development, transfer, cleaning), or exposed to, effluents (fusing, corona charging) from the various subsystems its physical and electrical characteristics must be stable to these interactions throughout its life. Much of the development in OPCs has been to extend the photoreceptor process lifetime. Today the most durable OPCs might be replaced after 100 000 or more imaging cycles.

37.2 Operational Considerations and Critical Materials Properties

The steps involved in the electrophotographic process were detailed previously [37.17, 19]. In this section we discuss some of the underlying physics in the formation of the electrostatic latent image. From the point of view of device physics the OPC is a large-area transducer configured as a belt or drum which converts optical information into a latent charge (primary) image. The photoreceptor material is a high-dielectric-strength insulator that is converted to the electrically conducting state with the application of a field and illumination. Phenomenologically, a high field is applied across the thin-film photoconductor (corona charging) and the film is exposed to radiation absorbed by the material. The photon energy is converted into charge carriers at or near the site of absorption and these drift under the influence of the field. Since the charged photoreceptor is an open-circuit device the surface potential decreases as the charges drift through the material and the photoreceptor is discharged.

In this section we focus on some of the key physics and materials issues governing the field-biased motion of electronic charge through the photoreceptor during light-induced xerographic discharge – the process which leads to latent image formation. Work done in optimizing the design and characterization of photoreceptors in order to service an evolving technology operated in a push–pull relationship with the growth in scientific understanding of photoinduced charge generation, injection, transport and trapping first in amorphous semiconductors and then, in what will concern us here, disordered molecular media.

Because the photoreceptor assembly is a sensitized large-area device, the materials which simultaneously optimize all the required properties have always been glassy coatings. (Polycrystalline media are largely unsuitable for a variety of reasons.) The inorganic materials used especially in light-lens copiers were initially amorphous chalcogenides [37.22, 23] and to a much lesser extent hydrogenated amorphous Si [37.24, 25]. Much of the early work was therefore focused on the physics and chemistry of amorphous semiconductors. In this case photogeneration and the subsequent transport of charge occur in precisely the same medium, and design latitude is clearly restricted by the particular combination of optical and transport characteristics of that single layer. In addition these inorganic films are relatively brittle and therefore unsuitable for applications which require a belt architecture, i. e., one in which the photoreceptor is required to bend around small rollers. On the other hand, polymer-based OPCs are inherently flexible and multilayer architectures are readily fabricated such that each layer can be optimized for a particular function such as photogeneration and charge transport [37.26]. In fact, all OPCs for high-end applications have been developed according to this principle with the photoconductor divided into a thin light-absorbing charge-generation layer (CGL) adjacent to the electrode and a thicker charge-transport layer (CTL) which transports holes. With this photoreceptor configuration the surface is negatively charged prior to exposure. The details of OPC architecture and the materials chosen for specific layer functions will be discussed later.

37.2.1 Dark Conductivity

The electrophotographic imaging process begins by applying a surface charge to the photoreceptor. A parallel-plate-capacitor model is appropriate and the surface potential and surface charge density are related through the capacitance per unit area as, $Q/A = (C/A)V = (\varepsilon\varepsilon_0/L)V$, where Q/A is the surface charge density ($\mathrm{C/cm^2}$), C/A is the capacitance per unit area ($\mathrm{F/cm^2}$), V is the surface potential (V), ε is the dielectric constant, ε_0 is the permittivity of free space, and L is the photoreceptor thickness. For a typical organic photoreceptor ($L = 25\,\mu\mathrm{m}$, $\varepsilon = 3$, and $V = -500\,\mathrm{V}$) C/A is $3.2\times10^{-10}\,\mathrm{F/cm^2}$, corresponding to $10^{12}\,\mathrm{charges/cm^2}$. Assuming the surface is composed of molecules with an area of $100\,\mathrm{Å}^2$ ($10^{-14}\,\mathrm{cm^2/molecule}$) only $\approx 1\%$ of the surface molecules are associated with the surface charge. The applied field (V/L) is $2\times10^5\,\mathrm{V/cm}$.

Electrophotographic imaging is enabled by a difference in surface potential between exposed and unexposed areas when the imaged area enters the development subsystem. Generally a potential difference of

at least 300 V is desired. The materials comprising the bulk of the OPC must be highly insulating. A resistivity of $10^{13}\ \Omega$ cm at a field of 2×10^5 V/cm will give rise to a dark decay rate of nearly 200 V/s. Since development typically occurs at a fraction of a second to one second after charging, the photoconductor resistivity needs to be $> 10^{13}\ \Omega$ cm. Organic polymers such as bisphenol-A polycarbonate meet this requirement and are the major component in OPCs.

There are other sources of dark conductivity such as electrode and/or surface charge injection, bulk charge generation, and/or charge detrapping [37.27–30]. Electrode injection is prevented with the interposition of a charge-injection blocking layer between the electrode and the photoconductive material. Surface injection is prevented by the chemical composition of the surface layer (which is typically a hole-transporting CTL). Detrapping can be a significant source of dark decay in OPCs and must be controlled by a balance between the process and the materials (the charge-generation material in particular).

37.2.2 Photodischarge–Charge Transport

The photoreceptor must retain charge in the dark and also be photosensitive enough to discharge exposed areas to half their initial charge potential when irradiated with (nominally for a mid-volume laser printer engine) 4–10 erg/cm^2. The xerographic gain or quantum efficiency of supply describes the fractional number of surface charges neutralized per absorbed photon. For a dual-layer OPC it is a complicated convolution of the quantum efficiency of generation in the charge-generation layer, CGL, the efficiency of carrier injection from the CGL to the CTL and the transport parameters of the CTL. For this discussion we will for the moment ignore issues of charge generation and focus on transport. In dual-layer OPCs the majority of photodischarge occurs via charge transport through the CTL with the CGL playing a minor role. In the transport lexicon [37.26, 31] there are essentially two parameters that constitute the figures of merit to characterize charge motion through the polymeric CTL. These are (1) the drift mobility μ, the measure of *how fast* the carrier moves per unit applied field, and (2) the normalized carrier range $\mu\tau$ (τ is the free-carrier lifetime against deep trapping), which is *how far* the injected carrier moves per unit field before becoming immobilized in a deep trap. The time for a photoinjected carrier to traverse the CTL is called the transit time. The transit time t_{tr}, and mobility μ are related to the specimen thickness L according to $t_{tr} = L/(\mu E)$. The importance of mobility as a critical parameter in the electrophotographic process can be understood as follows: for an increase in the exposure intensity d(F), the final decrease in surface potential d(V) is proportional to the number of injected carriers and the distance they travel within the CTL. During xerographic discharge, a charge of CV_0 (C is the CTL capacitance and V_0 the initial voltage) traverses the bulk and induces time-dependent variation in the electric field behind the leading edge of the injected carrier front. Thus, as the fastest carriers transit the CTL, the electric field behind them is reduced, and the carriers behind the leading edge transit at a lower field, which in turn makes their velocities lower. Thus during xerographic discharge the transit times of individual photoinjected carriers become dispersed over a wide range, typically about an order of magnitude. For discharge to proceed to completion, even in the complete absence of deep trapping, enough time is required for the slowest carriers in the packet to exit the layer before the photoreceptor reaches the development zone – nominally 0.3–1.0 s after exposure in mid-volume printers. The latter must be allowed for in practice. Thus, carrier mobility in this particular illustration should exceed 10^{-6} cm^2/(V s). For example, consider that a dual-layer photoreceptor with a 25 μm CTL, in which there are no deep traps, is subjected to a light flash intense enough to ultimately induce complete discharge (CV_0 of absorbed photons). Consider further that in this CTL the mobility $\mu = 10^{-6}$ cm^2/(V s) at $E = 10^4$ V/cm and that the device is initially charged to 1000 V. It can be calculated that under these conditions the device will have a residual voltage of 20 V after 0.3 s or 7 V after 1 s. Incomplete discharge, unless compensated for, might result in an inadequate toning potential and a toned density less than desired. Note further that, if the mobility is even lower, the results can become totally unacceptable. For example when the mobility is $\mu = 10^{-7}$ cm^2/(V s) under the conditions just described, the residual voltage a full second after exposure is 60 V, even in a completely trap-free CTL. In light of the foregoing illustration for the trap-free case, which sets the mobility benchmark, the effect of traps in the polymeric CTL must be of paramount concern. More precisely, we are concerned with traps whose release time at ambient temperature discernibly exceeds the period of a complete electrophotographic cycle. In the present context we take the latter as the operational definition of a deep trap. With such traps present it is clearly the case that, after repeated charge–expose cycling, some quantity of image degrading charge would remain immobilized in the bulk for times now exceeding the period of a complete xerographic cycle. If ρ is the density of uniformly trapped space charge in a CTL of thickness L and relative dielectric constant ε then there is an associated residual potential V_R, where $V_R = e\rho L^2/(2\varepsilon\varepsilon_0)$. Here

e is the electronic charge and ε_0 is the free-space permittivity. In a nonpolar dielectric medium ($\varepsilon = 3$) with a layer thickness of 25 μm, as few as 10^{13} electronic charges trapped per cm^3 already give rise to a residual of 19 V. Unless process control utilizing electronic feedback correction can be employed, bulk-trapping-induced space-charge buildup during cycling can result in severe image degradation. Residual potential arising from bulk-trapped space charge is a critical electrophotographic process parameter. A residual potential can be related to the normalized carrier range $\mu\tau$ in the weak-trapping limit from the physically plausible ansatz that the residual potential corresponds to that applied voltage for which the carrier range is about half the specimen thickness L. Thus, $\mu\tau$ is approximately $L^2/(2V_R)$. On this basis the tolerable trap density is defined by the requirement that the $\mu\tau$ product in practical devices should typically exceed $10^{-6}\,cm^2/V$. Even highly purified polymer will typically contain many chemical impurities in the 1–10 ppm range, which the foregoing calculations show is vastly higher than permissible trap levels. However, chemical impurities, even when present at relatively high concentration, can be rendered trap-inactive by employing molecular design principles to guide the overall choice of active materials. Concepts derived from scientific understanding of photogeneration, injection across interfaces, and electronic transport in disordered organic materials, in combination with the unique compositional flexibility characteristic of the organic solid state, were together responsible for making multilayer OPCs the dominant practical receptor technology for electrophotography.

As discussed above, successful electrophotographic imaging requires that charge transport through the CTL occur with little trapping in the time scale of the electrophotographic cycle. Here we discuss how this is accomplished in an *impure* organic-chemical-based CTL. The CTL is a glassy solid solution of a charge-transport-active moiety dispersed in a polymer binder. The charge-transport moiety can be molecular or a polymer component. The molecular solutions are called molecularly doped polymers (MDP). The transport-active component is typically 40–50% weight fraction of the CTL. Thus, *doping* in the present context is decidedly different from what doping refers to in conventional semiconductor physics [37.32]. In the latter case the dopant is typically introduced at very low concentration to control the relative proportion of electrons and holes in the bands of a semiconductor crystal leaving optical properties largely unaffected. In a semiconductor crystal at ambient temperature the dominant field driven transport mechanism is scattering perturbed motion of charge carriers in the bands. Under these circumstances, mobility decreases algebraically with

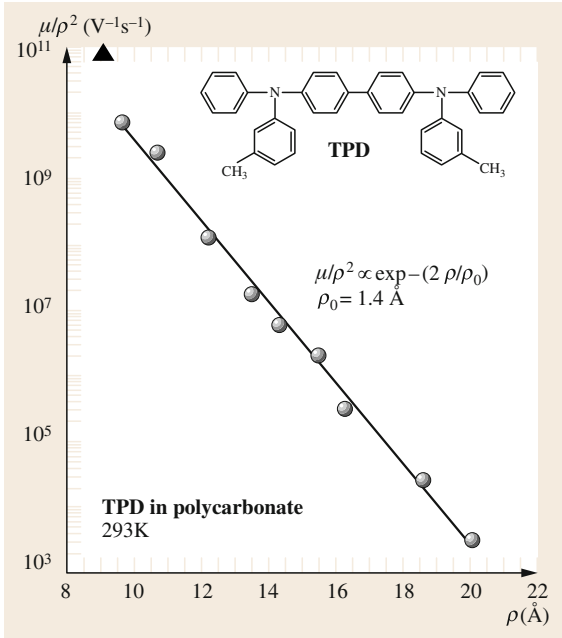


Fig. 37.2 Log (μ/ρ^2) versus average intersite separation ρ of TPD molecules in bisphenol-A polycarbonate ($E = 5 \times 10^4$ V/cm). The *triangle symbol* is the hole drift mobility in pure TPD film

increasing temperature, while carrier population in the bands is thermally activated. The convolution of these processes manifests in a thermally activated conductivity. On the other hand, in MDPs under equilibrium conditions the transport-active molecule is in the neutral state, and the glassy films are perfect insulators, that is, there are no free carriers present in thermal equilibrium [37.33]. These systems can nevertheless support relatively efficient charge transport under the action of an applied field when in contact with a charge reservoir. At fixed temperature the drift mobility of extrinsic carriers decreases exponentially with increasing average intersite separation of the active molecule (ρ) so the drift mobility of MDPs can be tuned over a broad range by simply adjusting concentration [37.34, 35]. This feature of tunability can readily accommodate a wide range of xerographic process speeds as described above and is therefore advantageous from a technological point of view [37.36]. This is illustrated in Fig. 37.2 for a common hole-transport material, N,N′-diphenyl-N,N′-bis(3-methylphenyl)-(1,1′-biphenyl)-4,4′-diamine (TPD, Fig. 37.3). The data is in conformity with a simple tunneling model where ρ_0 is the wavefunction localization radius and the transport states are sited on the TPD molecules. At 40–50 wt% doping approximately 10^4 *hops* are required for the hole to transit a 20 μm photoreceptor.

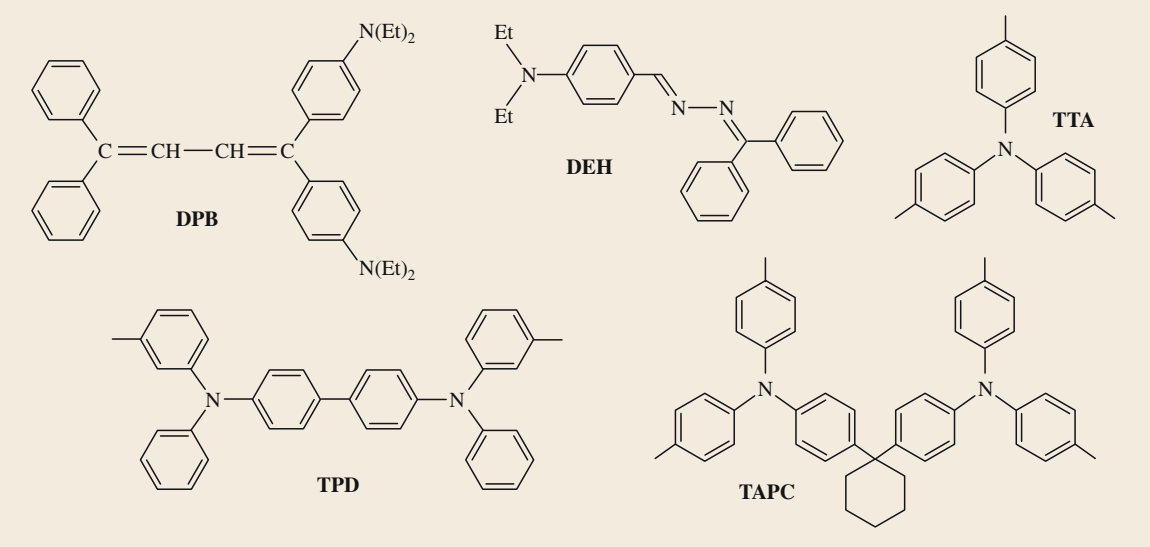


Fig. 37.3 Chemical structures of representative hole transport materials

Drift mobility of holes in CTL films is most conveniently measured by the canonical small-signal time-of-flight (TOF) technique [37.37, 38]. However, for a given composition of the CTL, and all other conditions analogous, the same drift motilities can be inferred by analyzing transport-limited xerographic discharge in bilayer photoreceptors [37.39, 40]. In TOF, the MDP film is prepared with semitransparent blocking contacts and maintained at a bias which is high enough to insure that the transit time of any excess injected carrier is shorter than the bulk dielectric relaxation time. The sample is exposed to a very short and weak pulse of strongly absorbed light (typically 337 nm from a nitrogen laser) incident on the positive electrode. Under these circumstances photoexcited transport molecules are oxidized at the positive electrode to the radical cation (the *hole* in transport terminology). This process sets up a chain of redox steps where electrons are progressively transferred from neutral molecules to their neighboring radical cation. The concentration of the advancing pulse of holes is low, such that the applied field remains uniform during their transit. In the ideal case the current from the advancing pulse of holes is constant until the leading edge reaches the counter-electrode, after which it rapidly decreases. The transit time can be extracted from the transient current and from this the mobility. For a given MDP composition drift mobilities are determined for films of known thickness as a function of applied field and temperature. The drift mobility in MDPs always has a thermally activated temperature dependence; thus the log of the drift mobility scales with inverse temperature, as illustrated in Fig. 37.4, where earlier measurements [37.34] have here [37.41] been extended to encompass the characteristic slope change displayed in the glass-transition region [37.42]. However, in a significant number of cases cited in the literature [37.5] the analogous scaling is with the square of the inverse temperature, (i. e., non-simple activation), as predicted by the disorder

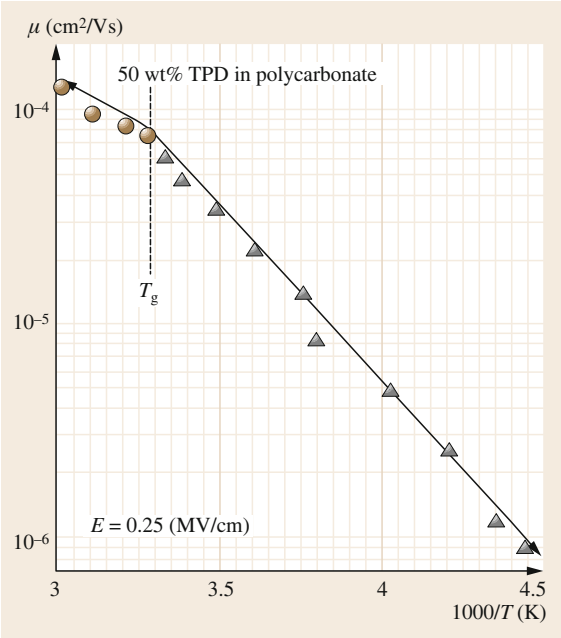


Fig. 37.4 Temperature dependence of the drift mobility in a film of 50 wt% TPD in bisphenol-A polycarbonate ($E = 2.5 \times 10^5$ V/cm). T_g is the glass transition temperature of the film

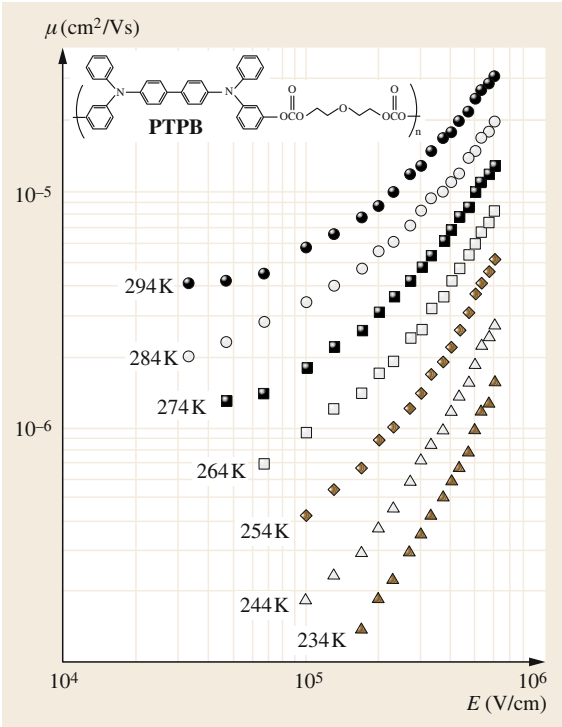


Fig. 37.5 Field dependence of the hole drift mobility in the polymer PTPB parametric in temperature

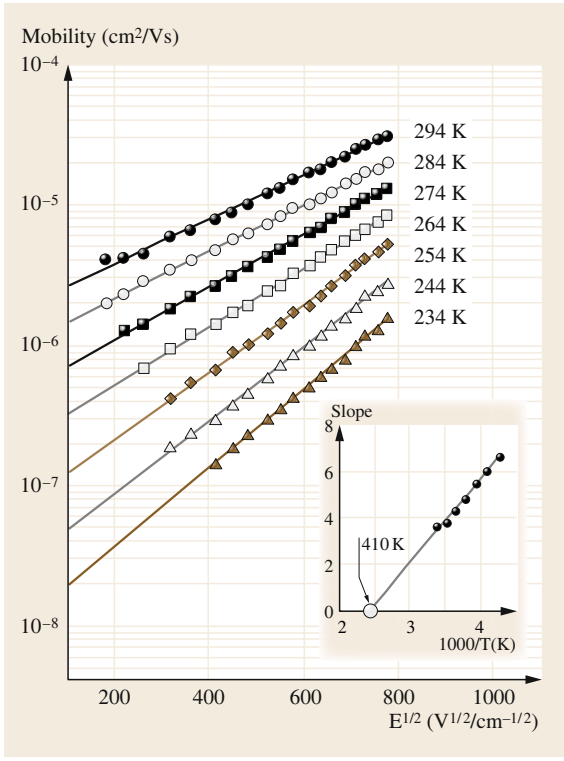


Fig. 37.6 Data set at each temperature in Fig. 37.5 replotted against $E^{1/2}$. The temperature dependence of the slope of each data set is plotted against reciprocal temperature in the *insert*

model [37.43]. The experimentally observed scaling with intersite separation, the temperature dependence, and the high degree of disorder in these systems clearly indicate that electronic transport must involve the field-biased hopping of carriers in an energetically inequivalent manifold of states sited on the transport-active molecules. Apart from their technological importance in electrophotography [37.44] and light-emitting displays [37.45], what has made MDPs a laboratory for the study of hopping transport is that a number of key secondary features in their transport behavior are in fact pervasive in other disordered molecular media [37.46]. There is then the suggestion of a common underlying mechanism susceptible to theoretical treatment. Thus, the field- and temperature-dependent behavior of molecular dispersions is also observed in polymers with transport-active pendant groups such as poly(N-vinylcarbazole) [37.47], in polymers in which small molecule moieties are incorporated in a main chain [37.48] in sigma-conjugated polymers [37.49, 50] and in certain poly π-conjugated systems [37.51, 52]. Figure 37.5 is a plot of the field dependence of the hole drift mobility in the polymer polytetraphenylbenzidine (PTPB) (see insert for the molecular structure) versus field and parametric in temperature over a 60 K range [37.53]. PTPB is a glassy polymer in which tetraphenylbenzidine (TPD) moieties are covalently bonded into a main chain [37.48]. The data is replotted in Fig. 37.6 to show the explicit dependence of the log mobility on the square root of the applied field. The field dependence is itself temperature dependent and clearly becomes stronger with decreasing temperature. The two TPD-based hole-transport media are closely related but qualitatively identical results are reproduced in polysilylenes and polygermylenes which have a sigma-conjugated backbone capable of supporting electron delocalization as inferred from analysis of absorption and emission spectra [37.54, 55]. For example, hole transport in PTPB can be represented in an Arrhenius plot over the temperature range illustrated in Fig. 37.7. In this purely phenomenological description the activation decreases with the square root of the applied field, as displayed in the insert. Precisely the same behavior is exhibited when the same phenomenological description is applied to hole drift mobility data in poly(methylphenylsilylene) (PMPS), as shown in Fig. 37.8. Combining drift mobility [37.56] and spectroscopic data [37.57] has in fact suggested that transport in these sigma-conjugated polymers involves the

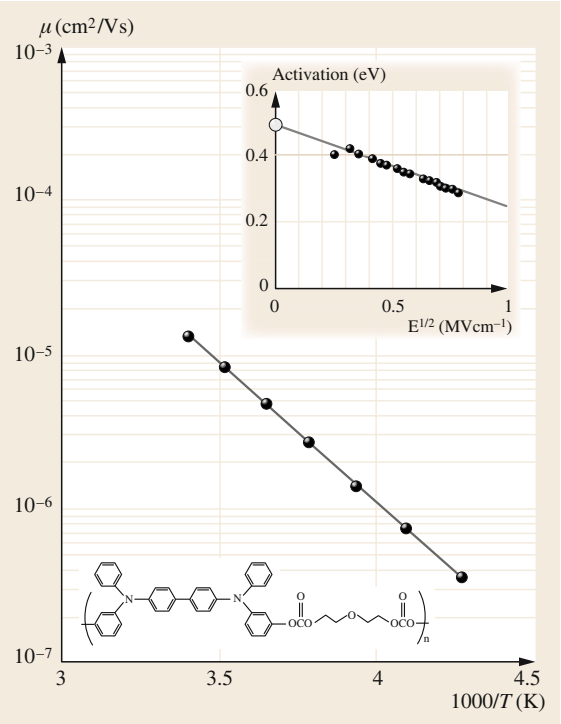


Fig. 37.7 Arrhenius representation of the temperature dependence of hole drift mobility in PTPB. The field dependence of activation is shown in the *insert*

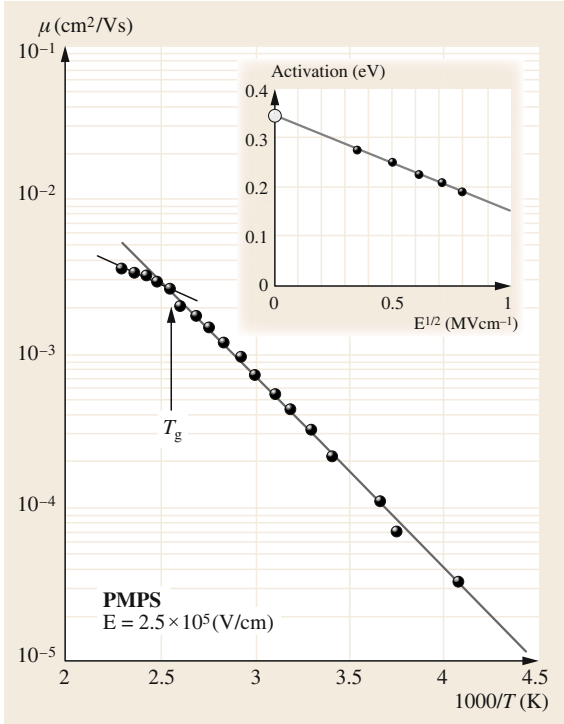


Fig. 37.8 Arrhenius representation of the temperature dependence of hole drift mobility in poly(methylphenylsilylene), PMPS. The field dependence of activation is shown in the insert. T_g is the glass transition temperature of the PMPS film

hopping of holes among chromophore-like main-chain segments of varying length whose function for transport is therefore analogous to the TPD sites in PTPB. This convoluted pattern of behavior first identified in MDPs and therefore clearly characteristic of hopping among discrete energetically in equivalent molecular sites has been reported in TOF experiments on π-conjugated systems notably the phenylene vinylenes. The sites, like those in polysilylenes, are interpreted to be domain-like backbone segments of varying conjugation length [37.51, 52]. These combined results and the recurrent pattern of behavior they reveal in systems of widely varying composition and morphology suggest that a transport theory must be founded on universally shared characteristics [37.58]. The theoretical literature stimulated by these observations is in fact extensive and far too detailed to be adequately described here. In the present context it suffices to comment briefly on certain general trends. The activated temperature dependence universal to all these materials systems arises because states in the hopping manifold are energetically inequivalent. Energetic inequivalence can be understood on the basis of site relaxation accompanying polaron formation [37.59–61]. Energetic inequivalence of sites is also a common feature of disordered solids resulting from site-to-site fluctuation of the static and dynamic electrostatic potential [37.43, 62, 63]. The combination of both effects cannot be excluded. In fact, there is a general need to develop the framework to distinguish the relative contributions of these key processes in the analysis of experimental data [37.64, 65]. A particularly challenging issue has been the attempt to account for the special features of the Poole–Frenkel-like field dependence illustrated in Fig. 37.6. Polaron models typically predict that the log mobility should scale linearly with the applied field [37.59, 66]. The disorder model based on analysis of Monte Carlo computer simulation of hopping on a finite lattice as originally proposed by *Bässler* [37.43] was in fact able to self-consistently account for a number of key experimental features in terms of a limited set of disorder parameters. On the other hand, it could only model the observed field dependence over a very narrow range. It was later recognized in the formulation of dipole disorder models [37.67, 68], that slow site-to-site variation, as distinct from fully random fluctuation, in the effective electrostatic potential could more properly account for the field-dependent behav-

ior of mobility commonly observed over several orders of magnitude [37.69–71]. The more general formulation of the disorder model developed by Dunlap and coworkers, treats the disorder potential as fully correlated and therefore slowly varying, and argues that such a supposition is physically plausible on general grounds in glassy solids. Further advances in the theories of hopping charge transport in organic materials can be found in review articles [37.72–74].

If every chemical impurity remaining in even the most scrupulously processed polymer film could act as a trap, organics as a materials class would be completely excluded as a basis for the design of xerographic photoreceptors. Trap-free transport in CTLs was achieved not by entirely eliminating impurities but by properly designing the transport-active moiety [37.75, 76]. The associated studies first carried out in this technologically critical context on molecularly doped systems [37.77] were instrumental in generally elucidating trap interactions in hopping systems and in first unambiguously illustrating and analyzing the mechanism of trap-controlled hopping transport [37.78]. It was already pointed out that hole transport is supported in MDPs when the neutral transport molecule is donor-like. For electrons the corresponding neutral transport molecule is acceptor-like in character. In each case, the CTLs are typically unipolar when analyzed by TOF or xerographic techniques. As it turns out the principle (simple in hindsight) which applies for example to hole-transport CTLs is to make the transport molecule significantly more donor-like than any of the resident impurity species. The limit is that the material must be stable to air oxidation. From an energetic point of view the impurity levels which then lie above the hole transport states are anti-traps. An important class of small molecules which have these desired characteristics are the aromatic amines and TPD is a prime example of a particularly strong donor in this class of molecules. The trap-free nature of these transport layers is xerographically apparent because there is no buildup of bulk space charge as measured by Kelvin-probe techniques [37.79] even after tens of thousands of (CV_0)s of holes are discharged through a photoreceptor with this CTL as a component. The trap-free nature of hole transport can also be independently demonstrated by combining TOF drift mobility and analysis of current–voltage measurements carried out on any transport layer fitted with a semitransparent blocking contact on one surface and an ohmic contact on the opposite face [37.80–82]. Figure 37.9 illustrates the result for the polymer PTPB in which TPD is covalently incorporated as the transport active unit [37.81]. Semitransparent Al on the films exposed surface is used for the TOF measurements of drift mobility. Dark injection occurs under positive bias from the gold-coated mica substrate onto which the PTPB film had initially been deposited. PTPB is highly insulating so that the bulk dielectric relaxation time is always much longer than the transit time of any excess injected carrier at all applied voltages. With the Al contact under positive bias, hole drift mobility is measured as a function of applied voltage V. From the measured drift mobility the trap-free space-charge-limited current density (TFSCLC), J, [37.83] that would be sustained by an ohmic contact under positive bias can be calculated to a goodapproximation, even when the drift mobility is itself field dependent [37.84, 85] using the expression $J = 9\varepsilon\varepsilon_0\mu V^2/(8L^3)$. These values are represented by open circles in Fig. 37.9. The electric field in the film is, of course, nonuniform and E in plots of this sort is simply a shorthand for V/L, where V is the externally applied voltage. It should be noted that, under the present conditions, the TFSCLC is in principle the maximum current that the bulk of a transport medium such as PTPB can demand from a contact in the steady state. It should also be noted that the

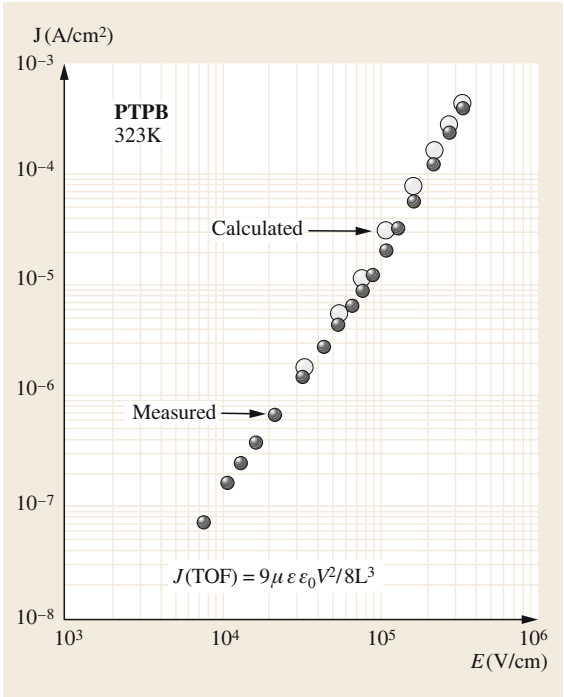


Fig. 37.9 The field dependence of the dark current for a PTPB film at 323 K measured with positive bias applied to an Au coated Mica substrate (*filled circles*). The trap free space charge limited current (*open circles*) is calculated from the hole drift mobilities obtained by the time of flight method

observation that an injecting contact under test is capable of sustaining a space-charge-limited dark current is a prima facie demonstration of ohmicity [37.26]. Filled circles represent the steady-state current measured when positive bias is applied to the substrate. That these calculated and measured currents coincide demonstrates that the injecting Au contact is ohmic and that the PTPB film is trap-free. Similar results have been obtained for other commonly used CTL MDPs such as TPD in bisphenol-A polycarbonate and in films of poly(methylphenylsilylene).

37.2.3 Photogeneration

The CGL material is designed to strongly absorb light from the exposure system (red to near-infrared LEDs and lasers are typical). In a typical dual-layer OPC the CGL material is a thin (submicron) coating of a dispersed pigment. Simply, the pigment material absorbs a photon and mobile charge is created which is *injected* into the CTL and then drifts under the influence of the applied field to discharge the device. Charge generation is a very complex topic [37.86]. In general terms, absorption of a photon produces an excited state (exciton) of a correlated hole–electron pair (valence band and conduction band, respectively) which has some mobility. Measured exciton ranges can be on the order of 0.1 μm. The exciton may then undergo recombination, with the release of energy to the environment, or it may be converted into charges. Conversion into charges can occur *spontaneously* or through chemical (impurities or dopants) or physical (crystal defects or the crystal surface) interactions. With the trapping of one of the charge carriers the other is free to drift under the influence of the applied field and geminate recombination is suppressed. The mobile carrier (assume holes) must drift to where it is in contact with the hole transport moieties of the CTL. In fact, in most cases the CTL penetrates and mixes with the CGL so that, even if the CGL is formulated without transport material, we can assume that these molecules are in molecular contact with the pigment particle surface. Now the *hole* is available to oxidize or *inject* into the transport moiety to produce the radical cation. Useful hole-transport materials have relatively low oxidation potentials (≈ 1 V versus the saturated calomel electrode (SCE)). The oxidation potential must be lower than that of the hole at the pigment surface but not so low that the material is unstable to air oxidation. The CTL hole is now transported through the CTL, discharging the photoreceptor, by sequential one-electron redox reactions between adjacent transport molecules as described in the previous section.

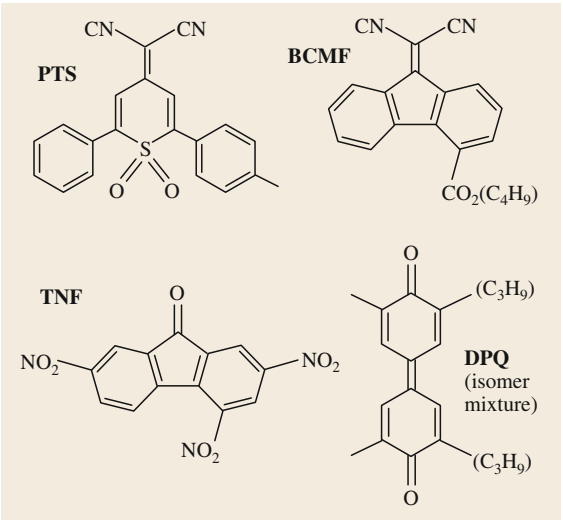


Fig. 37.10 Chemical structures of representative electron transport materials

The charge-generation mechanism is characterized as *intrinsic* when the hole is generated within the pigment particle without the influence of any externally added materials. In this case the hole is injected into the CTL with concomitant oxidation of a transport moiety at the pigment surface. *Sensitized* generation is said to take place when the photoconductivity is enhanced by an added material (dopant). In this case the dopant is available to react with the pigment exciton at the particle surface. In most cases the dopant is in fact the hole-transport moiety. Dopants may also influence charge generation by forming a (charge-transfer or other) complex which facilitates charge separation. Such complexes may be accessed by direct photoexcitation or via exciton diffusion. Sensitization of the photoconductivity of poly(N-vinyl carbazole) (PVK) by electron acceptors such as 2,4,7-trinitro-9-fluorenone (TNF) (Fig. 37.10) has been extensively studied [37.87–95]. Charge-generation mechanisms in pigment-based CGLs depend on the pigment molecular structure, crystal structure, morphology, and on the transport molecule used in the formulation.

The fundamental quantity of interest in charge generation is the quantum efficiency as a function of field. A strong field dependence is always observed, as shown in the example of Fig. 37.11. Several models have been proposed to explain charge-generation characteristics. These will be summarized here but more complete discussions can be found in *Pope* and *Swenberg* [37.86], *Popovic* [37.97], *Law* [37.3]; *Borsenberger* and *Weiss* [37.5], and *Weiss* et al. [37.7]. One of the early models to be applied to organic

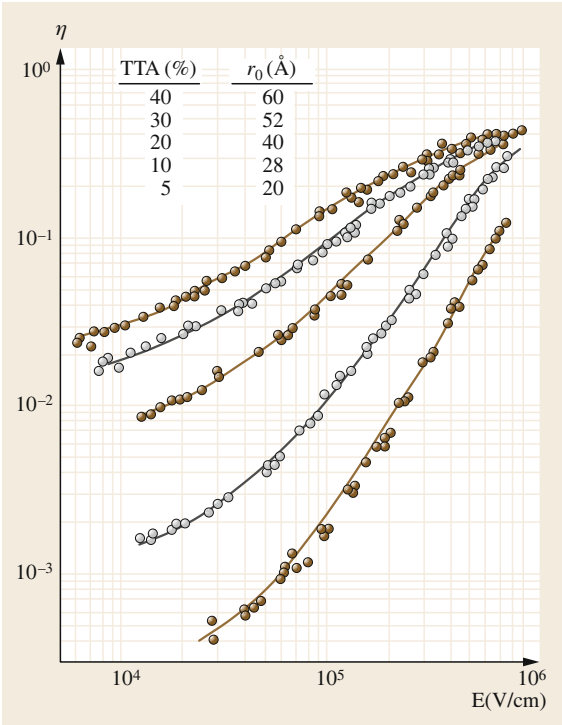


Fig. 37.11 Quantum efficiency of the field dependence of generation for a dual layer dye-polymer aggregate based OPC. The solid lines were calculated from the Onsager model with $\eta_0 = 0.60$ and r_0 as shown. (After [37.96])

photoconductors was the *Onsager* model of geminate recombination [37.98]. The 1938 theory gives the probability that a charge pair in thermal equilibrium will separate under the influence of an applied field in competition with geminate recombination. In this model charge separation occurs from an upper vibrational state to produce a *geminate* pair of charges which either recombine due to their Coulombic attraction or are separated under the influence of the applied field. The parameters in this model are the field-dependent quantum efficiency for geminate-pair formation, η_0, and the initial separation distance of the geminate pair (r_0) [37.99]. The data in Fig. 37.11 is for a dual-layer dye–polymer-aggregate-based OPC with varying concentrations of charge-transport material (tri-*p*-tolylamine, TTA) in the CTL [37.96]. The data has been fit to the 1938 model with $\eta_0 = 0.60$ and r_0 as shown. The model has been refined using a modified basis, also due to *Onsager* [37.100]. Further refinements have involved modifications of the distribution function assumed for the separation distances [37.101, 102]. This model has been successfully used to describe charge generation in amorphous materials [37.103]. However, these models predict a wavelength-dependent quantum efficiency and a photoconduction threshold which is at higher energy than the absorption, neither of which is observed in pigment-based generation materials. A model for charge generation in these materials invokes decay (radiative or radiationless) of the initially formed excited state followed by a competition between field-dependent charge separation and internal conversion [37.104]. Sensitized generation occurs with the reaction of the thermalized excited state with a dopant. Studies of the field dependence of pigment fluorescence quenching and lifetime have been used to sort out the details with specific generation materials [37.97, 105–107].

The quantum efficiency for carrier generation can be defined in several ways, as described above. The efficiency of geminate-pair formation refers to the initially produced charge pair and this is also the efficiency which has been obtained with fluorescence quenching studies. On the other hand, the generation efficiency determined with xerographic discharge always includes some contribution from charge transport because a measured decrease in photoreceptor surface potential is due to charge motion. In addition, very-high-intensity exposures may give rise to (bimolecular) recombination of positive and negative carriers during transport [37.108–112]. Quantum-efficiency determinations from xerographic discharge are carried out with very short, low-light-intensity continuous exposures to minimize the influence of charge transport and recombination (see the following section).

37.3 OPC Characterization

Photoreceptors may be characterized by many techniques. On the one hand one wishes to determine the fundamental characteristics of the materials, while on the other hand one wishes to determine electrophotographic performance-based characteristics of the photoreceptor. The quantum efficiency of charge generation and, depending on the model chosen, the initial separation distance of the charge pair, are common characterizations. Electrophotographic performance characterizations include dark decay, photoinduced discharge sensitivity, and residual potential as well as their stability to electrical-only cycling.

37.3.1 Dark Decay

The quantity of interest is the rate of dark discharge (V/s) as a function of field and environment. However, the dark decay characteristics of a dark-adapted photoreceptor when incrementally corona charged to higher and higher fields will be significantly different from what is observed in the electrophotographic process. This is because OPCs seldom behave as perfect capacitors and there is typically a depletion charging component where previously trapped charges are mobilized and drift in the photoreceptor on the application of the corona charge [37.28–30, 113]. Thus, in addition to field and environment, the observed dark decay characteristics depend on prior exposure history (trapped charges), corona charger characteristics, and timing. Thus, a given photoreceptor will have a higher dark decay rate if the photoreceptor is charged exactly to the desired initial potential (V_0) and the dark decay rate determined immediately, as opposed to charging well above the desired potential and measuring when the surface potential decreases to the desired potential. Dark decay measurements which are meaningful to the electrophotographic process must be carried out in an apparatus in which the exposures, charging, and timing elements are fixed. Dark decay is an important characteristic of an OPC because, although process control will attempt to maintain V_0, the imaging system in a printer will in general not tolerate a significant V_0 drop.

37.3.2 Photosensitivity

Measurements to determine fundamental characteristics (quantum efficiency of generation materials, mobility, etc.) are often carried on photoreceptors that differ from what might be eventually used in a printer. The layer of interest (CGL or CTL) may be isolated or the device may be modified with the application of a vacuum-deposited electrode (Au is common) on the free surface. Common fundamental characterizations include quantum efficiency as a function of field and wavelength, and mobility. Mobility was discussed in Sect. 37.2.2. The quantum efficiency of carrier generation has several definitions as applied to OPCs. On the one hand it can refer to the number of ion pairs generated per incident or absorbed photon. This fundamental characteristic might be determined experimentally by fluorescence quenching, or calculated based on a model such as the Onsager model discussed above. On the other hand a photodischarge efficiency can be determined based on the decrease in surface potential per photon absorbed or incident. The photoinduced discharge (PIDC) method is commonly used to characterize OPCs as a function of field, wavelength, temperature, humidity, exposure intensity, etc. The photoreceptor is corona charged to apply the desired field and is photodischarged. Using the parallel-plate-capacitor model, $Q = CV$ (terms as defined previously), the change in surface charge density can be calculated from the decrease in the surface potential. The change in surface charge density per unit area per photon incident or absorbed is a measure of photosensitivity. This measure includes the efficiency of charge generation, injection into the CTL, and transport through the CTL. Because of carrier bimolecular recombination and range limitations the quantum efficiency determined using this method may depend on the light intensity and the extent of photodischarge. The photodischarge method is carried out with either continuous or flash exposures. With continuous exposures ($J/(cm^2\,s)$) the surface potential is monitored as a function of time. One measure of OPC photosensitivity which relates to the electrophotographic process is the exposure required to discharge the photoreceptor a fixed amount, often 50%, from an initial potential. This is typically carried out using exposures (wavelength and duration) relevant to a particular electrophotographic process. The photoreceptor sensitivity is usually reported in terms of J/cm^2 (the inverse is often used to report spectral sensitivity). A related method is to determine the initial photodischarge rate (dV/dt). This can be related to the decrease in surface charge density through the capacitance, $dQ/dt = C dV/dt$. The light intensity is kept low and the discharge time short to avoid space-charge perturbations during charge transport and to keep the field essentially constant. The ratio of the decrease in surface charge density to the exposure ($photons/cm^2$) is a measure of photodischarge efficiency. In fact the photodischarge per exposure ($V/(J\,cm^2)$) is another commonly used metric for OPC photosensitivity. When determined as a function of field, data obtained in this manner can be analyzed in terms of the Onsager models. Because photosensitivity depends on field (surface potential per unit OPC thickness) and the exposure produced decrease in surface charge density (related to the change in surface potential through $\Delta Q = C\Delta V$), meaningful comparisons can only be made between OPCs of similar thickness, charged to similar initial potentials, and exposed to similar discharged potentials.

For a process using a flash exposure the photoreceptor discharge is determined at several exposure levels (with the surface potential determined at a fixed time after the exposure) and the data is plotted as surface potential (V) versus log(exposure, J/cm^2). An arbitrary

photodischarge point (typically 50%) can be determined from this plot for comparison purposes. The utility of this characterization is that the exposure characteristics and the timing for reading the discharged surface potential directly relate to the printing process. As discussed above, short high-intensity exposures may give rise to Langevan recombination [37.110, 112, 114], resulting in decreased sensitivity (reciprocity failure) relative to low-intensity exposures.

An often overlooked sensitivity factor is that there are more photons per energy unit (J) as the wavelength increases. For example, there are about 30% more photons/J at 820 nm relative to 630 nm. Thus, for a photoreceptor with a given quantum efficiency for carrier generation the sensitivity will increase with increasing wavelength of the exposure source.

37.3.3 Electrical-Only Cycling

Electrophotographic performance characterizations are necessarily carried out on the photoreceptor in an apparatus which has the key elements of the electrophotographic process. The most *relevant* apparatus would of course be the fully configured printer. However, the use of such a complex device is often undesirable, or unnecessarily, for screening purposes. In the laboratory it is common to use *electrical-only* test fixtures in which the process always includes charge and expose, and may include erase (or unique process elements), with the surface potential being determined after the application of each. With an apparatus such as this it is possible to determine process-relevant characteristics such as: chargeability, dark decay, photosensitivity, and residual potential, all as a function of cycling. Of course large changes in any of these characteristics is generally undesirable and long cycling (10–100 kcycles) is often necessary to adequately determine the electrical stability of the photoreceptor to the electrophotographic process. Such stability is a necessary, but not sufficient, condition for eventual commercialization, and extensive testing in the machine for which the photoreceptor was designed is always the final step in the development process. Other factors, often connected to the device physical characteristics (wear rate, corona chemical sensitivity, layer adhesion, seam strength, etc.) may in fact determine the eventual success or failure of a particular OPC in a printer.

37.4 OPC Architecture and Composition

As we have described, OPCs are large-area thin-film devices that are insulating in the dark and become conductive when exposed to light. A high field is applied by exposure to an air corona (or similar charging method) and the photoreceptor is image-wise exposed to produce an electrostatic latent image. The surface potential in the exposed areas is determined by the exposure intensity and the OPC photosensitivity while the surface potential in the unexposed areas is determined by the dark conductivity. The development system deposits charged marking particles (toner) on the photoreceptor surface according to the difference in surface potential between the exposed and unexposed areas. The key photoelectrical processes which enable latent image formation are photoinduced charge generation and the transport of the generated charge through the device. Organic photoreceptors accomplish this in a multilayer architecture with specific materials (monomers and polymers) designed and optimized for specialized functions: binder, charge generation, charge transport, dark conductivity, etc. In this section we will discuss the architecture and composition of organic photoreceptors with details on the function and composition of each layer.

37.4.1 OPC Architecture

The choice of photoreceptor thickness is a balance between capacitance, surface potential, and internal field. Capacitance relates to surface charge density, which determines the achievable developed toner density for toners with a given surface charge per area. Surface potential, in combination with the development system, provides the potential difference to attract toner to the photoreceptor surface. The internal field enables efficient charge generation and transport. In most electrophotographic printing machines the balance is achieved with an OPC thickness of $\approx 20-30\,\mu m$ and a surface potential of 500–1000 V.

A schematic cross section of a fully configured modern electrophotographic photoreceptor is shown in Fig. 37.12. In its most simple form the photoreceptor consists of a conductive substrate and a single layer of photoconductive material to accomplish charge generation and transport. The advantage of this structure is in its simplicity, but the disadvantage is that not all characteristics will be optimum. In the modern embodiment the photoreceptor is constructed with separate layers for charge generation (CGL) and charge

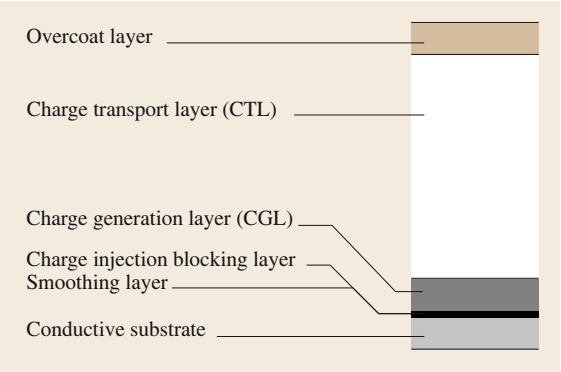


Fig. 37.12 A schematic cross section of a fully configured *modern* electrophotographic photoreceptor

transport (CTL). Other layers shown in Fig. 37.12 include: substrate, smoothing layer, charge-injection blocking layer, charge-generation layer (CGL), charge-transport layer (CTL), and protective overcoat layer. Additional layers, such as adhesive interlayers, and a back-side anti-curl layer may also be present. OPCs with separate CGL and CTL are termed dual-layer. Each of these layers may be a complex formulation of materials. The advantage in separating layers is that each layer can be optimized for its specialized function. This adds materials expense and manufacturing complexity but has the advantage of superior performance.

In a single-layer photoreceptor when exposed with strongly absorbed radiation directed at the free surface, charge generation will occur near the free surface and the photodischarge will be dominated by the transport of positive or negative charge through the bulk of the layer. The polarity of the mobile charge will be the same as the applied surface potential. In most commercial dual-layer photoreceptors the architecture is substrate–CGL–CTL. Because the CTL is formulated for hole transport, dual-layer OPCs are charged with a negative surface potential. Positive corona charging has some advantages over negative (better uniformity and less ozone production). A single-layer OPC can function with either a positive or negative surface charge depending on the exposure direction and the transport characteristics of the materials used in the formulation. Positive-charging dual-layer photoreceptors with the architecture substrate–CTL–CGL have been commercialized but these require a protective overcoat to prevent the CGL from being damaged by contact with the electrophotographic process elements. The typical substrate–CGL–CTL architecture will function with positive charging if the CTL is formulated for electron transport. Electron-transporting CTLs have been widely studied by not yet commercialized.

In a multilayer photoreceptor the principle is that the functions of each layer are optimized. Thus, with separate CGL and CTL the former is designed for strong absorption of the wavelength of light used in the process and efficient charge generation, and the latter is optimized for the injection of photogenerated charge from the CGL and transport of that charge to the free surface. In general, most charge-generation materials efficiently transport only one polarity of charge (electrons or holes depending on the material) so that it is generally desirable to have the CGL as thin as possible while retaining the required optical density at the exposure wavelength. It is also necessary that the CGL be uniform over the entire area of the photoreceptor to maintain consistent sensitivity for the entire photoreceptor surface. Submicron CGLs are typical of the photoreceptors used in most modern digital printers. The CTL determines the device capacitance and a thickness of $\approx 25\,\mu m$ is typical. However, the CTL thickness also influences the ultimate latent-image resolution in a digital printer because holes transiting the CTL experience mutual Coulombic repulsion [37.13, 115]. It has been shown that for optimum resolution the photoreceptor thickness should be about one quarter of the pixel size [37.13, 14]. Thus, for 600 dpi imaging ($\approx 40\,\mu m$ pixel size) the ideal photoreceptor is on the order of $10\,\mu m$. Thin OPCs have not yet been introduced into digital printers but one might anticipate that this will occur as toners get smaller and the processes of toner transfer and fusing are refined to minimize image disruption [37.116].

37.4.2 Coating Technologies

Hopper coating of continuous webs and dip coating of drums are the two technologies commonly used to fabricate photoreceptors. As with any large-area photoelectrical device absolute cleanliness is required to prevent artifacts in the coated layers. In a multilayer device one must carefully control the interfaces between the layers. For example, a key interface is that between the CGL and CTL. Here it is important that the two layers adhere but do not mix to such a degree that the desired characteristics of either layer are compromised. Attaining the desired degree of mixing is a process of optimization depending on the materials, solvents, and coating method.

In hopper coating the coating solution is continuously pumped through a slot onto the moving web. Film thickness is controlled by the pumping rate and solids content of the coating solution. Precise control of drying temperature, humidity, and air flow enable optimization of the coated layers. Curing of the coated film involves removal of solvent and annealing of the

polymer binder(s). In a hopper coating process several layers can be coated in a single pass with sequential hoppers as long as the previous layer is adequately cured. Hopper coatings can be carried out at a very high speed (m/s) to produce a roll of many thousands of feet. To coat subsequent layers the previously coated substrate is re-run through the coating machine. Because the hopper coating process involves contact of the coated surfaces, and winding of the coated web, there is the possibility of film damage. Final inspection for film defects must be carried out and damaged, or imperfectly coated, regions must be removed before construction of the final photoreceptor loops.

In a dip coating process the substrate and coated layers are dipped into a solution or dispersion of the next layer to be coated and slowly withdrawn and cured. The coated layer thickness is determined by the rheological characteristics of the coating solution and the rate of withdrawal. Under these conditions there is the opportunity for extensive interaction between the previously coated layer and the solution into which it is dipped. Thus, formulations which are suitable for hopper coating may be inadequate for dip coating even though the chemical compositions of the final coated layers might be identical after curing. Another consideration with dip coating is that components of the previously coated layers which are soluble in the solvent used in the subsequent dip will contaminate the dip reservoir. Low levels of contamination may be insignificant, but since the amount of fluid removed in each coating operation is a small fraction of the total volume one must make sure that such contamination does not degrade performance.

Other processes such as ring coating (a dipped drum is withdrawn through an annulus which acts as a doctor blade to control the wet coating thickness), spray coating, vacuum coating, etc. may be used for specific applications and materials. In all layers coated from solution it is often necessary to add a small quantity of surface-active agent as a coating aid to enable smooth, defect-free coatings.

In recent studies on charge transport in free-standing doped polymer films it has been reported that the surface exposed to the air during drying may have degraded transport characteristics, and a depleted concentration of charge transport material was suggested as the cause [37.117, 118]. This finding demonstrates that solvent coated layers are not necessarily of uniform composition. Inhomogeneous composition of a charge-transport layer could lead to degraded electrophotographic performance. Thus, in analyzing the electrophotographic performance of an OPC one must consider not only the composition of each layer, but also its uniformity both spatially and compositionally particularly at the interface with another layer.

In the following sections we will discuss details around the compositions of each of the layers which might be found in an organic photoreceptor for electrophotographic printing. The order of discussion is in the order in which the typical photoreceptor would be constructed: substrate and ground layer, smoothing layer and charge-blocking layer, charge-generation layer (CGL), charge-transport layer (CTL), and protective overcoat layer.

37.4.3 Substrate and Conductive Layer

A common element to all electrophotographic photoreceptors is a substrate on which all the layers are deposited. In a research environment a small square of glass or quartz is convenient, but in commercial applications the substrate is invariably either a polymer film or a metal drum. Some of the material characteristics which might be considered in the choice of a polymer film are: stiffness, toughness, transparency, core set, and surface friction. Polymeric photoreceptor substrates are often 3–7 mil. A common polymer film material is the polyester poly(ethylene terephthalate) (PET). PET film is very durable, optically transparent, and readily available in a wide range of thickness. Since it is highly insulating it must be coated with a conductive material to act as the ground plane. PET or other polymer films are available in large rolls and the coating operations are usually done on the roll rather than cut-to-size sheets. Although a solution-coated conductive layer is possible, the most common ground layers are metals such as Ni, Al, Cr, and Ti, which are applied in a continuous sputtering or similar vacuum technique. The metallized rolls are subsequently coated with the photoconductive layers. If a metal drum is the chosen substrate it is almost invariably aluminum (1–10 mm wall thickness). The conductive metal serves as the ground layer but typically it receives further treatment before coating the photoconductive layers. The treatment may be chemical such as anodization to form a thin hole-blocking layer, or physical such as turning or sanding. Because these secondary surface treatments add cost, a smoothing layer is often used instead.

The electrical conductivity requirements of the conductive layer are modest. The higher the process speed the greater the conductivity required for adequate surface charge accumulation. Analysis by *Chen* [37.119] demonstrated that the result of inadequate conductivity in corona charging of a photoreceptor is a nonuniform surface potential. The surface potential is decreased depending on the distance from the actual grounding

contact. For a typical situation the upper limit of the sheet resistance is $\approx 10^4\ \Omega$ per square.

Alternative approaches such as *seamless* webs and conductive plastic drums have been developed but have not yet been commercialized.

37.4.4 Smoothing Layer and Charge-Blocking Layer

For xerographic applications it is important that the conductive layer does not inject charge into the photoreceptor in the dark. However, the metals used as grounding layers all have the potential to inject holes into the photoreceptor CGL via direct contact with the charge-generation material or the hole-transport material. The injection and transport of charge in the dark would cause a loss of surface potential (dark decay). Since the formation of a latent image relies on a difference between the rate of dark- and light-induced surface potential decrease, a high dark decay rate means that the process must be amenable to a small imaging potential. Another aspect of dark injection from the conductive layer is that such phenomena often occur in small localized *charge-deficient spots*. In a process where the discharged areas are developed small deposits of toner are found in what should be non-toned background. To counter such processes it is common to interpose a charge- (typically hole-)blocking layer between the ground layer and the CGL. With drum photoreceptors it may be the roughness of the metal surface which initiates the charge injection. In this case it may be necessary to interpose a relatively thick *smoothing layer* between the metal surface and the CGL.

In principle, any good film-forming insulating material will suffice as a charge-blocking layer as long as adhesion to the ground layer and CGL are adequate. Several types of materials have been used for this purpose including polysiloxanes and nylons. Since these materials are insulating, in an electrophotographic charge/expose cycle there will be a residual potential due to the field remaining across the layer. This will build up with cycling depending on the layer conductivity and the process cycle time. Although residual potentials typically decrease when the photoreceptor is rested in the dark, in practice such relaxation is usually incomplete and in subsequent cycling the residual builds up at a faster rate. The physics and chemistry of such hysteresis is not well understood. At some point the residual potential may cause image degradation and necessitate photoreceptor replacement. Thus, the blocking layer must be thick enough to cover the ground layer uniformly but not so thick as to cause an undesirable residual potential. Generally such a layer is coated at around 1 μm or less. On the other hand, if a thicker layer is required (perhaps combining the functions of both charge blocking and surface smoothing) this can be accomplished with a charge transporting insulating layer. With the conventional multilayer OPC architecture (holes transported through the CTL), this layer must transport electrons injected from the CGL to the electrode and block hole injections from the electrode into the OPC. Such polymer films have been described with naphthalene bisimides as the electron transporting moiety [37.120, 121].

With an aluminum drum, a thin anodized layer is often used to prevent charge injection. However, anodization is relatively expensive, and it has been reported that a simple chemical process (dilute alkali treatment followed by immersion in boiling water) produces an effective blocking layer [37.122]. Making the drum surface smooth on a submicron scale is expensive and an alternative is to use a relatively thick smoothing/blocking layer to cover the surface irregularities. Smoothing layers are typically metal oxides (titanium for example) in a polymer binder. The oxide loading is high to impart sufficient conductivity to prevent a residual potential. If the metal oxide does not inject holes it also serves as the charge-blocking layer. Such layers are up to around 10 μm thick.

Imaging with monochromatic radiation in digital printers has led to a new imaging artifact caused by interference between the incoming and reflected light. In dual-layer OPCs with the CTL uppermost, if the CGL incompletely absorbs the exposing light it can be reflected from the ground layer. The interference pattern that is produced can be visible in the developed image as a *wood grain* pattern. One way to counter this is to scatter the incoming light before it's reflected. There are many patents describing techniques for carrying this out including roughening of the surface of the conductive layer.

37.4.5 Charge-Generation Layer (CGL)

The CGL is where the imaging light is absorbed and charges are generated. The first major consideration in choice of material is that it strongly absorb the wavelengths of light emitted by the exposure system. In copiers light is reflected from a document onto the charged photoreceptor. Since an original may have colored components the light source must cover the visible range (400–700 nm) and the photoreceptor must have broad sensitivity to visible light. Common light sources in copiers are fluorescent or xenon lamps. Both sources are generally filtered to remove ultraviolet and infrared light. In digital printers the exposing source may be an LED array (680–780 nm) or a scanned laser

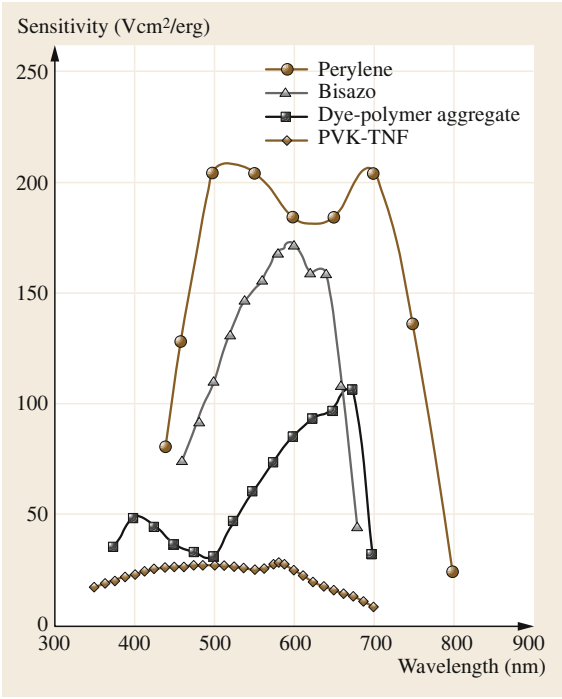


Fig. 37.13 Sensitivity action spectra of visible sensitive OPCs. The chemical structures of the CGL materials are given in Fig. 37.15

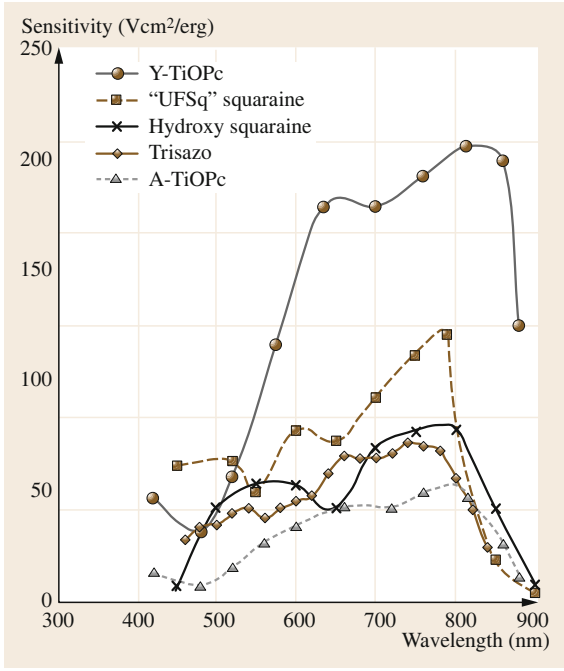


Fig. 37.14 Sensitivity action spectra of near-infrared sensitive OPCs. The chemical structures of the CGL materials are given in Fig. 37.16

(633 nm from He−Ne or 740, 780, and 820 nm from laser diodes).

Most modern commercial photoreceptors utilize pigments as charge-generation materials. Pigment-based CGLs comprise submicron crystals of the charge-generation material suspended in a polymer binder. Pigments are milled to the desired size and stabilized with appropriate additives and binders to produce a coating dispersion. Charge-generation materials are generally classified according to their chemical identity. Materials which absorb primarily in the visible region are dye–polymer aggregate, PVK–TNF charge-transfer complex, perylenes, and bisazos. Phthalocyanines, trisazo, and squarylium pigments are utilized for near-infrared absorption. Figures 37.13 and 37.14 show the sensitivity action spectra and Figs. 37.15 and 37.16 the chemical structures of these charge-generation materials. The sensitivity action spectra in these figures are taken from literature data and the sensitivities converted into consistent units ($V/(cm^2\,erg)$) based on information given in the cited publication. Because the data are from OPCs with different thickness, exposure intensity, exposure time, initial potential, and exposed potential, the sensitivities are not directly comparable. However, the wavelength ranges are accurate. When the exposure is through the CTL (in a negative-charging dual-layer OPC) the photosensitivity will be decreased at wavelengths below ≈ 400 nm because of light absorption by the transport material.

Perylenes, azos, and phthalocyanines are best known for their use as colorants and it is because these materials were readily available that they were investigated as charge-generation materials. A complication with pigments is that some exhibit polymorphism. Over many decades the colorant industry has developed technologies for controlling polymorphism to obtain pigments with specific colorant characteristics and similar technologies have been developed to obtain polymorphs with the appropriate absorption as well as charge-generation characteristics for electrophotographic applications. Thus, in many cases to prepare a CGL it is first necessary to convert the pigment-generation material into a dispersion of submicron particles of the desired polymorph. Examples of charge-generation materials are discussed below.

The donor–acceptor charge-transfer (CT) complex between poly(N-vinyl carbazole) (PVK) and 2,4,7-trinitrofluroren-9-one (TNF) was the first commercialized organic photoreceptor [37.123, 124]. The photoconductivity characteristics have been extensively studied [37.87–89, 92–94] and reviewed [37.125–128]. With the addition of TNF the carrier-generation efficiency of PVK is increased and extended into the region

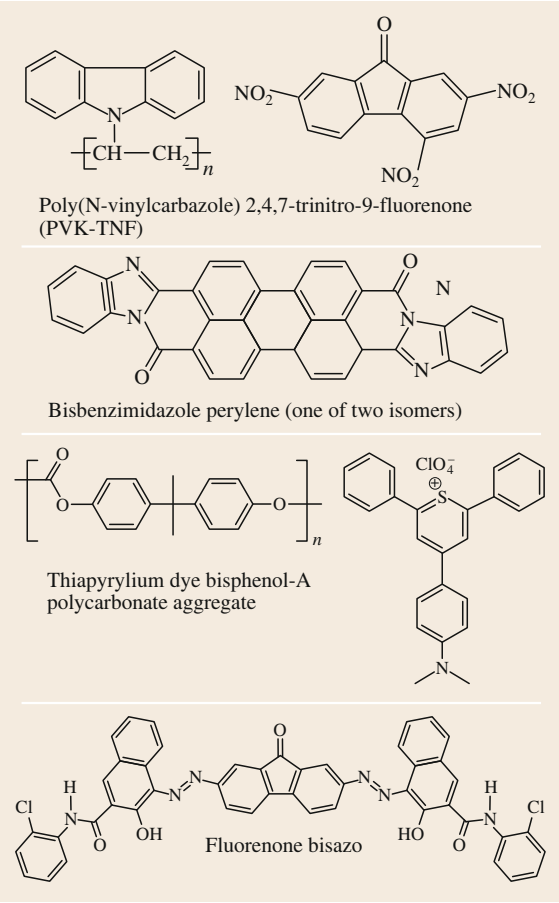


Fig. 37.15 Chemical structures of visible sensitive charge generation materials

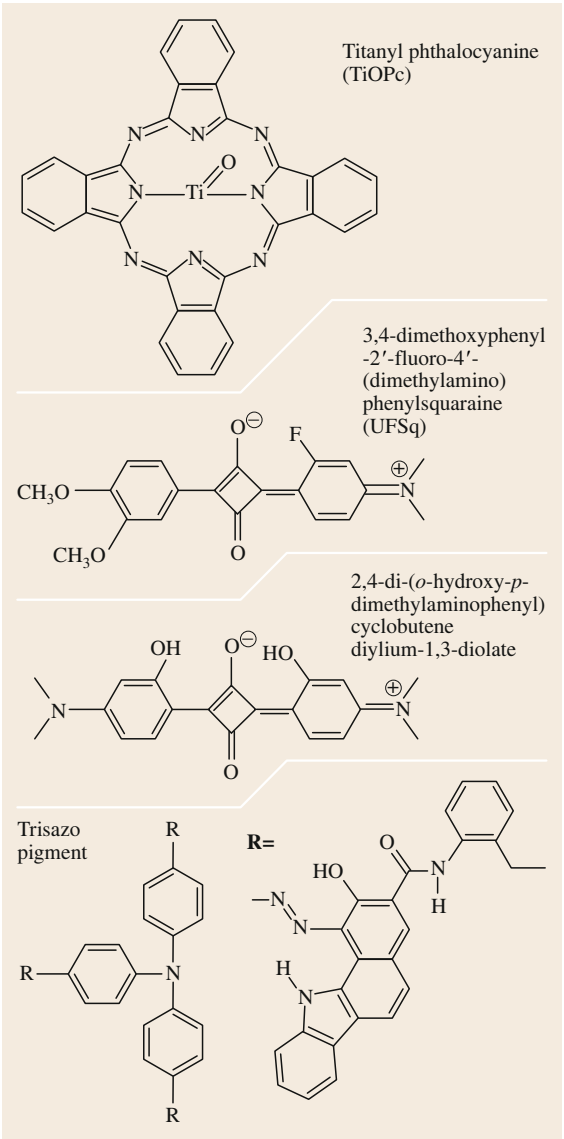


Fig. 37.16 Chemical structures of near-infrared sensitive charge generation materials

of CT complex absorption. Based on field quenching of exciplex fluorescence a mechanism for PVK–acceptor systems was proposed involving carrier generation from a nonrelaxed exciplex which is accessed by either direct absorption into the CT state or via an encounter complex between excited PVK and TNF [37.92–95]. As used in the IBM copiers I and II a 1 : 1 molar ratio of PVK to TNF was coated as a single 20 μm layer and electron-dominated transport was used to create the latent image (negative surface charge) with a He−Ne laser exposure (632.8 nm) [37.124]. An exposure of 5.5 $\mu J/cm^2$ was used to decrease the surface potential from −750 to −200 V. The quantum efficiency based on incident photons was ≈ 0.2 while the maximum efficiency was ≈ 0.3 at ≈ 400 nm. The photosensitivity action spectrum for this OPC is shown in Fig. 37.11 [37.123]. Using the photoinduced discharge technique [37.90] the maximum quantum efficiency (charges transported through the device per photon absorbed) was found to be 0.14 at 375 nm and a field of 10^6 V/cm. The PVK–TNF-based OPC had several drawbacks: low photosensitivity, short lifetime in the electrophotographic process, and concerns about the safety of TNF. It was eventually replaced with dual-layer OPCs with CGL materials based on pigments. Acceptor-doped PVK continues to receive attention particularly with C_{60} and related molecules as the acceptor [37.129].

The dye–polymer-aggregate based CGL is unique in that the active material is formed in situ during the coating process. A solution of thiapyrylium dye (2,6-diphenyl-4(4-dimethylaminophenyl)thiapyrylium per-

chlorate) and bisphenol-A polycarbonate is coated to form a CGL in which the dye and polymer spontaneously aggregate [37.130–132]. The aggregate is a dense filamentary structure which has an absorption spectrum red-shifted from that of the dye. The aggregate filaments transport electrons with high field trapping due to carrier immobilization at filament dead ends. The photosensitivity of the aggregate is several orders of magnitude larger than that of the nonaggregated dye [37.133]. To prepare a CGL the formulation also includes hole-transporting materials. Because the CGL transports both holes and electrons it is possible to use the aggregate as a single layer or as the CGL in a dual-layer photoreceptor. Because the aggregate absorption drops rapidly at wavelengths longer than 700 nm this material is only useful for visible-light exposure systems, Fig. 37.13. In this example the CGL (2 μm) comprised dye and tri-*p*-tolylamine (TTA) hole-transport material in polycarbonate, and the CTL (13 μm) was a solid solution of 40 wt% tri-*p*-tolylamine in a polyester binder [37.96]. This was charged to −600 V and exposed with very low-intensity continuous irradiation to obtain the initial rate of photodischarge. At 680 nm the photogeneration efficiency is 0.34 per absorbed photon. The quantum efficiency (Onsager model) was 0.6 and the thermalization distance was dependent on the concentration of hole-transport material ranging from 200 nm (5 wt%) to 600 nm (40 wt%) [37.96], Fig. 37.11.

Azo compounds contain the −N=N− functionality. Bisazo compounds have the general structure R′−N=N−R−N=N−R″. Trisazo compounds have the general structure R(−N=N−R′)(−N=N−R″)(−N=N−R‴). The R substituents generally have extended conjugation to provide the colors which have made bisazo compounds important dyes and pigments for many years. Bisazo CGLs generally absorb in the visible and trisazo into the near-infrared. An early example of a bisazo pigment-based OPC (chlorodiane blue, 4,4′-bis(1″-azo-2″-hydroxy-3″-naphthanilide)-3,3′-dichlorobiphenyl) was used in the IBM 3800 printer. This photoreceptor had over four times the photosensitivity of the PVK–TNF-based OPC [37.134]. Azo charge-generation materials are often synthesized from the reaction of an aromatic amine with an o-hydroxyaromatic carboxylic acid such as naphthol-AS. As CGL materials these materials are used as dispersed pigments. Although bisazo pigments do not generally exhibit polymorphism they can exist as hydroxy-azo or keto-hydrazone tautomers. In some cases, it has been found that the active form is the keto-hydrazone tautomer [37.135]. Because the synthesis of azo compounds is relatively easy there is a considerable literature on the effects of chemical structure on electrophotographic characteristics and carrier-generation efficiency [37.136–140]. In azo pigments charge is generated extrinsically with the pigment exciton oxidizing a hole-transport molecule at the pigment surface [37.140, 141] and it was demonstrated that penetration of the CTL into the CGL occurs during the coating process [37.142]. The photosensitivity action spectrum for a fluorenone bisazo pigment is shown in Fig. 37.13 [37.141]. In this OPC the CGL (0.17 μm) consisted of the pigment in a binder of poly(vinyl butyral) (10 : 4 weight ratio) and the CTL (17.4 μm) a solid solution of hydrazone transport material, p-diethylaminobenzaldehyde diphenylhydrazone (DEH), in polycarbonate (9 : 10 weight ratio). The initial potential was −800 V and the exposure required for a discharge to −400 V determined. The quantum yield of carrier generation from the Onsager model was 0.52 with a thermalization length of 1400 nm. Time-resolved transient absorption spectroscopy has been used to study the formation and decay of the oxidized transport material [37.143]. With a bisazo-based CGL (chlorodiane blue) [37.144] and trisazo-based CGL it was found that the charge-generation efficiency depended on the oxidation potential of the CTL transport materials [37.138]. In the most favorable case the bisazo had a quantum efficiency of 0.25 and the trisazo about 0.5. This was interpreted in terms of the energy gap between the hole at the pigment surface and the transport material. In a study using a series of bisazo pigment CGLs and a common CTL it was found that neither electron donating nor withdrawing substituents favor high photosensitivity [37.139]. Photosensitivity was found to depend inversely on the crystallinity of the pigment, which was interpreted in terms of pigment surface area. The development of high-sensitivity azo-based CGLs has been reviewed by *Murayama* [37.145]. The trisazo pigment-based OPC has an absorption which extends into the near-infrared, making it suitable for laser exposures. Figure 37.14 shows the photosensitivity action spectrum of an OPC with a trisazo pigment CGL [37.146]. The CGL (0.33 μm) consisted of the pigment in poly(vinyl butyral) binder (10 : 4 weight ratio) and the CTL (18.5 μm) was a solid solution of DEH transport material in polycarbonate (9 : 10 weight ratio). The initial surface potential was −700 V and the exposure to discharge to −300 V used to determine the photosensitivity at each wavelength. A photoreceptor with CGL (0.12 μm) and CTL (19.5 μm) had a quantum efficiency for charge generation of 0.46 from 500−800 nm [37.146].

Perylenes used as CGL materials are diimides of perylene-3,4,9,10-tetracarboxylic acid. One example is the N,N′-bis(2-phenethyl)-perylene-3,4:9,10-bis(dicarboximide). In this case the CGL was pre-

pared by vacuum deposition of an amorphous film (0.1 μm). Exposure to dichloromethane (as when overcoating with the charge transport layer) converts the film into a crystalline form with enhanced absorption at 620 nm [37.147]. The mechanism of carrier generation of this pigment in the presence of tri-*p*-tolylamine has been studied by studying the field dependence of quenching of perylene fluorescence [37.148]. The results were interpreted in terms of both intrinsic (directly from the perylene singlet state) and sensitized (interaction of holes on the pigment surface with the amine) charge generation. A bisbenzimidazole perylene has been developed as a CGL [37.149–151]. Dual-layer OPCs with a CGL of 0.1 μm vacuum-deposited bisbenzimidazole perylene, and a 15 μm CTL of 35 wt% TPD in polycarbonate, exhibit high sensitivity ($E_{1/2}$ 2 erg/cm^2 for −800 V_0) out to 700 nm, Fig. 37.13. Further research has involved the synthesis and study of novel materials such as dimeric perylene pigments [37.152].

Phthalocyanines have received considerable attention as CGL materials for near-infrared exposure systems. Metal-free (H_2Pc) [37.153–156], titanyl (TiOPc) [37.157], and hydroxygallium (HOGaPc) [37.158] phthalocyanines have been extensively studied as CGL pigments. Although these materials are relatively easy to synthesize their intractability makes them difficult to purify. Except for halogen substituents on the aromatic rings there has been little research into the effect of chemical structure on performance. Instead, much of the development has been to understand and control polymorphism. Compared with solution absorption spectra, the spectra of many CGL pigments are broadened with enhanced absorption in the red and near-infrared. With specific combinations of milling technique, solvent exposure, and thermal treatment many phthalocyanine polymorphs can be produced [37.159–162]. Polymorphs have been characterized by visible and near-infrared absorption spectroscopy, x-ray diffraction, nuclear magnetic resonance (NMR) spectroscopy, infrared spectroscopy, and infra-red (IR) absorption spectroscopy. Examples of TiOPc polymorph absorption spectra are shown in Fig. 37.17 [37.157]. Polymorphs may have very different photogeneration characteristics as CGLs in OPCs. The materials shown in Fig. 37.17 were fabricated into OPCs: CGL, 0.4 μm, pigment and poly(vinyl butyral) 1 : 1 weight ratio; CTL, 20 μm, 43 wt% triarylamine hole-transport material, 4-(4-methylstyryl)-4′,4″-dimethoxytriphenylamine, in polycarbonate. The photosensitivity action spectra for OPCs from the Y- and A-forms of TiOPc are shown in Fig. 37.14. The OPC with the Y-form had a photosensitivity for half photodischarge (600−850 nm) of 0.75 erg/cm^2 (initial potential −600 V) while the

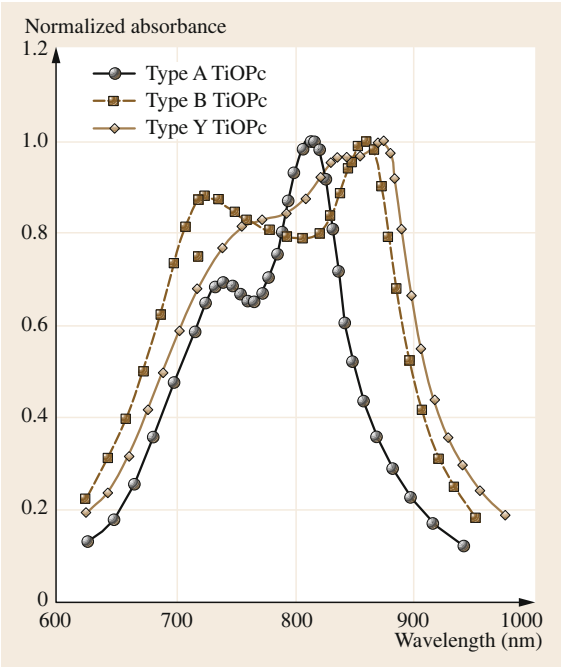


Fig. 37.17 Absorption spectra of polymorphs of TiOPc. (After [37.157])

A-form was much less sensitive. In another study, photoreceptors formulated with α-, β-, τ-, and X-forms of H_2Pc had generation efficiencies as follows (5×10^5 V/cm): $\tau_2 \approx 0.1$, $X_2 \approx 0.09$, $X_1 \approx 0.03$, $\beta_1 \approx 0.01$, $\alpha_2 \approx 0.005$, and $\alpha_1 \approx 0.001$ [37.163]. In the preparation of a generation material a common procedure is to prepare the pigment in the amorphous form by acid pasting or with a specific milling technique, and then to convert the pigment to the desired polymorph with a specific solvent treatment. The desired polymorph must of course be stable to all the subsequent operations. In addition to the dominant influence of crystal form on electrophotographic performance, studies have implicated water in carrier generation of Y-form TiOPc [37.157, 164]. A recent paper has reviewed the carrier generation mechanisms in several phthalocyanines [37.165]. In all the pigments studied (x−H_2Pc, TiOPc (I) (low sensitivity), TiOPc (IV) (also known as Y, high sensitivity), HOGaPc (V)), fluorescence quenching implicates a neutral excited state (not a charge-transfer state) as the precursor to carrier generation. Studies of the effect of field on fluorescence amplitude and lifetime reveal that with HOGaPc and TiOPc (I and IV) carrier generation comes from a non-relaxed excited state and TiOPc (IV) has an additional generation channel from a trapped state. With x−H_2Pc trapped excitons appear to be the major source of charge carriers.

Squarylium (squaraine) pigments are a class of materials which were first synthesized in the 1960s. Squaryliums (2,4-di-(*o*-hydroxy-*p*-dimethylaminophenyl)cyclobutene and 1,4-bis-(4-dimethylamino-2-methylphenyl diylium-1,3-diolate)) were the first pigments to be utilized as a near-infrared-sensitive CGL in an IBM printer [37.144, 166, 167]. The photoreceptor based on the hydroxy squarylium with a CTL (21 μm, DEH in bisphenol-A polycarbonate) had high photosensitivity from 500–800 nm (≈ 3 erg/cm^2 for a −700 to −200 V photodischarge), Fig. 37.14 [37.167]. This class of pigment has been extensively investigated for use as pigment-dispersion CGLs for near-infrared exposure systems [37.149, 168–174]. A nonsymmetrical squaraine (UFSq)-based OPC with CGL, 0.4 μm UFSq in poly(vinyl formal) (80 : 20 weight ratio), and CTL 26 μm 40 wt% TPD in bisphenol-A polycarbonate, exhibited an $E_{1/2}$ photosensitivity of 1.9 erg/cm^2 at 790 nm (−980 V_0) when optimally purified [37.174], Fig. 37.14.

37.4.6 Charge-Transport Layer (CTL)

In all current commercial organic photoreceptors holes are the carriers that dominate photodischarge of the device. Hole transport is carried out by aromatic amines which are either doped into, or are incorporated as part of, the binder polymer. The key molecular characteristic is that aromatic amines have relatively low oxidation potentials. The oxidation potential of the transport moiety must be lower than that of the charge-generation material such that photogenerated holes are injected from the generation material into the charge-transport layer. In a chemical sense the key reaction is oxidation of the charge-transport material by the photogenerated hole on the surface of the generation material. If the oxidation potential of the transport material is too low it will undergo air oxidation, if it is too high its oxidation will require an activation energy and will be inefficient. Molecular orbital studies suggest that molecules with little change in geometry with oxidation will be favorable for hole transport [37.175]. Aromatic amines with oxidation potentials between about 0.8–1.2 eV (relative to SCE) are generally useful as hole-transport materials. The synthesis and study of hole-transport molecules is a very active area. Figure 37.3 shows the chemical structures of some common hole-transport materials:

- 1,1-Bis(*p*-diethylaminophenyl)-4,4-diphenyl-1,3-butadiene (DPB)
- *p*-Diethylaminobenzaldehyde diphenylhydrazone (DEH)
- tri-*p*-Tolylamine (TTA)
- N,N′-Diphenyl-N,N′-bis(3-methylphenyl)-(1,1′-biphenyl)-4,4′-diamine (TPD)
- 1,1-Bis(di-4-tolylaminophenyl)cyclohexane (TAPC).

These are often referred to by acronym and classified according to one of their structural functionalities: arylmethane, diarylamine, enamine, triarylamine, heterocyclic, butadiene, and hydrazone. A more complete listing is available in *Borsenberger* and *Weiss* [37.5]. A wide variety of new hole transporting materials for electrophotographic photoreceptors have been reported. These include carbazole-containing enamines [37.176], indolo[3,2-b]carbazole derivatives [37.177], and phenyl-1,2,3,4-tetrahydroquinoline [37.178], among many others.

Charge transport was discussed previously (Sect. 37.2.2). Since transport involves orbital overlap between neighboring charge-transport moieties their concentration needs to be as high as possible. In practice, molecular transport materials are doped, singly or as a mixture, at 40–50% by weight into the binder polymer. The purpose of the binder polymer is to provide an inert, robust, substrate for the transport material. The mobility characteristics of a large number of hole-transport materials have been studied and the results interpreted according to a model where the dissolved transport molecules are disordered in energy and position [37.179–181]. Binder effects on hole transport have been observed. It appears that the presence in the CTL of materials with highly polar substituents decrease hole mobility but in general these effects are not well understood [37.182–187].

Since the CTL is typically the surface that is exposed to the various electrophotographic process elements it must be formulated to resist damage and wear. The glass-transition temperature (T_g) of the CTL is an important characteristic. Too low and the material will be soft and easily damaged, too high and it will be brittle. Generally a T_g around 60 °C is adequate. Polycarbonates and polyesters are two classes of materials commonly used as CTL binder polymers. Exposure of the CTL to the effluents of the corona charging system, UV light, reactive chemicals (ozone and nitrogen oxides), ions, etc., results in surface oxidation and degradation of both the physical and electrical characteristics of the CTL [37.188].

Dual-layer devices in which electron transport dominates have been studied extensively but never commercialized. Examples of commercialized electron-transport-dominated photoreceptors are the single-layer dye–polymer aggregate and the PVK–TNF charge-transfer complex. Electron-transport materials are designed to be easily reduced. The reduction potential

must be lower than that of the photogenerated electron on the charge-generation material so that it is transferred to the transport material. However, it is believed that oxygen may act as an electron trap so the reduction potential of the transport material must be lower than that required to reduce oxygen dissolved in the polymer film. Figure 37.10 shows the chemical structures of some electron-transport materials. As with hole-transport materials acronyms based on molecular functionality are common:

- 1,1-Dioxo-2-(4-methylphenyl)-6-phenyl-4-(dicyanomethylidene)thiopyran (PTS),
- *n*-Butyl 9-dicyanomethylenefluorene-4-carboxylate (BCMF),
- 2,4,7-Trinitro-9-fluorenone (TNF),
- 3,3′-Dimethyl-5,5′-di-*t*-butyl-4,4′-diphenoquinone and 3,5′-Dimethyl-3′,5-di-*t*-butyl-4,4′-diphenoquinone isomer mixture (DPQ).

A more complete table can be found in *Borsenberger* and *Weiss* [37.5]. In general the charge mobilities observed with electron-transport materials are much lower than for hole-transport materials. For this reason dual-layer OPCs with electron-transport-based CTLs have not been commercialized. Analysis of the mobility characteristics of electron transport with the disorder model implicates the highly polar nature of many electron-transport molecules [37.189–192].

Both hole- and electron-transport materials have been prepared with the transport moieties as part of a polymer (main chain or pendant) which would serve as a binder with hole-transport functionality. The advantage of a transport-active polymer is that it is a single material with improved physical characteristics relative to the corresponding molecularly doped polymer. In practice, charge transport is generally not enhanced in these materials and their chemical complexity makes them expensive to synthesize. Also, they are still susceptable to the damaging effects of the charging system in the electrophotographic process. Many polymers have been synthesized utilizing hydrazones [37.193–195] or other hole-transporting functionalities [37.196, 197].

Backing Layer

In a web photoreceptor the substrate–photoreceptor may tend to curl due to the residual stress of the coated layers. Such curl can make it difficult to fabricate the web into loops, may degrade loop tracking in the machine, or cause the loop edge to touch other process elements. Any of these attributes would make the photoreceptor unusable. One way to counter film curl is to coat a polymer layer on the back side to compensate for the stresses in the photoreceptor layers. The back-coat polymer must have adequate physical characteristics (friction, wear, etc.) to withstand cycling in the process. The extra coating adds expense but its addition is relatively commonplace. Any suitable polymer, such as a polycarbonate, may be used.

37.4.7 Overcoat Layer

Overcoat layers have received increased attention to lengthen the photoreceptor life by preventing wear, minimizing the effects of surface contacts, the effects of corona effluents (ozone and nitrogen oxides), and liquid attack in liquid-toner-based printers. As with the charge-blocking layer, the simplest approach is to coat a thin insulating polymer onto the photoreceptor surface. The thickness cannot be more than a few μm if one wants to avoid large residual potentials. Several approaches have been developed. One approach is to coat a second CTL with a changed composition to maximize polymer toughness (often by lowering the concentration of transport material or adding inorganic filler particles). Other approaches are to overcoat the photoreceptor with a silsesquioxane sol-gel several μm thick [37.198, 199], or with a submicron refractory layer of diamond-like carbon [37.200] or AlZrN [37.201]. However, the thinness of the refractory materials limits their utility for scratch and wear resistance. Furthermore, depending on their chemical make up they may have sensitivity to corona gasses.

Silsesquioxane overcoats applied in a sol–gel process have received attention [37.202, 203]. A commercial application utilized a sol–gel to overcoat a photoconductor with an uppermost CGL (positive-charging OPC). Because the CGL is sensitive to wear and corona chemicals an overcoat was essential to its implementation in a commercial photoreceptor. The sol is prepared and subsequently hopper-coated and cured at elevated temperature to effect crosslinking. To avoid a residual potential some conductivity is necessary but this cannot be so high that the latent image is degraded particularly under high humidity conditions [37.204]. The conductivity can be controlled with an ionic dopant [37.202] or by the incorporation of a hole-transporting functionality into the sol-gel matrix [37.203, 205].

A polyurethane overcoat in a positive-charging dual-layer OPC has been described for use in a liquid-toner-based printer [37.206]. The CGL was based on TiOPc pigment and, to avoid disruption of the CGL during the overcoat coating operation, an aqueous polyurethane dispersion was used.

A tough, durable, overcoat layer is described in a Konica Minolta Research Report [37.207]. The overcoat is a curable resin chemically bonded to inorganic nanoparticle filler. After one million prints at high

humidity in an electrophotographic printer the OPC thickness decreased by 1.5 μm, but the printed images showed no degredation in quality relative to a new OPC.

37.5 Photoreceptor Fabrication

As discussed above the two most common photoreceptor configurations are as a loop or a metal drum. After coating the layers of a drum photoreceptor it is ready to use in the process. Generally the drums are inspected for visible defects and samples are taken for process testing and quality control. Since the substrate is conductive the ground contact is easily made by means of a conductive brush contacting the edge or inside of the drum when it is mounted in the machine. On the other hand, after the appropriate layers are coated on a web several more steps are required for fabrication of the photoreceptor. The web is rolled onto a spool at the end of the coating process and subsequently slit to the desired width and rewound. The coated roll is then inspected for visible defects and defective areas are marked for removal. Defective areas are removed when the photoreceptor is cut to the desired length. The cut sheets are then ultrasonically welded to produce a photoconductor loop. Because the conductive layer is buried between the substrate and the coated layers, making a ground contact is not trivial. It is possible to leave one edge of the ground layer uncoated but, because the thin metal would be readily damaged or worn away with use, one typically finds a conductive stripe coated on one edge of the photoreceptor. The only requirement is that the conductive stripe penetrate the coated layers and contact the metal. The ground contact in the machine is often made with a conductive brush contacting the conductive stripe. The printing machine must keep track of the splice for several reasons: cleaning blade, degraded photoconductivity due to the ultrasonic welding, etc. This is sometimes accomplished by applying marks to the back of the photoreceptor, which are detected by the machine.

Fabricating a photoreceptor loop entails several operations not required in the fabrication of drum photoreceptors (slitting, cutting, conductive striping, welding, backside marking, etc.). Web-coating a photoreceptor is an expensive process but many thousands of feet of photoreceptor can be produced in a single coating event. Furthermore all the loop-processing elements are relatively straightforward and inexpensive. The key factor is that web-coating is a continuous process while drum-coating is a batch process. Photoreceptor loops are generally about ten times less costly to produce than drums.

There are advantages and disadvantages to both seamed loop and drum configurations in a printing machine. Loops can be made small for desktop printers or large for high-speed commercial printers. A large web (current commercial webs are as large as ≈ 10 ft in circumference) provides more surface area for imaging (several images per loop cycle) and more space for the various electrophotographic process elements. In addition, exposures through a semitransparent conductive layer are possible. Disadvantages are the presence of a seam which must be tracked and makes some processes, such as cleaning with a blade, problematic. Also, the tendency of a web to flutter or move necessitates both tracking and positioning to achieve adequate print registration. Metal drums are seamless and rigid, making it easier to control drum position. However, drums typically have a diameter of 35–300 mm and have limited imaging area and space for process elements. All things being equal a loop which produces several images per cycle will produce more images over its lifetime than a drum. Ultimately, the choice of loop or drum is a printing system issue.

37.6 Summary

Organic photoreceptors have been utilized in electrophotography since 1970. In this chapter we have briefly reviewed the development of organic photoreceptors from the standpoint of the device architecture, materials formulation, and electrical characteristics with emphasis on how the materials used impact on specific device functions.

The use of organic materials has enabled the development of electrophotographic technologies into a multi-billion-dollar business. Today's organic photoreceptor is a large-area photoelectric device capable of very high spatial resolution in the formation of a large-area image, with pixel resolution as small as 20 μm (1200 dpi), to meet demanding image-quality re-

quirements. It is truly an amazing success story. Many of the principles and indeed many of the same materials are being applied to today's highly publicized new technologies including organic light-emitting diodes (OLEDs), organic photovoltaic devices (OPV), and organic transistors. In this chapter we have described the device requirements and shown how organic photoreceptor development has progressed through a combination of architecture manipulation and materials design.

Organic photoreceptors have to satisfy many diverse requirements to be successfully employed in electrophotography as it is practiced today. These requirements have been met in large part by the principles of: (1) functional separation, and (2) materials design for the enhancement of specific electrical and mechanical characteristics. The concept of functional separation was key to organic photoreceptor development. Rather than try to produce a device with all the desirable characteristics in a single layer it is more productive to have a multilayer architecture with each layer being optimized for a specific characteristic. Not surprisingly these layers are commonly named after those characteristics which are paramount: the charge-generation layer (CGL), the charge-transport layer (CTL), the charge-blocking layer, the protective overcoat layer, etc. In this chapter we have described the characteristics of each of these layers and demonstrated how materials have been chosen and new materials synthesized to optimize their characteristics.

It is worthwhile to take a minute to review the electrophotographic process and consider the characteristics which have been built into the device:

- Spatial uniformity. The thickness of each of the coated layers (each varies from submicron to a few tens of μm) and the total thickness (controlled mainly by the CTL) must be very uniform so that the device capacitance (a major factor in determining surface charging characteristics) is uniform (on the order of a few percent variability). The CGL generation characteristics must be uniform, which means that the optical density of the CGL at the exposure wavelength must be uniform, to produce a uniform surface potential after exposure (on the order of a few percent variability). Thickness must be uniform for the smallest pixel (20 μm for 1200 dpi imaging) over the entire area of the device ($> 1000\,cm^2$). The ability to coat large areas of such thin layers to this degree of uniformity is a great technological achievement.
- Chemical resistance. Photoreceptors must resist chemical attack. Because corona charging is the most common method of applying a surface potential, the photoreceptor must be stable to the acid and oxidizing species produced in the corona. These chemicals (ozone, NO_x, and HNO_x) are highly reactive and the layer that is exposed (usually the CTL) must not be degraded. This is achieved by judicious choice of materials, the addition of stabilizers (antioxidants) to the CTL, or a protective overcoat.
- Electrical uniformity and stability. The photoreceptor dark electrical and photoelectrical characteristics (surface potentials with no exposure, imaging exposure, and erase exposure) must be very uniform and stable to prolonged electrophotographic cycling. The initial (dark) and exposed potentials must be the same for each pixel over the entire device area and must remain stable with cycling. If these characteristics drift uniformly with cycling, machine process control must be implemented to maintain image quality. If these characteristics drift nonuniformly the photoreceptor must be discarded when image artifacts (ghosting) appear.
- Photosensitivity. The sensitivity of the photoreceptor to light, often expressed as the energy needed to produce a defined decrease in surface potential, must be mated with the exposure used in the process in terms of exposure energy and wavelength. This has been achieved by the synthesis and pigment form manipulation of charge-generation materials. Today's high-sensitivity photoreceptors can approach a quantum efficiency of unity for charge generation (at a typical applied field of $\approx 20\,V/cm$). But, most importantly photoreceptors with lower sensitivity can be fabricated depending on the needs of the imaging system.
- Photodischarge rate (switching time). The time it takes for photogenerated carriers to transit the device limits the printing process speed. With the synthesis of highly stable hole-transport materials and the principle of molecularly doped polymers it has been possible to prepare devices with hole transit times that permit printing speeds approaching 200 pages/min.
- Mechanical strength. The surface of the photoreceptor that is exposed to the abrasive process elements (often the development and cleaning systems), must not wear too rapidly or unevenly and must be stable to crazing or other forms of damage. Typically the CTL is the surface layer, and considering that it is typically a polymer with as much as 50% by weight of charge-transport molecules, it is remarkable that materials can be chosen such that significant wear occurs only after tens of thousands of cycles. Where warranted, very tough protective overcoats have been developed which have little effect on the overall OPC device performance.

Today multilayer organic photoreceptors that meet all of the requirements listed above are made commercially on a huge scale. This success is due in large measure to the skill of the scientists and engineers who design the materials and processes for photoreceptor fabrication, and who then integrate it into the electrophotographic printing system such that acceptable and consistent image quality is obtained for a photoreceptor life, which is routinely 200 000 imaging cycles or longer if protective overcoat layers are used. This is an amazing story, an organic photoreceptor device with *adjustable* sensitivity at any desired wavelength from the visible to the near-infrared, with stable dark and photoelectrical characteristics over many thousands of imaging cycles, such that the characteristics of a 40 μm pixel area are identical over an area which might be greater than 1000 cm^2.

Acknowledgments. Martin A. Abkowitz received his Ph.D. in Physics from Syracuse University in 1964. In 1965 he joined the Xerox Corp. Webster Research Center where he was a Principal Scientist until his retirement in 1969. After retirement he was a Visiting Scientist at the University of Rochester. He was a Fellow of the American Physical Society, had 174 publications including 35 U.S. patents, and made over 250 contributed and invited presentations at international conferences. Martin passed away on January 30, 2014. The scientific community has lost an outstanding physicist and I, his co-author, have lost a dear friend.

References

37.1 J. Mort, D.M. Pai: *Photoconductivity and Related Phenomena* (Elsevier, New York 1976)
37.2 N.V. Joshi: *Photoconductivity Art, Science, and Technology* (Marcel Dekker, New York 1990)
37.3 K.-Y. Law: Chem. Rev. **93**, 449 (1993)
37.4 D.S. Weiss, M. Abkowitz: Chem. Rev. **110**, 479 (2010)
37.5 P.M. Borsenberger, D.S. Weiss: *Organic Photoreceptors for Xerography* (Marcel Dekker, New York 1998)
37.6 P.M. Borsenberger, D.S. Weiss: Photoreceptors: Organic photoreceptors. In: *Handbook of Imaging Materials*, ed. by A.S. Diamond, D.S. Weiss (Marcel Dekker, New York 2002) p. 369
37.7 D.S. Weiss, J.R. Cowdery, R.H. Young: Electrophotography. In: *Electron Transfer in Chemistry*, ed. by V. Balzani (Wiley-VCH, Weinheim 2001) p. 379
37.8 R.M. Schaffert, C.D. Oughton: J. Opt. Soc. Am. **38**, 991 (1948)
37.9 J.H. Dessauer, G.R. Mott, H. Bogdonoff: Photogr. Eng. **6**, 250 (1955)
37.10 J. Mort, I. Chen: Appl. Solid State Sci. **5**, 69 (1975)
37.11 J. Mort: *The Anatomy of Xerography* (McFarland, Jefferson 1989)
37.12 C.F. Carlson: Electrophotography, US Patent 2 297 691 (1942)
37.13 I. Chen: J. Imaging Sci. **34**, 15 (1990)
37.14 I. Chen: Nature of latent images formed on single layer organic photoreceptors, Proc. IS&T NIP18: 2002 Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 2002) p. 404
37.15 S. Jeyadev, D.M. Pai: J. Imaging Sci. Technol. **40**, 327 (1996)
37.16 J. Yi, R.B. Wells: J. Imaging Sci. Technol. **50**, 193 (2006)
37.17 D.M. Pai, B.E. Springett: Rev. Mod. Phys. **65**, 163 (1993)
37.18 L.B. Schein: *Electrophotography and Development Physics* (Laplacian, Morgan Hill 1996)
37.19 R.M. Schaffert: *Electrophotography* (Focal, New York 1975)
37.20 M.E. Scharfe: *Electrophotography Principles and Optimization* (Wiley, New York 1984)
37.21 E.M. Williams: *The Physics and Technology of Xerographic Processes* (Wiley-Interscience, New York 1984)
37.22 R.G. Enck, G. Pfister: Amorphous chalcogenides. In: *Photoconductivity and Related Phenomena*, ed. by J. Mort, D.M. Pai (Elsevier, New York 1976) p. 215
37.23 S.O. Kasap: Photoreceptors: The chalcogenides. In: *Handbook of Imaging Materials*, ed. by A.S. Diamond, D.S. Weiss (Marcel Dekker, New York 2002) p. 329
37.24 J. Mort: Applications of amorphous silicon and related materials in electronic imaging. In: *Handbook of Imaging Materials*, ed. by A.S. Diamond, D.S. Weiss (Marcel Dekker, New York 2002) p. 629
37.25 R. Joslyn: Photoreceptors: Recent imaging applications for amorphous silicon. In: *Handbook of Imaging Materials*, ed. by A.S. Diamond, D.S. Weiss (Marcel Dekker, New York 2002) p. 425
37.26 A. Rose: *Photoconductivity and Related Processes* (Interscience, New York 1963)
37.27 J.C. Scott, G.S. Lo: Dark decay in organic photoconductors, Proc. 6th Int. Symp. Adv. Non-Impact Print. Technol., ed. by R.J. Nash (Society Imaging Science and Technology, Springfield 1991) p. 403
37.28 E. Montrimas, T. Lozovski, J. Sidaravicius, Z. Tokarski: J. Imaging Sci. Technol. **49**, 326 (2005)
37.29 D.S. Weiss, B.R. Benwood, D.L. Troendle: J. Imaging Sci. Technol. **51**, 520 (2007)
37.30 K. Seino, S. Iino, H. Hirahara, T. Konuma, I. Yoshida, S. Kaieda: J. Imaging Sci. Technol. **56**, 020503 (2012)
37.31 R.H. Bube: *Photoelectronic Properties of Semiconductors* (Cambridge Univ. Press, Cambridge 1992)
37.32 S.M. Sze: *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York 1981)

37.33 M.A. Abkowitz, M. Stolka: Electronic transport in polymeric photoreceptors: PVK in polysilylenes, Proc. Int. Symp. Polym. Adv. Technol., ed. by M. Lewin (VCH, Weinheim 1988) p. 225
37.34 D.M. Pai, J.F. Yanus, M. Stolka: J. Phys. Chem. **88**, 4714 (1984)
37.35 M. Stolka, J.F. Yanus, D.M. Pai: J. Phys. Chem. **88**, 4707 (1984)
37.36 M. Stolka, M.A. Abkowitz: Mater. Res. Soc. Symp. **277**, 3 (1992)
37.37 F.K. Dolezalek: Experimental techniques. In: *Photoconductivity and Related Phenomena*, ed. by J. Mort, D.M. Pai (Elsevier, New York 1976) p. 27
37.38 A.R. Melnyk, D.M. Pai: Photoconductivity measurements. In: *Physical Methods in Chemistry*, ed. by B.W. Rossiter, J.F. Hamilton, R.C. Baetzold (Wiley, New York 1992) p. 321
37.39 I. Chen: J. Appl. Phys. **43**, 1137 (1972)
37.40 I. Chen, J. Mort: J. Appl. Phys. **43**, 1164 (1972)
37.41 M. Abkowitz: J. Reinf. Plast. Compos. **16**, 1303 (1997)
37.42 M.A. Abkowitz, M. Stolka, M. Morgan: J. Appl. Phys. **52**, 3453 (1981)
37.43 H. Bässler: Phys. Status Solidi (b) **107**, 9 (1984)
37.44 M. Stolka, M.A. Abkowitz: TBD a contribution to the success of organic materials. In: *Practical Applications of Organic Electronic Materials*, ed. by M. Iwamoto, I.S. Pu, S. Taniguchi (Science Forum, Tokyo 1994) p. 265
37.45 C.W. Tang, S.A. Van Slyke: Appl. Phys. Lett. **51**, 913 (1987)
37.46 M.A. Abkowitz, H. Bässler, M. Stolka: Philos. Mag. B **63**, 201 (1991)
37.47 W.D. Gill: J. Appl. Phys. **43**, 5033 (1972)
37.48 M.A. Abkowitz, J.S. Facci, W.W. Limburg, J. Janus: Phys. Rev. B Cond. Matter **46**, 6705 (1992)
37.49 R.G. Kepler, J.M. Zeigler, L.A. Harrah, S.R. Kurtz: Phys. Rev. B **35**, 2818 (1987)
37.50 M.A. Abkowitz, M. Stolka: Synth. Met. **50**, 395 (1992)
37.51 D. Hertel, H. Baessler, U. Scherf, H.H. Horhold: J. Chem. Phys. **110**, 9214 (1999)
37.52 C. Im, H. Baessler, H. Rost, H.H. Horhold: J. Chem. Phys. **113**, 3802 (2000)
37.53 J.S. Facci, M.A. Abkowitz, W.W. Limburg, D. Renfer, J. Yanus: Mol. Cryst. Liq. Cryst. **194**, 55 (1991)
37.54 J. Michl, J.W. Downing, T. Karatsu, A.J. McKinley, G. Poggi, G.M. Wallraff, R. Sooriyakumaran: Pure Appl. Chem. **60**, 959 (1988)
37.55 R.D. Miller, J. Michl: Chem. Rev. **89**, 1359 (1989)
37.56 M.A. Abkowitz, M. Stolka: Polym. Prepr. (ACS Div. Polym. Chem.) **31**, 254 (1990)
37.57 R.D. Miller, J.R. Rabolt, R. Sooriyakumaran, G.N. Fickes, B.L. Farmer, H. Kuzmany: ACS Polym. Prepr. (ACS Div. Polym. Chem.) **28**, 422 (1987)
37.58 M.A. Abkowitz, M. Stolka: Philos. Mag. Lett. **58**, 239 (1988)
37.59 D. Emin: Adv. Phys. **24**, 305 (1975)
37.60 L.B. Schein, A. Rosenberg, S.L. Rice: J. Appl. Phys. **60**, 4287 (1986)
37.61 L.B. Schein: Mol. Cryst. Liq. Cryst. **183**, 41 (1990)
37.62 H. Bässler: Phys. Status Solidi **175**, 15 (1993)
37.63 H. Baessler, G. Schoenherr, M.A. Abkowitz, D.M. Pai: Phys. Rev. B Cond. Matter **26**, 3105 (1982)
37.64 P.E. Parris, V.M. Kenkre, D.H. Dunlap: Phys. Rev. Lett. **87**, 126601 (2001)
37.65 D.E. Dunlap, P.E. Parris, V.M. Kenkre: Proc. SPIE **3799**, 88 (1999)
37.66 L.B. Schein: Philos. Mag. B **65**, 795 (1992)
37.67 S.V. Novikov, A.V. Vannikov: J. Phys. Chem. **99**, 14573 (1995)
37.68 S.V. Novikov, D.H. Dunlap, V.M. Kenkre: Proc. SPIE **3471**, 181 (1998)
37.69 D.H. Dunlap: Phys. Rev. B **52**, 939 (1995)
37.70 D.H. Dunlap, P.E. Parris, V.M. Kenkre: Phys. Rev. Lett. **77**, 542 (1996)
37.71 S.V. Novikov, D.H. Dunlap, V.M. Kenkre, P.E. Parris, A.V. Vannikov: Phys. Rev. Lett. **81**, 4472 (1998)
37.72 D. Hertel, H. Bässler: Chem. Phys. Chem. **9**, 666 (2008)
37.73 S.V. Novikov, A.V. Vannikov: J. Phys. Chem. C **113**, 2532 (2009)
37.74 S.D. Baranovskii: Phys. Status Solidi (b) **251**, 487 (2014)
37.75 M.A. Abkowitz: Philos. Mag. B **65**, 817 (1992)
37.76 M. Stolka, M.A. Abkowitz: Synth. Met. **54**, 417 (1993)
37.77 H.J. Yuh, D. Abramsohn, M. Stolka: Philos. Mag. Lett. **55**, 277 (1987)
37.78 F.W. Schmidlin: Phys. Rev. B **16**, 2362 (1977)
37.79 M.A. Abkowitz, R.C. Enck: J. Phys. Colloq. C **4**(1), 443 (1981)
37.80 M.A. Abkowitz: Characterization of metal interfaces to molecular media from analysis of transient and steady state electrical measurements. In: *Conjugated Polymer and Molecular Interfaces*, ed. by W.R. Salaneck, K. Seki, A. Kahn, J.-J. Pireaux (Marcel Dekker, New York 2002) p. 545
37.81 M.A. Abkowitz, J.S. Facci, M. Stolka: Chem. Phys. **177**, 783 (1993)
37.82 M.A. Abkowitz, J.S. Facci, M. Stolka: Appl. Phys. Lett. **63**, 1892 (1993)
37.83 M.A. Lampert, P. Mark: *Current Injection in Solids* (Academic, New York 1970) p. 17
37.84 P.N. Murgatroyd: J. Phys. D. **3**, 151 (1970)
37.85 R.H. Young: Philos. Mag. Lett. **70**, 331 (1994)
37.86 M. Pope, C.E. Swenberg: *Electronic Processes in Organic Crystals* (Clarendon, Oxford 1982)
37.87 Y. Hayashi, M. Kuroda, A. Inami: Bull. Chem. Soc. Jpn. **39**, 1660 (1966)
37.88 H. Hoegl: J. Phys. Chem. **69**, 755 (1965)
37.89 P.J. Regensburger: Photochem. Photobiol. **8**, 429 (1968)
37.90 H. Hoegl, H.G. Barchietto, D. Tar: Photochem. Photobiol. **16**, 335 (1972)
37.91 P.J. Melz: J. Chem. Phys. **57**, 1694 (1972)
37.92 M. Yokoyama, Y. Endo, H. Mikawa: Bull. Chem. Soc. Jpn. **49**, 1538 (1976)
37.93 M. Yokoyama, Y. Endo, A. Matsubara, H. Mikawa: J. Chem. Phys. **75**, 3006 (1981)
37.94 M. Yokoyama, S. Shimokihara, A. Matsubara, H. Mikawa: J. Chem. Phys. **76**, 724 (1982)
37.95 M. Yokoyama, H. Mikawa: Photogr. Sci. Eng. **26**, 143 (1982)

37.96 M.B. O'Regan, P.M. Borsenberger, E.H. Magin, T. Zubil: J. Imaging Sci. Technol. **40**, 1 (1996)
37.97 Z.D. Popovic: Carrier generation mechanisms in organic photoreceptors, Proc. 9th Int. Congr. Adv. Non-Impact Print. Technol. (Society Imaging Science and Technology, Springfield 1993) p. 591
37.98 L. Onsager: Phys. Rev. **54**, 554 (1938)
37.99 P.M. Borsenberger, A.I. Ateya: J. Appl. Phys. **50**, 909 (1979)
37.100 L. Onsager: J. Chem. Phys. **2**, 599 (1934)
37.101 C. Braun: J. Chem. Phys. **80**, 4157 (1984)
37.102 S.N. Smirnov, C. Braun: J. Imaging Sci. Technol. **43**, 425 (1999)
37.103 D.M. Pai: Geminate recombination in some amorphous materials. In: *Physics of Disordered Materials*, ed. by D. Adler, H. Fritzsche, S.R. Ovshinsky (Plenum, New York 1985) p. 579
37.104 J. Noolandi, K.M. Hong: J. Chem. Phys. **70**, 3230 (1979)
37.105 T. Niimi, M. Umeda: J. Appl. Phys. **74**, 465 (1993)
37.106 Z.D. Popovic: J. Chem. Phys. **76**, 2714 (1982)
37.107 Z.D. Popovic: Chem. Phys. **86**, 311 (1984)
37.108 R.C. Hughes: J. Chem. Phys. **58**, 2212 (1973)
37.109 J.W. Kerr, G.H.S. Rokos: J. Phys. D Appl. Phys. **10**, 1151 (1977)
37.110 I. Chen: J. Appl. Phys. **49**, 1162 (1978)
37.111 W. Mey, E.I.P. Walker, D.C. Hoesterey: J. Appl. Phys. **50**, 8090 (1979)
37.112 R.H. Young: J. Appl. Phys. **60**, 272 (1986)
37.113 J.-Y. Moisan, B. André, R. Lever: Chem. Phys. **153**, 305 (1991)
37.114 A.V. Buettner, W. Mey: Photogr. Sci. Eng. **26**, 80 (1982)
37.115 K. Aizawa, M. Takeshima, H. Kawakami: A study of 1-dot latent image potential, Proc. NIP17: Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 2001) p. 572
37.116 T. Iwamatsu, T. Toyoshima, A. Nobuyuki, Y. Mutou, Y. Nakajima: A study of high resolution latent image forming and development, Proc. IS&T NIP 15: 1999 Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 1999) p. 732
37.117 D.H. Dunlap, L.B. Schein, A. Tyutnev, V. Saenko, E.D. Poshidaev, P.E. Parris, D.S. Weiss: J. Phys. Chem. C **114**, 9076 (2010)
37.118 D.S. Weiss, A.P. Tyutnev, E.D. Pozhidaev: J. Phys. Chem. C **119**, 5334 (2015)
37.119 I. Chen: J. Imaging Sci. Technol. **37**, 396 (1993)
37.120 M.F. Molaire, W.T. Ferrar, R.G. Galipo, M. Sykes, T. Zubil: Cross-linked naphthalene bisimide „E-barrier" layer compositions using blocked isocyanate chemistry, Proc. NIP24: Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 2008) p. 184
37.121 W.T. Ferrar, D.S. Weiss, M.F. Molaire, X. Jin, L.J. Sorriero: Polym. Prepr. (ACS Div. Polym. Chem.) **50**, 518 (2009)
37.122 Y.I. Seo, D.-G. Kim, Y.D. Kim, B.J. Lee, K.H. Lee: Scripta Materialia **59**, 889 (2008)
37.123 R.M. Schaffert: IBM J. Res. Dev. **15**, 75 (1971)
37.124 U. Vahtra, R.F. Wolter: IBM J. Res. Dev. **22**, 34 (1978)
37.125 J.M. Pearson: Pure Appl. Chem. **49**, 463 (1977)
37.126 R.C. Penwell, B.N. Ganguly, T.W. Smith: J. Polym. Sci. Macromol. Rev. **13**, 63 (1978)
37.127 W.D. Gill: Polymeric photoconductors. In: *Photoconductivity and Related Phenomena*, ed. by J. Mort, D.M. Pai (Elsevier, New York 1976) p. 303
37.128 M. Hatano, K. Tanikawa: Prog. Org. Coat. **6**, 65 (1978)
37.129 F. Li, Y. Li, Z. Guo, Y. Mo, L. Fan, F. Bai, D. Zhu: Solid State Commun. **107**, 189 (1998)
37.130 W.J. DuImage, W.A. Light, S.J. Marino, C.D. Salzberg, D.L. Smith, W.J. Staudenmayer: J. Appl. Phys. **49**, 5543 (1978)
37.131 P.M. Borsenberger, A. Chowdry, D.C. Hoesterey, W. Mey: J. Appl. Phys. **49**, 5555 (1978)
37.132 J.M. Perlstein: Structure and charge generation in low-dimensional organic molecular self-assemblies. In: *Electrical Properties of Polymers*, ed. by D.A. Seanor (Academic, New York 1982) p. 59
37.133 P.M. Borsenberger, D.C. Hoesterey: J. Appl. Phys. **51**, 4248 (1980)
37.134 D. McMurtry, M. Tinghitella, R. Svendsen: IBM J. Res. Dev. **28**, 257 (1984)
37.135 J. Pacansky, R.J. Waltman: J. Am. Chem. Soc. **114**, 5813 (1992)
37.136 G. DiPaola-Baranyi, C.K. Hsiao, A.M. Hor: J. Imaging Sci. **34**, 224 (1990)
37.137 K.-Y. Law, I.W. Tarnawskyj: J. Imaging Sci. Technol. **37**, 22 (1993)
37.138 M. Umeda, T. Shimada, T. Aruga, T. Niimi, M. Sasaki: J. Phys. Chem. **97**, 8531 (1993)
37.139 K.-Y. Law, I.W. Tarnawskyj, Z.D. Popovic: J. Imaging Sci. Technol. **38**, 118 (1994)
37.140 M. Umeda: J. Imaging Sci. Technol. **43**, 254 (1999)
37.141 M. Umeda, T. Niimi, M. Hashimoto: Jpn. J. Appl. Phys. **29**, 2746 (1990)
37.142 T. Niimi, U. Umeda: J. Appl. Phys. **76**, 1269 (1994)
37.143 K. Takeshita, Y. Sasaki, T. Shoda, T. Murayama: Time-resolved absorption study on the photocarrier generation process in layered organic photoreceptors: A role of delocalized holes in photocarrier generation, Proc. IS&T NIP19: 2003 Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 2003) p. 683
37.144 P.J. Melz, R.B. Champ, L.S. Chang, C. Chiou, G.S. Keller, L.C. Liclican, R.R. Neiman, M.D. Shattuck, W.J. Weiche: Photogr. Sci. Eng. **21**, 73 (1977)
37.145 T. Murayama: The design of high performance organic photoconductors, Proc. IS&T NIP17: Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 2001) p. 557
37.146 M. Umeda, M. Hashimoto: J. Appl. Phys. **72**, 117 (1992)
37.147 E.H. Magin, P.M. Borsenberger: J. Appl. Phys. **73**, 787 (1993)
37.148 Z.D. Popovic, R. Cowdery, I.M. Khan, A.-M. Hor, J. Goodman: J. Imaging Sci. Technol. **43**, 266 (1999)
37.149 R.O. Loutfy, A.M. Hor, P. Kazmaier: Properties and application of organic photoconductive materials: Molecular design of organic photoconductive polycyclic aromatic diimides compounds, Proc.

32nd Symp. Macromol., ed. by T. Saegusa, T. Higashimura, A. Abe (Blackwell Scientific, Oxford 1988) p. 437
37.150 R.O. Loutfy, A.M. Hor, P. Kazmaier, M. Tam: J. Imaging Sci. **33**, 151 (1989)
37.151 Z.D. Popovic, A.-M. Hor, R.O. Loutfy: Chem. Phys. **127**, 451 (1988)
37.152 J.M. Duff, G. Allen, A.-M. Hor, S. Gardner: Synthesis, spectroscopy and photoconductivity of dimetric perylene bisimide pigments, Proc. IS&T NIP15: 1999 Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 1999) p. 655
37.153 A. Kakuta, Y. Mori, S. Takano, M. Sawada, I. Shibuya: J. Imaging Sci. Technol. **11**, 7 (1985)
37.154 A. Shimada, M. Anzai, A. Kakuta, T. Kawanishi: IEEE Trans. Ind. Appl. **1A23**, 804 (1987)
37.155 Y. Kanemitsu, S. Imamura: J. Appl. Phys. **67**, 3728 (1990)
37.156 T. Enokida, R. Hirohashi, S. Mizukami: J. Imaging Sci. **35**, 235 (1991)
37.157 Y. Fujimaki, H. Tadokoro, Y. Oda, H. Yoshioka, T. Homma, H. Moriguchi, K. Watanabe, A. Konishita, N. Hirose, A. Itami, S. Ikeuchi: J. Imaging Technol. **17**, 202 (1991)
37.158 K. Daimon, K. Nukada, Y. Sakaguchi, R. Igarashi: J. Imaging Sci. Technol. **40**, 249 (1996)
37.159 T. Enokida, R. Hirohashi, T. Nakamura: J. Imaging Sci. **34**, 234 (1990)
37.160 S. Takano, Y. Mimura, N. Matsui, K. Utsugi, T. Gotoh, C. Tani, K. Tateishi, N. Ohde: J. Imaging Sci. Technol. **17**, 46 (1991)
37.161 T.I. Martin, J.D. Mayo, C.A. Jennings, S. Gardner, C.K. Hsiao: Solvent induced transformations of the polymorphs of oxytitanium phthalocyanine (TiOPc), Proc. IS&T 11th Int. Congr. Adv. Non-Impact Print. Technol. (Society Imaging Science and Technology, Springfield 1995) p. 30
37.162 D. Liang, W. Peng, Y. Wang: Adv. Mater. **24**, 5249 (2012)
37.163 Y. Kanemitsu, A. Yamamoto, H. Funada, Y. Masumoto: J. Appl. Phys. **69**, 7333 (1991)
37.164 Z.D. Popovic, M.I. Khan, S.J. Atherton, A.-M. Hor, J.L. Goodman: Time-resolved fluorescence quenching and carrier generation in titanyl phthalocyanine (TiOPc)-humidity effects. In: *Electrical and Related Properties of Organic Solids*, ed. by R.W. Munn, A. Miniewicz, B. Kuchta (Kluwer Academic, Dordrecht 1997) p. 207
37.165 Z.D. Popovic, M.I. Khan, A.-M. Hor, J.L. Goodman, J.F. Graham: Study of the photoconductivity mechanism in phthalocyanine pigment particles by electric field modulated time resolved fluorescence, Proc. IS&Ts NIP19: 2003 Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 2003) p. 687
37.166 R.E. Wingard: IEEE Ind. Appl. **37**, 1251 (1982)
37.167 R.B. Champ: Proc. SPIE **759**, 40 (1987)
37.168 R.O. Loutfy, C.K. Hsiao, P.M. Kazmaier: Photogr. Sci. Eng. **27**, 5 (1983)
37.169 K.-Y. Law: J. Imaging Sci. **31**, 83 (1987)
37.170 K.-Y. Law, F.C. Bailey: J. Imaging Sci. **31**, 172 (1987)
37.171 R.O. Loutfy, A.-M. Hor, C.-K. Hsiao, G. Baranyi, P. Kazmaier: Pure Appl. Chem. **60**, 1047 (1988)
37.172 P.M. Kazmaier, R. Burt, G. Di-Paola-Baranyi, C.-K. Hsiao, R.O. Loutfy, T.I. Martin, G.K. Hamer, T.L. Bluhm, M.G. Taylor: J. Imaging Sci. **32**, 1 (1988)
37.173 G. DiPaola-Baranyi, C.K. Hsiao, P.M. Kazmaier, R. Burt, R.O. Loutfy, T.I. Martin: J. Imaging Sci. **32**, 60 (1988)
37.174 K.-Y. Law: Chem. Mater. **4**, 605 (1992)
37.175 K. Sakanoue, M. Motoda, M. Sugimoto, S. Sakaki: J. Phys. Chem. **103**, 5551 (1999)
37.176 A. Matoliukstyte, E. Burbulis, J.V. Grazulevicius, V. Gaidelis, V. Jankauskas: Synth. Met. **158**, 462 (2008)
37.177 D. Velasco, V. Jankauskas, J. Stumbraite, J.V. Grazulevicius, V. Getautis: Synth. Met. **159**, 654 (2009)
37.178 V. Getautis, J. Stumbraite, V. Gaidelis, V. Jankauskas, A. Kliucius, V. Paulauskas: Synth. Met. **157**, 35 (2007)
37.179 P.M. Borsenberger, E.H. Magin, M. Van der Auweraer, F.C. De Schryver: Phys. Status Solidi (a) **140**, 9 (1993)
37.180 P.M. Borsenberger, R. Richert, H. Bässler: Phys. Rev. B **47**, 4289 (1993)
37.181 M. Van der Auweraer, F.C. De Schryver, P.M. Borsenberger, H. Bässler: Adv. Mater. **6**, 199 (1994)
37.182 P.M. Borsenberger, H. Bässler: Phys. Status Solidi (b) **170**, 291 (1992)
37.183 A. Dieckmann, H. Bässler, P.M. Borsenberger: J. Chem. Phys. **99**, 8136 (1993)
37.184 P.M. Borsenberger, J.J. Fitzgerald: J. Phys. Chem. **97**, 4815 (1993)
37.185 R.H. Young: Philos. Mag. B **72**, 435 (1995)
37.186 A. Fujii, T. Shoda, S. Aramaki, T. Murayama: J. Imaging Sci. Technol. **43**, 430 (1999)
37.187 T. Shoda, T. Murayama, A. Fujii: Relationship between molecular properties of hole transport molecules and field dependence of hole mobility, Proc. NIP17: Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 2001) p. 550
37.188 K. Nauka, S. Chang, H.T. Ng: J. Imaging Sci. Technol. **54**, 050304-1 (2010)
37.189 P.M. Borsenberger, W.T. Gruenbaum: J. Polym. Sci. Polym. Phys. **34**, 575 (1996)
37.190 P.M. Borsenberger, E.H. Magin, M.R. Detty: J. Imaging Sci. Technol. **39**, 12 (1994)
37.191 P.M. Borsenberger, H.-C. Kan, E.H. Magin, W.B. Vreeland: J. Imaging Sci. Technol. **39**, 6 (1995)
37.192 P.M. Borsenberger, W.T. Gruenbaum, M.B. O'Regan, L.J. Rossi: J. Polym. Sci. Polym. Phys. **33**, 2143 (1995)
37.193 V. Getautis, J.V. Grazulevicius, M. Daskeviciene, T. Malinauskas, D. Jankunaite, V. Gaidelis, V. Jankauskas, J. Sidaravicius, Z. Tokarski: Polymer **46**, 7918 (2005)
37.194 V. Getautis, M. Daskeviciene, T. Malinauskas, V. Gaidelis, V. Jankauskas, Z. Tokarski: Synth. Met. **155**, 599 (2005)

37.195 V. Getautis, J.V. Grazulevicius, M. Daskeviciene, T. Malinauskas, V. Jankauskas, J. Sidaravicius, A. Undzenas: Eur. Polym. J. **43**, 3597 (2007)
37.196 J.V. Grazulevicius: Polym. Adv. Technol. **17**, 694 (2006)
37.197 R. Laurinaviciute, J. Ostrauskate, J.V. Grazulevicius, V. Jankauskas: Des. Monomers Polym. **17**, 255 (2014)
37.198 W. Witt, C. Roth, O. Gelsen, R. Lischewski, H.-J. Humpert, M. Dohle: Silica sol/gel based nano structured hard coats for organic photoconductors, Proc. NIP23: Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 2007) p. 440
37.199 D.S. Weiss, W.T. Ferrar, J.R. Corvan, L.G. Parton, G. Miller: J. Imaging Sci. Technol. **43**, 280 (1999)
37.200 M.-S. Hwant, E.-T. Kim, J.-B. Kim, C. Lee: Diam. Relat. Mater. **10**, 2063 (2001)
37.201 Y.C. Chan, X.S. Miao, E.Y.B. Pun: J. Mater. Res. **13**, 2042 (1998)
37.202 W.T. Ferrar, D.S. Weiss, J.R. Cowdery-Corvan, L.G. Parton: J. Imaging Sci. Technol. **44**, 429 (1999)
37.203 X. Jin, D.S. Weiss, L.J. Sorriero, W.T. Ferrar: J. Imaging Sci. Technol. **47**, 361 (2003)
37.204 D.S. Weiss, J.R. Cowdery, W.T. Ferrar, R.H. Young: J. Imaging Sci. Technol. **40**, 322 (1996)
37.205 W.T. Ferrar, W.T. Gruenbaum, X. Jin, M.F. Molaire, J.R. Cowdery, D.S. Weiss: Silicon **1**, 13 (2009)
37.206 N.-J. Lee, H.-R. Joo, K.-Y. Yon, Y. No: Development of positive charging multi-layered organic photoconductor for liquid electrophotographic process, Proc. IS&T NIP19: 2003 Int. Conf. Digit. Print. Technol. (Society Imaging Science and Technology, Springfield 2003) p. 670
37.207 Konica Minolta Research Report: Highly Durable Photoconductor Drums Stress-Resistant Overcoat Layer (2006) http://www.konicaminolta.com/about/research/publishing/opc.html

38. Luminescent Materials

Andy Edgar

This chapter surveys the field of solid-state luminescent materials, beginning with a discussion of the different ways in which luminescence can be excited. The internal energy-level structures of luminescent ions and centres, particularly rare-earth ions, are then discussed before the effects of the vibrating host lattice are included. Having set the theoretical framework in place, the chapter then proceeds to discuss the specific excitation process for photo-stimulated luminescence and thermally stimulated luminescence before concluding by surveying current applications, including phosphors for compact fluorescent and LED lighting, long-term persistent phosphors, x-ray storage phosphors, and scintillators.

38.1 **Luminescent Centres** 999
38.1.1 Rare-Earth Ions 999
38.1.2 Transition-Metal Ions 1000
38.1.3 s^2 Ions 1001
38.1.4 Semiconductors 1001
38.2 **Interaction with the Lattice** 1001
38.3 **Thermally Stimulated Luminescence** 1003
38.4 **Optically (Photo-)Stimulated Luminescence** 1004
38.5 **Experimental Techniques – Photoluminescence** 1005
38.6 **Applications** 1006
38.6.1 Compact Fluorescent Lamps 1006
38.6.2 Phosphors for Solid State Lighting 1006
38.6.3 Long-Persistence Phosphors 1007
38.6.4 X-Ray Imaging 1008
38.6.5 Phosphors for Displays 1009
38.6.6 Scintillators 1010
38.7 **Representative Phosphors** 1011
References 1011

Luminescent materials are substances which convert an incident energy input into the emission of electromagnetic waves in the ultraviolet (UV), visible or infrared regions of the spectrum, over and above that due to black-body emission. A wide range of energy sources can stimulate luminescence, and their diversity provides a convenient classification scheme for luminescence phenomena, which is summarised in Table 38.1. Photoluminescence, where the luminescence is stimulated by UV or visible light, is a widely used materials science technique for characterising dopants and impurities, and finds applications in lighting technologies such as fluorescent and solid state lamps. Radioluminescence involves excitation by ionising radiation, and is used in scintillators for nuclear particle detection; the special case of stimulation by energetic electrons is called cathodoluminescence, the name arising from early atomic physics experiments involving gas discharges. Electroluminescence involves collisional excitation by internal electrons accelerated by an applied electric field, and with a much lower energy than in the case of cathodoluminescence. Electroluminescence finds applications in panel lighting used in some liquid-crystal display (LCD) back-planes, in inorganic light-emitting diodes (LEDs) and organic light-emitting diodes (OLEDs).

There are other forms of luminescence which we mention for completeness but will not discuss further: bioluminescence and chemiluminescence where the energy input is from chemical or biochemical reactions, sonoluminescence (sound wave excitation), and triboluminescence (strain or fracture excitation). Both organic and inorganic systems can display luminescence, but here we focus primarily on inorganic systems.

There are several books which describe the luminescence and spectroscopy of inorganic materials [38.1–4] and applications to phosphors and scintillators [38.5, 6] in more depth than is possible in a short article.

The terms phosphorescence and fluorescence are often used in connection with luminescent materials. This classification is based on the time-domain response of the luminescent system. Figure 38.1 shows a generic

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_38

Table 38.1 Types of luminescence

Designation	Excitation	Trigger	Acronym
Photo-luminescence	UV, visible photons	–	PL
Radio-luminescence	X-ray, gamma rays, charged particles	–	RL
Cathodo-luminescence	Energetic electrons	–	CL
Electro-luminescence	Electric field	–	EL
Thermo-luminescence	Photons, charged particles	Heat	TSL
Optically/photo-stimulated luminescence	Photons, charged particles	Visible/IR photons	OSL, PSL

photoluminescent system where incident UV radiation excites a system from a ground state 1 with probability per unit time P_{12} into an excited state 2. The system decays with probability P_{23} to a luminescing level 3, from which there is a probability P_{31R} and P_{31NR} of radiative and nonradiative decay, respectively, to the ground state. Nonradiative decay generally involves phonon emission.

First suppose that the transition probabilities are such that $P_{12}, P_{21} \gg P_{31}(= P_{31R} + P_{31NR})$. If the optical pumping (excitation) is abruptly stopped, the population of the luminescing state 3 decays as,

$$N_3 = N_3(0)\exp(-P_{31}t) \tag{38.1}$$

and the rate of luminescent energy emission is $-(h\nu P_{31R}/P_{31})\mathrm{d}N_3/\mathrm{d}t$ for an energy difference $E_{31} = h\nu$ between states 1 and 3. Hence the luminescence intensity at distance r from the sample is,

$$I = (h\nu)N_3(0)P_{31R}\frac{\exp(-P_{31}t)}{4\pi r^2} \tag{38.2}$$

and the characteristic luminescence lifetime is $\tau = (P_{31})^{-1}$. Thus the lifetime is governed by both radiative and nonradiative processes, whilst the intensity of the luminescence depends on the relative magnitude of P_{31R}.

This discussion provides the basis for understanding the terms fluorescence and phosphorescence applied to luminescent materials. A material is often classified as one or the other according to the relative magnitude of $\tau = (P_{31})^{-1}$, with 10 ns being set in a relatively arbitrary way as the boundary between a *fast* fluorescent system and a *slow* phosphorescent one. For comparison, the theoretical lifetime for spontaneous emission for a strongly allowed hydrogen atom $2\mathrm{p} \rightarrow 1\mathrm{s}$ transition is about 0.2 ns.

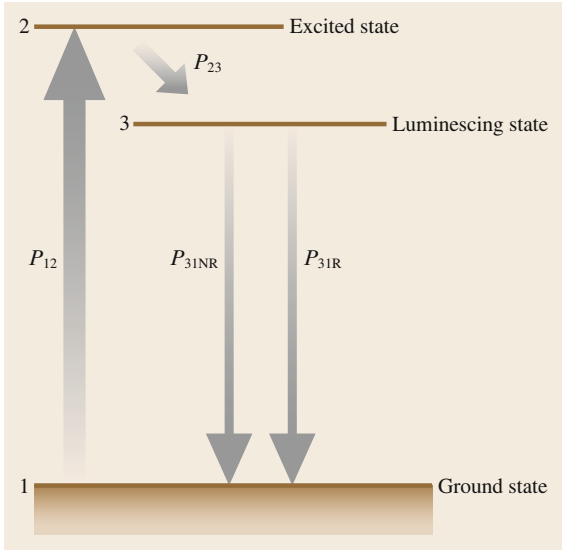


Fig. 38.1 Optical pumping cycle for a generic photoluminescent system

However, by this definition phosphorescence can also arise from luminescent states with short lifetimes which are populated through ones with long lifetimes. In Fig. 38.1, if state 2 is long-lived in the sense that $P_{23} \ll P_{31}$ then the measured lifetime for emission from the luminescing state will be $\tau = (P_{23})^{-1}$, and the system will be labelled phosphorescent, even though the luminescing level itself has a very short lifetime. In consequence, a second classification [38.2] is based on whether or not the luminescing level is fed by a metastable state which sets the lifetime. Sometimes the metastable state is a long-lived intermediate form of energy storage which can be triggered by an external stimulus to undergo a transition to a fluorescent level. Thus in Table 38.1, thermally stimulated and optically stimulated luminescence involve a metastable state consisting of trapped electrons and holes, which can be triggered to recombine by heating or by optical stimulation; the recombination energy is transferred to a fluorescing centre.

Overall, the fluorescence/phosphorescence classification is somewhat nebulous, and it is debateable whether the classification is necessary or desirable.

38.1 Luminescent Centres

A wide variety of centres give rise to luminescence in semiconductors and insulating materials, including rare-earth ions, transition-metal ions, excitons, donor–acceptor pairs, and ions with a d^{10} or s^2 electronic configuration ground state. Organic molecules also display luminescence on two different time scales, fast fluorescence associated with spin-allowed singlet-to-singlet transitions, and slow phosphorescence from triplet-to-singlet transitions. Some luminescence spectra consist of broad emission bands arising from the interaction between the electronic system of the luminescent centre and the vibrations of the atoms or ions, which surround it; the broad bands arise from simultaneous transitions of both electronic and vibrational systems. For others, such as the rare earths, the spectra comprise sharp lines arising from purely electronic transitions, and the effect of the environment is felt mainly through their effects on the lifetimes of the states. Thus in discussing the physical background to luminescence, it is simplest to start with a discussion of rare-earth luminescence, as an example where the effect of vibrations can be initially ignored.

38.1.1 Rare-Earth Ions

The trivalent rare-earth ions have n electrons ($n = 1-14$) in the 4f shell. In a free ion, the eigenstates resulting from the various atomic interactions are labelled by the total spin $\boldsymbol{S}$ and orbital angular momenta $\boldsymbol{L}$. Spin–orbit coupling breaks up each L, S multiplet of degeneracy $(2S+1)(2L+1)$ into sub-multiplets labelled by the total angular momentum $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$, where J can range from $L-S$ to $L+S$. The $4f^n$ orbitals lie within the outer $5s^2$ and $5p^6$ filled shells, which partly shield them from the effects of a crystalline environment. The effects of the latter are quantitatively described by the *crystal field* [38.3, 7], and this term in the Hamiltonian splits the J multiplets into $2J+1$ sublevels, the so-called crystal-field splitting. Some of these crystal-field levels may still be two- or threefold degenerate, depending upon the symmetry of the environment. Odd-electron systems always have at least twofold (Kramers) degeneracy. The resulting energy-level structures are complicated, and are summarised in the classic *Dieke diagram* [38.7]. The original has been updated by *Carnall* et al. [38.8] and is reproduced in many books and papers on rare-earth ion spectroscopy [38.9–11] which should be consulted for a more extensive presentation of the subject than is possible here. In Fig. 38.2, we show a schematic version (not accurately to scale) of the diagram appropriate for Pr^{3+} with $n = 2$, which serves for our discussion. The crystal-field splitting is usually smaller than the spin–orbit splitting, and is illustrated schematically by the vertical extent of the bands in Fig. 38.2. The multiplet labels follow the usual $^{2S+1}L_J$ scheme.

In the figure we also show a generic highest excited band which does not belong to the $4f^2$ configuration. In the specific case of the rare earths this can be the excited-state configuration resulting from promotion of one electron to the 5d state, giving an overall $4f^{n-1}5d^1$ configuration; more generally it could be a so-called charge transfer band, which corresponds to the transfer of one electron from the ligands to the luminescent ion. The relative location and importance of these bands varies with the luminescent ion and the crystalline environment, but they play an important role in the excitation of luminescence.

Excitation and luminescence transitions within the various levels in Fig. 38.2 are governed by the golden rule of quantum mechanics [38.12] for interactions with the electromagnetic field; in summary the probability of a transition between two states i and j is proportional to square of the matrix element $< i|H|j >$,

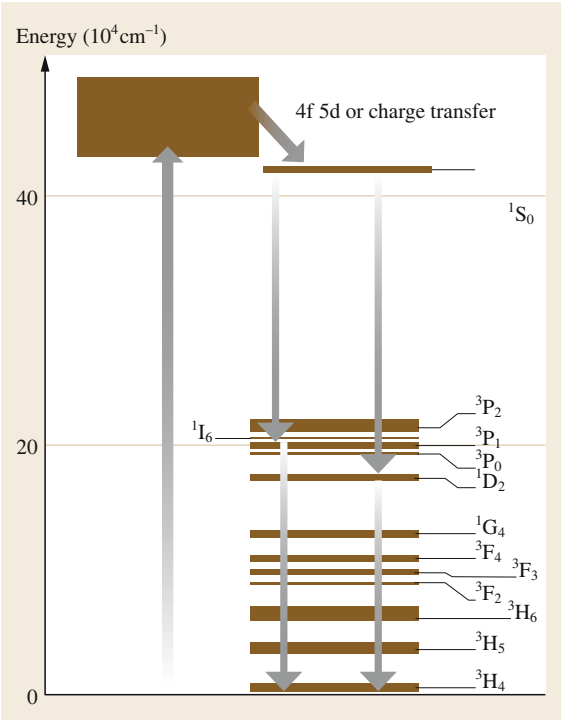


Fig. 38.2 Schematic energy-level diagram for Pr^{3+}

where H is the time-dependent perturbation Hamiltonian representing the interaction of the electrons with the electromagnetic field. The proportionality constant contains the light intensity in the case of excitation. The perturbation can be expanded in a power series involving electric and magnetic multipoles of the electronic system, but of these the electric dipole term is dominant, with the magnetic dipole term being much smaller by a factor of more than five orders of magnitude. Since the electric dipole operator $\boldsymbol{er}$ has odd parity, the matrix element for transitions $r_{ij} = < i|\boldsymbol{er}|j >$ is necessarily zero unless i and j have opposite parity. This is the most important selection rule governing luminescence: transitions between states of the same parity have zero transition probability and so are *forbidden* (Laporte's rule). In the case of the rare earths, all states of a single $4f^n$ configuration have the same parity, and so all optical transitions within the configuration are strictly forbidden. But this rule is relaxed by several considerations. First, if the crystalline environment lacks inversion symmetry, the crystal field admixes a small fraction of the excited configurations (e.g., $4f^15d^1$ for Pr^{3+}) of opposite parity into the ground configuration, which makes such transitions weakly allowed. Secondly, the selection rule for magnetic dipole transitions is that they are allowed between states of the same parity, although they are typically about five orders of magnitude weaker than for electric dipole ones. Finally, odd-parity vibrations and an electron–phonon interaction produce a similar configuration admixture effect to static lattice odd-parity mixing although this effect is more important for 3d ions than for 4f ones.

With only weak transitions possible within the 4f configuration, one might wonder how it would be possible to optically excite any significant luminescence. The answer lies in the 5d or charge-transfer bands which either lie at higher energies, or overlap with the upper levels of the 4f configuration. These give rise to strong absorption and efficient pumping. Relaxation can occur via the parity-allowed transitions to the upper levels of the 4f configuration, and from there via single or multiple radiative emissions back to the ground state. Because these intra-configurational transitions are only weakly allowed, the lifetimes are generally quite long, of the order of μs–ms. Figure 38.2 shows some of the observed transitions in the case of the Pr^{3+} ion. There are further constraints on possible transitions which arise from an analysis of the angular momenta of the initial and final states. For example, a transition between two states both of which have $J = 0$ is forbidden since there is no angular momentum change as required for a photon; similarly for dipole transitions we require $\Delta J = 0, \pm 1$.

The transition probability per second for spontaneous emission [38.13] is given by

$$P_{ij} = \frac{64\pi^4\nu^3}{3hc^3}\left|r_{ij}\right|^2 , \tag{38.3}$$

where ν is the frequency of the transition, h is Planck's constant, c is the velocity of light, and $|r_{ij}|$ is the matrix element of the electric dipole operator $\boldsymbol{er}$ between the two states i and j, and e is the electronic charge. For absorption, this must be multiplied by N, the mean number of photons with energy $h\nu$, which thus incorporates the effect of the incident beam intensity. Experimentally, one measures an absorption coefficient k as a function of energy $k(E)$ [38.13] which is linked to P_{ij} through,

$$\int k(E)\mathrm{d}E = N_i\left(\frac{\pi e^2 h}{nmc}\right)\left(\frac{n^2+2}{3}\right)^2 f_{ij} , \tag{38.4}$$

where n is the refractive index of the crystal environment, m is the electronic mass, N_i is the concentration of the luminescent centres, and f is the *oscillator strength* for the transition. For both absorption and emission, the dimensionless oscillator strength f_{ij} defined as [38.13]

$$f_{ij} = \frac{8\pi^2 m\nu}{3he^2}\left|r_{ij}\right|^2 \tag{38.5}$$

is often quoted to compare the relative transition probabilities. For an electron harmonic oscillator, $\sum f_{ij} = 1$, and so oscillator strengths of the order of 0.1–1 are strongly allowed transitions.

38.1.2 Transition-Metal Ions

Transition-metal ions from the 3d series are characterised by a much stronger interaction with the crystalline environment than the 4f ions since there is no equivalent of screening by the 5s, 5p outer shells. In addition, the spin–orbit coupling is weaker, and so the order of perturbation is reversed: the atomic L, S multiplets are split by the crystal field, with spin–orbit coupling being a smaller interaction. Intra-configurational transitions are again strictly forbidden, but become weakly allowed by inter-configurational mixing through odd-parity crystal fields, and by odd-parity vibrations. The result of this is that the strongest selection rule after parity is that transitions should have $\Delta S = 0$, since the electric dipole operator does not involve spin. The other major difference, again due to the strength of the crystal-field interaction, is that transitions which are purely electronic, the so-called zero-phonon lines, are rarely observed. Rather what are seen

are broad bands which correspond to the simultaneous excitation of an electronic transition and vibrational transitions, which overlap to give the broad observed bands. In particular, transitions involving odd-parity vibrations have a high transition probability through the effect of configuration admixing. This will be considered in a following section. The most commonly observed luminescent ions are those from the d^3 configuration (Cr^{3+}, Mn^{4+}) and from the d^5 configuration (Mn^{2+}). *Avram* and *Brick* [38.14] give a recent and comprehensive account of transition metal ion spectroscopy and fluorescence.

38.1.3 s^2 Ions

The $5s^2$ (e.g., Sn^{2+} and Sb^{3+}) ions and $6s^2$ (e.g., Tl^{+}, Pb^{2+}, Bi^{3+}) ions are of considerable importance because transitions to and from the excited s^1p^1 states are Laporte-allowed. The interaction of the p state with the crystalline environment can be very strong, and so broad spectra are often observed.

38.1.4 Semiconductors

Luminescence in inorganic semiconductors is dominated by near-band-gap luminescence arising from recombination of electrons and holes. This process is most efficient in direct-band-gap materials such as ZnS and GaP rather than indirect-gap materials such as Si and Ge because the transition probability requires conservation of wavevector, but the photon wavevector is ≈ 0 on the scale of the Brillouin zone. Hence creation or destruction of a phonon is required for band-to-band luminescence in indirect-gap materials, which is less probable. The near-edge emission may correspond to luminescence from a variety of shallow energy-level structures such as free or trapped excitons, or from donor–acceptor recombination. These are both example of electronic systems with spatially extensive wavefunctions, in contrast to the atomically localised 3d and 4f wavefunctions considered earlier. However, it is also possible to observe *deep-level* luminescence from transition-metal ions and rare earths in semiconductors provided that the electron affinities and band-gap energies are such that the pertinent energy levels fall in the band gap. In recent years, recombination emission in structured PN junctions based on GaN and InGaN which emit in the UV/blue spectral region has become of substantial practical importance in solid state lighting. Since inorganic semiconductors are discussed elsewhere in this volume, we shall not consider them further here, except as pump sources for down-conversion phosphors in Sect. 38.6.2.

Organic semiconductors are the basis for organic light emitting diodes (OLEDs) which are now finding applications in television and mobile phone displays, and in lighting. The advantages of OLEDS are the simple processing required, and ease of large area production and elelctrode patterning, together with a wide range of available emission colours. The operation is in principle similar to that of inorganic diodes but with very different materials. A diode structure is is engineered with an emissive polymer or small molecule layer sandwiched between between a polymer which is a good hole transporting material and one which is a good electron transporting material. The two transport layers inject charge carriers into the emissive layer, where they form excitons, which subsequently decay through luminescent molecular units in the emissive layer. OLEDs are described in more detail in Chap. 51.

38.2 Interaction with the Lattice

For rare-earth ions, the interaction with the vibrations of the crystal lattice can be ignored for most purposes; the observed luminescence spectrum consists of sets of sharp electronic transitions. But for other luminescent ions which interact strongly with the vibrating ions of the surrounding crystal, the incorporation of the latter is critical to explaining the observed spectra. The simplest model of ion–lattice interactions is to consider only the N nearest neighbour ions and their atomic displacements X_n, Y_n, Z_n, $(n = 1, N)$ in Cartesian coordinates. The vibrational Hamiltonian involves cross terms in these coordinates, but may be transformed to harmonic form if symmetry-adapted forms of these coordinates (*normal modes*) are used instead of the actual displacements. For example, the so-called breathing mode Q_b, for an octahedrally coordinated ion takes the form

$$Q_b = \frac{(Z_1 - Z_2 + X_3 - X_4 + Y_5 - Y_6)}{6} . \qquad (38.6)$$

If all the other modes have zero amplitude, the ions move radially towards or away from the central luminescent ion. The key point in considering the influence of the crystal lattice is that the vibrational potential energy is just the variation of the electronic energy with ionic displacement, or equivalently with the normal modes (within the spirit of the Born–Oppenheimer

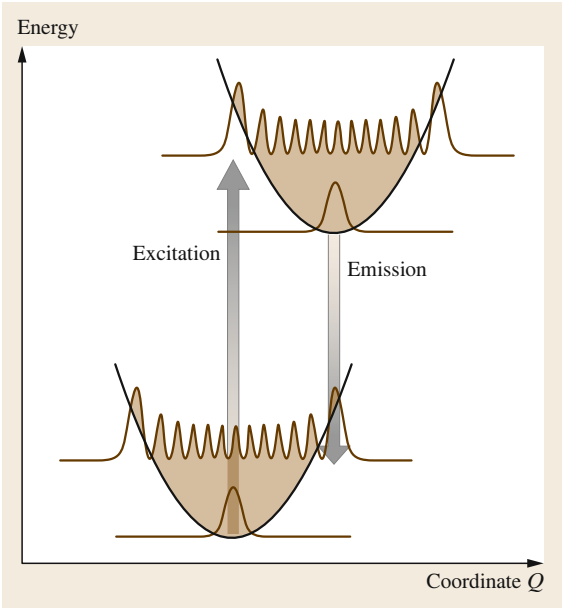


Fig. 38.3 Configuration coordinate diagram for excitation/emission cycle

approximation). Put another way, the crystal field depends on the ion positions so that the electron and lattice quantum-mechanical systems are linked through this electron–lattice coupling. We can therefore expect a difference in the harmonic vibrational potential from one electronic state to another, so that it will in general have the form $(1/2)m\omega_i^2(Q-Q_{0i})^2$; i. e., both the magnitude of the potential and the position of the minima Q_{0i} will depend on the electronic state i. The vibrational states in the harmonic approximation are just the usual simple harmonic oscillator states with energies $(n+1/2)h\nu_i$, with n an integer. Thus we arrive at a *configuration coordinate* diagram such as that shown in Fig. 38.3, with the potential energies for a ground state g and excited state e being offset parabolas of different curvature. The square of the vibrational wavefunction is shown for the ground state and that with $n=11$. We note that for a classical oscillator there would be a peak in the probability function at the extreme lengths of travel, corresponding to the maxima for $n=11$ at the outer limits of the wavefunction in Fig. 38.3.

We consider first luminescence from the excited electronic state/lowest vibrational state. The transition probability may be calculated using the Franck–Condon principle, that is we assume that the duration of the electronic transition is much shorter than a vibrational period, so that the Q remains constant during a transition. The transition is therefore taken to be *vertical* on the configuration coordinate diagram. The transition probability is just $\langle\phi_e|\phi_g\rangle^2$, the square

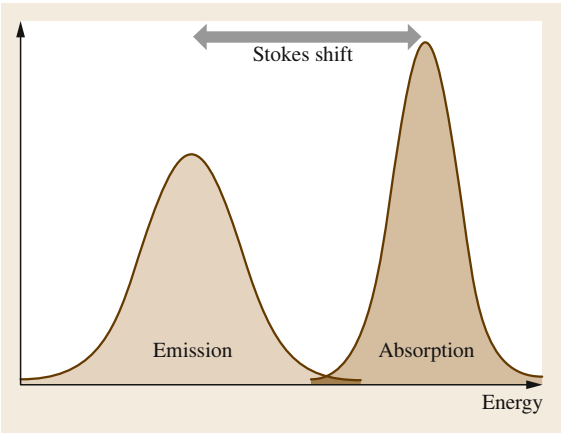


Fig. 38.4 Absorption and emission line shapes for strong electron–lattice coupling

of the vibrational wavefunction overlap, multiplied by the electronic transition probability considered earlier. From Fig. 38.3, this overlap will be a maximum for some vibrational state other than the ground state, unless the positions of the minima of the two potential energy curves coincide accidentally. (In Fig. 38.3, this maximum would be for the $n=11$ ground vibrational state). Put another way, the maximum transition probability is not for the zero-phonon transition (no change in vibrational state), but corresponds to the creation of a finite number of phonons. A range of transitions is allowed, and the result in a semiclassical analysis, allowing for finite line widths, is a Gaussian-shaped band. The analysis has to be extended to include finite temperatures and other modes, as described in detail by *Struck* and *Fonger* [38.15], but the overall result is that the emission line shape is approximately a Gaussian centred at an energy lower than that of the difference between the minima of the two potential curves. The same argument can be applied to the excitation process, and again an approximately Gaussian line shape results but this time centred on an energy above that of the difference in potential-energy minima. Figure 38.4 shows the overall result; the difference between the maxima of the two curves, known as the Stokes shift, is an indicator of the degree of electron–lattice coupling.

It is clear from the diagram that, in cases where there is strong electron–lattice coupling of this type, that: (a) there will be very little intensity in the zero-phonon line, and (b) there will be a large Stokes shift between the energies for maximum absorption and maximum emission. Thus the luminescence spectra of transition-metal ions, colour centres, and closed-shell ground-state ions (s^2, d^0), which have strong interactions with the lattice in the excited states, are typically broad bands with only occasionally weak, sharp zero-

phonon lines being observed. For luminescence from within the 4f states of the rare earths, the reverse is true; we are in the weak-coupling regime, and zero-phonon lines are the predominant features of the spectrum.

38.3 Thermally Stimulated Luminescence

Thermally stimulated luminescence (TSL), or simply thermoluminescence (TL), refers to luminescence induced by thermally stimulated recombination of trapped electrons and holes in materials which have been subject to prior irradiation. The irradiation, which may be in the form of UV light, x-rays, gamma rays, or energetic electrons, creates free electrons and holes, most of which promptly recombine, but some of which are locally trapped at defect centres such as impurities and vacancies. If the trap binding energies are sufficiently large, thermal promotion of the electron or hole to the conduction band or valence band, respectively, is improbable at the irradiation temperature, and so these charge carriers remain trapped after irradiation. However, if the sample is then heated, thermally assisted recombination becomes increasingly probable, and the result is an initially increasing light output with increasing temperature until the traps are depleted, whereupon the light intensity drops. The curve of light intensity versus temperature $I(T)$ is known as a glow curve, and may be analysed to extract the trap depths and concentrations. A comprehensive review of the field has been given by *McKeever* [38.16], and a shorter discussion by *Vij* [38.1].

The process is shown schematically for a simple system comprising a single electron trapping level T and a single recombination centre R in Fig. 38.5.

Irradiation results in a trapped electron at trap T and a trapped hole at R. The trapped hole binding energy is larger than that of the trapped electron, so the latter is depopulated first, with a probability P which has the form,

$$P = s \exp\left(\frac{-E}{k_B T}\right), \tag{38.7}$$

where s and E are the attempt frequency and the activation energy respectively. We have shown the hole trap and recombination centre as being one and the same in Fig. 38.5, but it is also possible that the recombination energy is transferred to a separate luminescent centre.

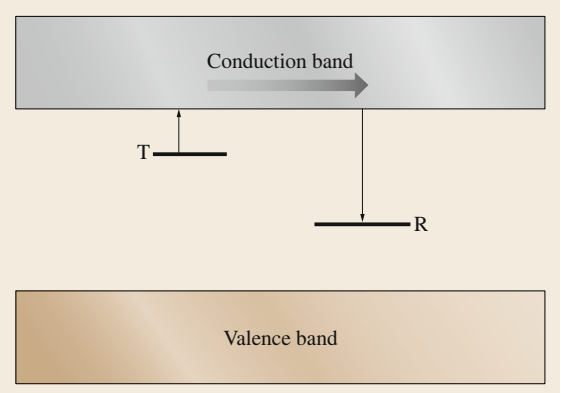


Fig. 38.5 Thermally stimulated luminescence process

In the measurement process, the sample is heated at a fast and linear rate, typically 1–10 K/s, whilst the light emission is monitored by a sensitive filter/photomultiplier combination, or by a monochromator system. The resulting glow curve may be fitted to a theoretical curve to extract the trap parameters. Generally the light output is quite weak, and TSL systems have sensitivity as one of their prime design factors, so cooled detectors and photon counting are commonplace. Above about 400 °C, thermal radiation from the sample/heater is a problem and must be eliminated by filtering or by subtraction of a glow curve recorded using a thermally bleached sample.

The main uses of TSL are in determining trap depths and irradiation doses. In archaeological and geological applications, a comparison is made between accumulated natural dose and a dose from a radioisotope; by combining this with a measurement of the activity of the surroundings, a date since last thermal or optical erasure of the object can be determined. The nature of the traps is often poorly understood for these chemically complex samples. For medical dosimetry, room-temperature stable traps with high thermoluminescent output are required. At present, the material of choice is LiF doped with a few hundred ppm of Mg and Ti.

The mathematical form of the TSL glow curve depends on the physical model used for the TL process. In the simplest case, assuming first-order kinetics [38.1, 16], the light intensity I at temperature T is given by,

$$I(T) = n_0 s \exp\left(\frac{-E}{k_B T}\right) \times \exp\left[-\frac{s}{\beta}\int_{T_0}^{T} \exp\left(\frac{-E}{k_B T}\right) \mathrm{d}T\right], \tag{38.8}$$

where n_0 is the number of occupied traps at time $t = t_0$ when the temperature is T_0, β is the heating rate in K/s,

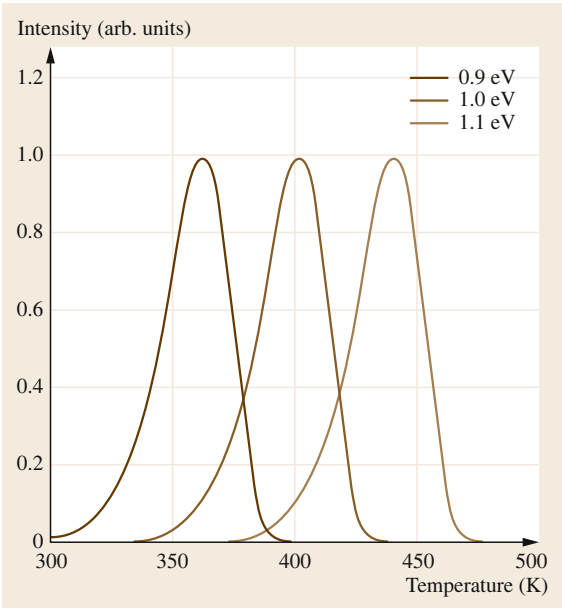


Fig. 38.6 Computed glow curves for first-order kinetics for fixed escape frequency and various trap depths

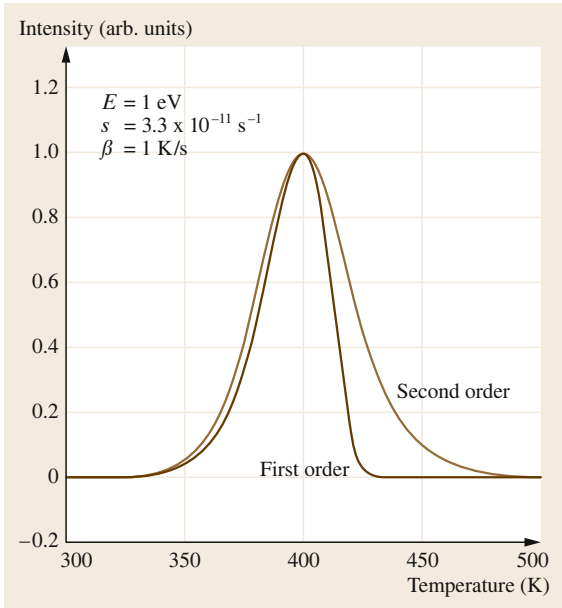


Fig. 38.7 Comparison of first- and second-order glow curves

and k_B is Boltzmann's constant. The result is a glow curve whose peak position varies approximately linearly with trap depth E, as shown in Fig. 38.6 (for fixed escaped frequency $s = 3.3 \times 10^{11}\,\mathrm{s}^{-1}$ and heating rate $1\,\mathrm{K/s}$).

Of course, the peak position also depends on the escape frequency s, but is less sensitive to s than to E. For second-order kinetics [38.1, 16], the glow curve equation becomes

$$I(T) = \frac{\left(\frac{n_0^2 s}{N}\right) \exp\left(\frac{-E}{k_B T}\right)}{\left[1 + \left(\frac{n_0 s}{N\beta}\right) \int_{T_0}^{T} \exp\left(\frac{-E}{k_B T}\right) \mathrm{d}T\right]^2} , \qquad (38.9)$$

where N is the number of available traps. The two forms both give glow curves of similar shape, but the first-order form shows an asymmetric form with a sharp fall off on the high-temperature side of the peak, whilst the second-order form is more symmetric as shown in Fig. 38.7. The parameters are extracted by least-squares fitting of these expressions to the experimental glow curves, which is cumbersome due to the integral, but *Kitis* et al. [38.17] have given analytical approximations to (38.8) and (38.9). In practice, many glow curves do not follow first- or second-order kinetics precisely. *Furetta* [38.18] describes the simulation of glow curves for a variety of TL analytical models.

38.4 Optically (Photo-)Stimulated Luminescence

Thermally stimulated luminescence is sometimes also accompanied by optically stimulated luminescence (OSL), in which one of the trapped carriers is excited by optical stimulation to a level from which it can recombine with the conjugate carrier by tunnelling, or completely to one of the bands so that recombination is achieved through what is essentially a photoconductivity effect. For OSL to be significant, there must be an appreciable optical transition probability, so not all TSL centres are OSL active. The stimulation energy measured in OSL generally differs from that determined from TSL because of the Franck–Condon principle. The OSL effect finds practical application in dosimetry, e.g., [38.19, 20] and in x-ray storage phosphors used for medical imaging, as described later.

38.5 Experimental Techniques – Photoluminescence

A typical traditional photoluminescence measuring system involves a broad-spectrum source, either a combined tungsten filament for the visible spectrum and deuterium lamp for the UV, or a xenon flash lamp. The lamp emission is passed through a grating monochromator and so selectively excites the luminescence. Band-pass or band-edge filters are generally required to eliminate unwanted second- and higher-order diffraction maxima from the grating. The luminescence is efficiently gathered by low-f/number optics and fed to a second grating monochromator, also equipped with filters, to monitor and analyse the luminescence. The final detector may be a photomultiplier, or preferably a charge-coupled device or photodiode array for improved data collection efficiency at multiple wavelengths.

This arrangement relies on good monochromation/filtering to remove what is sometimes a relatively strong component of scattered light from the beam analysed by the emission monochromator. An alternative method of removing scattered light is to use time discrimination, by replacing the source by a xenon flash lamp. The flash has a duration of about 10 μs, so any scattered light has decayed away to an insignificant level when the emitted beam is sampled some time (0.1−10 ms) after the flash. A timed electronic gate is used to sample the emission immediately after a flash and just before the next flash; the difference between these two sampled signals gives the short-term luminescence whilst the second sample alone gives the long-term luminescence with *long* and *short* being relative to the flash repetition period. Of course, luminescence with lifetimes shorter than the pulse width ($\approx 10\,\mu s$) cannot be readily measured with this system. The luminescence intensity is normalised with respect to the excitation intensity by steering a sample of the excitation beam to a rhodamine dye cell which has a quantum efficiency of essentially unity for wavelengths below about 630 nm. The fluorescence from the dye is measured with a second photomultiplier.

To minimise the effect of scattered light, a conventional laser with its intrinsically narrow linewidth and high intensity is a very convenient replacement for a broad-spectrum lamp, but suffers from the disadvantage of a fixed wavelength. Typical lasers of interest are nitrogen (pulsed), argon (UV lines, or frequency-doubled visible lines), krypton, and the new generations of GaN/GaInN blue/violet/UV laser diodes. For rare-earth spectroscopy, or other systems which are characterised by narrow absorption lines, it is very useful to have a scanning dye laser as the excitation source. This *selective excitation* facility enables *tagging* of particular luminescent levels with excited states belonging to the same centre, so that a picture of the energy-level structure of each luminescent centre can be built up in cases where several such centres contribute to the overall luminescence.

For decay kinetics on faster time scales, fluorimeters such as those developed initially by the Spex company (now Horiba) use a fast modulator and phase-sensitive detection to measure the phase shift between fluorescence and excitation; it is claimed that fluorescence decays can be measured with a resolution of 25 ps this way. An alternative is the time-correlated single-photon-counting technique which can measure decay constants in the ps–ns range. In this method, the excitation comes from a fast laser pulse, and the light level reaching the photomultiplier or micro-channel plate detector is reduced to such a low level that less than one photon per excitation pulse is detected. The time delay between the photon detection and the time of the pulse is measured, and a histogram produced of numbers of detected photons versus arrival time taken over a large number of excitation pulses. For efficient data collection, a high repetition rate and fast-pulse laser are required, often a Ti–sapphire laser. The wavelength for Ti–sapphire is too large for stimulating many materials directly with single-photon excitation, but stimulation is nonetheless possible by a two-photon excitation process, or by the use of a nonlinear crystal acting as a frequency doubler to produce laser output at one half the wavelength of the basic laser. Alternatively, a range of fixed wavelength LEDs and laser diodes with pulse widths in the range of ns to ps have become available.

There are a number of specialist techniques, such as hole-burning, fluorescence line-narrowing, and photon echo methods associated with the use of lasers with either very narrow line widths or short pulse duration which have developed in a parallel way to techniques first introduced in nuclear magnetic resonance, and which are mainly used to investigate the dynamics and quantum mechanics of the luminescent species rather than the material in which they are contained. *Meijerink* includes a short discussion of these specialist techniques in the book edited by *Vij* [38.1].

38.6 Applications

The largest market for luminescent materials has traditionally been in the areas of lighting, through fluorescent tubes, and in cathode-ray-tube screen phosphors for image display. Both of these areas can now be regarded as mature in terms of materials development, and in the case of cathode ray tubes, in decline as alternative display technologies have appeared. However, new discoveries in the past decade, and the advent of new technologies, have rekindled interest in phosphor materials. Some of the current areas of activity in applications are outlined below.

38.6.1 Compact Fluorescent Lamps

Compact Fluorescent Lamps (CFLs) are simply condensed versions of the traditional fluorescent lamp with a built-in ballast/starter in the lamp base. The tube contains an argon/mercury gas mixture, and an electric discharge between two electrodes results in the emission of (predominantly) 254 nm UV light from the ionised mercury. This light falls on the powder phosphor mixture on the inside surface of the glass tube, and excites luminescence from the rare earth ions contained within the phosphors. Typically Eu^{2+} is used for blue emission, Eu^{3+} for red, and Ce-activated Tb^{3+} for green. In the latter case, the Ce^{3+} ion efficiently absorbs the UV and the energy migrates to the Tb^{3+} ions. A variety [38.6] of host crystals are used, typically alkaline earth aluminates and yttria, selected on the basis of their performance and stability at high temperatures under intense UV irradiation and ion bombardment. A major challenge is that of good colour rendition, that is to achieve a light output that matches the spectrum of a conventional tungsten bulb in the visible region (which is close to the continuous spectrum expected for a black body radiator) with the multiline output of a rare-earth phosphor. Although CFLs offer a substantially higher energy efficiency and lifetime compared to conventional tungsten bulbs, albeit at a much higher cost, they face strong competion from the newer solid state LED technology as described below. The efficiencies of LED and CFL lamps are similar, but CFL's suffer from the disadvantages of having a slow start characteristic (as mercury liquid is progressively converted into gas as the lamp warms up), they are not readily dimmable, and they are an environmental hazard when they come to the end of their useful life due to the (small) mercury content.

38.6.2 Phosphors for Solid State Lighting

The development of blue, violet and UV LEDs based on GaN, InGaN and other semiconductors and alloys has laid the foundation for the new industry of solid state lighting, where high powered blue or UV LED excites intense visible light emission from an adjacent phosphor or phosphors [38.21, 22]. These so called *white-light LEDs* are making large inroads into the domestic and industrial lighting market as high-efficiency replacements for traditional incandescent bulbs, and are at present in direct competition with CFLs. The most common commercial design for a *white LED* uses visible blue LED emission to stimulate yellow luminescence from a single phosphor, so that when that yellow light is mixed with residual blue light from the LED, the result is simulated white light. For UV emitting LEDs, the emission is used to stimulate separate blue, red and green phosphors, or a single phosphor containing red, green and blue emitting centres, whose collective output approximates white light. The phosphors used for white light LEDs are quite different to those used for CFL's because of the need for efficient optical pumping in the range of 370–450 nm rather than 254 nm.

The most common single phosphor/single centre technology is based around a YAG:Ce garnet phosphor powder, embedded in a transparent polymer resin, which converts some of the emission from a 465 nm blue GaInN LED into an orange/yellow emission centred at 550 nm; the combination of blue and yellow simulates white light. A very strong crystal field splitting in the 5d state of Ce^{3+} shifts the lowest emitting level from the UV or blue where it normally occurs, to the yellow/orange region of the spectrum. YAG:Ce satisfies the considerations of optical and thermal stability in high power applications, but the colour rendition is not optimal, particularly in the red spectral region. Other single phosphors [38.21, 22] which have been investigated for this application include compositional variations on YAG:Ce, silicates, borates, sulphides, all with various rare earth dopings. The nitrides, oxynitrides and other complex nitrides [38.23, 24] are of particular interest from a colour rendition and thermal stability perspective. However, the original YAG:Ce has undergone performance improvements which should maintain its dominance in single phosphor white LEDs. The pumping and light extraction efficiency has been improved, for example through the development of nanoparticle forms such as transparent glass ceramics [38.25] which minimise back-scattering.

Whilst the single phosphor/single centre phosphor implementation has many advantages, including simplicity, improved colour rendition is obtained by using more phosphors or centres, with optimised balance between the various emissions, for example with blue diode pumped $SrGa_2S_4$:Eu (green), SrY_2S_4:Eu (red),

complemented by residual blue LED light. Recent developments based on 375 nm UV LEDs have used multiple emissions to achieve even better colour balance; for example *Kim* et al. [38.26] use $Sr_3MgSi_2O_8:Eu^{2+}$, Mn^{2+}, which has blue (Eu^{2+}), yellow (Eu^{2+}) and red emissions (Mn^{2+}). The europium ions presumably occupy two different sites with very different crystal field splittings. A major problem with multi-phosphor white LEDs is finding a satisfactory red emitter – usually Eu^{3+} fills this role, but it is not efficiently pumped by the available LED wavelengths, and so Mn^{2+} or Mn^{4+} are also under consideration [38.27].

Other directions in research include the possible use of photoluminescent quantum dots (QDs) [38.28] so as to improve colour rendition and achieve a warm white light output. QDs are nanometer-sized particles, in this context they are based on a semiconductor such as CdSe or CdS whose band gap lies in the visible region. The quantum confinement effect means that the energy of recombination emission in these direct band gap materials has a wavelength which is dependent on particle size, resulting in a size-tuneable phosphor. The emission colour can also be modified by chemical modification of the nanoparticle surface with luminescent ions such as rare earths; the recombination energy of electron-hole pairs excited by LED emission can migrate to the surface and excite these ions. The surface layer is a significant volume fraction for a nanoparticle, and so the surface emission can dominate the volume emission. The combination of tailored absorption by adjusting the nanoparticle size and composition, and tailored emission by chemical modification of the surface, gives a very flexible design palette. However, the toxicity of cadmium is an issue, and so other semiconductors which can be fabricated as nanoparticles are under active consideration.

The efficiencies of white LEDs are comparable with CFLs and (currently) about five times larger than for incandescent bulbs, and with a lifetime which is about twenty five times longer. The major problems relate to the cost, colour rendition, and stability at the high temperatures produced by incident optical powers of the order of tens of watts/cm^2, but continuing improvements can be expected in all of these aspects in contrast to the mature state of CFL development.

38.6.3 Long-Persistence Phosphors

It has been discovered in the past twenty years that some materials show a very long-lived but intense afterglow, arising from thermal emission of charge carriers from deep traps, followed by electron–hole recombination at or near a luminescent ion, i. e., room-temperature thermoluminescence. The persistence is for periods of several hours, much longer than the well-known ZnS:Cu persistent phosphor. This new class of materials [38.29], have widespread applications in areas such as signage and passive emergency lighting for public buildings, aircraft cabins etc. The material is activated or *charged* by the blue/UV content of solar radiation or fluorescent indoor lighting during normal conditions; in subsequent dark conditions the energy is released as an afterglow. No power supply is required, the operation is entirely passive, and so a high degree of reliability is assured. Generally the phosphor powder ($\approx 10\,\mu$m grain size) is mixed with a resin binder and applied as a thick (≈ 1 mm) coating. The first materials which were used were rare-earth-doped strontium or calcium aluminate [$CaAl_2O_4$:Eu/Nd (blue), $SrAl_2O_4$:Eu/Dy (green), $Sr_4Al_{14}O_{25}$:Eu/Dy (blue–green), and Y_2O_2S:Eu/Mg/Ti (orange)]. Figures 38.8 and 38.9 show the decay and emission/excitation characteristics of a commercial phosphor strip of this type, based on Luminova $SrAl_2O_4$:Eu/Dy powder sourced from the Nemoto company. The decay is clearly non–exponential, but after the initial rapid decay the intensity may be fitted with an exponential with a decay constant of 3.1 h.

The persistent phosphor effect has now been found in a wide variety of materials [38.30, 31] including the nitrides and oxynitrides (also of interest for LED lighting), although the aluminates remain as the most significant materials for all-round performance from a commercial perspective. The mechanism must involve

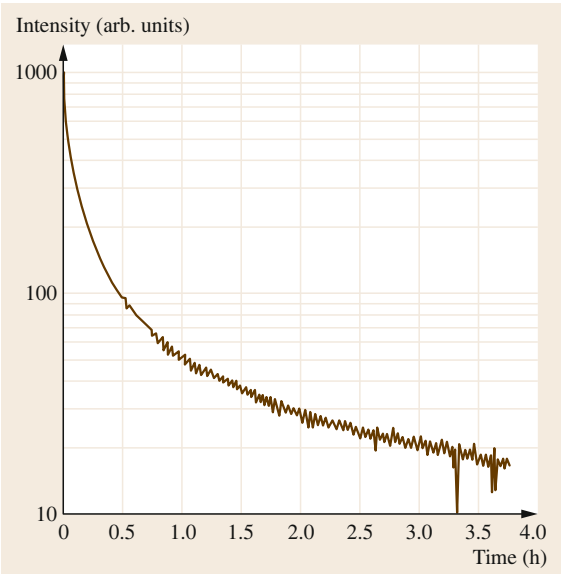


Fig. 38.8 Dark decay of persistent luminescence in a commercial lighting strip based on Nemoto $SrAl_2O_4$:Eu/Dy material

shallow charge carrier traps from which electrons or holes can recombine with the conjugate charge carrier through thermal excitation to the appropriate band, by tunnelling, or by thermally assisted tunnelling. A useful persistent phosphor requires both a substantial brightness, and a time constant for recombination which is of the order of hours. The traps which are responsible for the long-lived decays in these systems have not been clearly identified, and it is likely that there are several involved. It would seem that progress in identifying the mechanisms and traps will require corellated optical, ESR and TSL studies.

38.6.4 X-Ray Imaging

X-ray images have traditionally been recorded with phototographic film since the time of Roentgen, but it was noted in the very early days that photographic film itself was not very sensitive to x-rays. However, if a phosphor is used as an intermediate x-ray to visible photon converter, the sensitivity is increased by approximately two orders of magnitude. Consequently, x-ray film cassettes are usually outfitted with two intensifying screens which sandwich the photographic film. The operating principle of the screens is that an incident x-ray photon generates a large number (tens of thousands) of electron-hole pairs through the photoelectric effect or Compton scattering. When the electrons and holes recombine, the recombination energy can be efficiently transferred to a luminescent centre, resulting in the emission of a large number of visible light photons. The most common phosphor material [38.2] is powdered gadolinium oxysulphide doped with terbium, embedded in an organic binder, and which emits in the green region of the spectrum. However, this long standing technique of phosphor-enhanced film for x-ray radiography is being superceded by a variety of new methods, some of which are phosphor or scintillator based. Among these, optically stimulated or photo-stimulated luminescence is the basis for an x-ray imaging technology known commercially as computed radiography (CR). Several other techniques such as those based on amorphous selenium photoconductor/flat panels and a-Si arrays, and scintillator CCD/complementary metal–oxide–semiconductor (CMOS) arrays have also emerged in recent years, but CR remains as a substantive technology on the basis of numbers of installed units in medical and dental radiography. The many advantages of CR over photographic film, and the principles of the method are detailed in [38.32, 33], but the basic mechanism is that incident x-rays create electron–hole pairs in the material. Most of these promptly recombine, but some are trapped at defects and impurities, and remain trapped for periods of hours after the x-ray source is turned off. In an x-ray storage phosphor (XRSP), one of the trapped carriers is optically stimulable, and can be excited to the conduction band or valence band, or to a level from which it can recombine by tunnelling with the conjugate trapped carrier. The resulting recombination energy is transferred to a luminescent ion, and the intensity of the photo-stimulated luminescence (PSL) is in direct proportion to the incident x-ray intensity. The dominant material used in current XRSP systems is $BaFBr_{1-x}I_x:Eu^{2+}$, where the electron traps are F-centres, the hole traps are unidentified, and the luminescence is the 5d–4f transition of the Eu^{2+} ion. One disadvantage of systems based on this (powder) phosphor is that when the image is extracted with a raster-scanned He–Ne laser beam, light scattering from the powder grains means that material outside the focussed laser spot is also stimulated, limiting the spatial resolution to around 200 μm, which is inadequate for applications such as mammography. Several ways to overcome this are currently under development. CsBr:Eu is also an x-ray storage phosphor [38.34] and can be grown by vapour deposition in a columnar form. The columnar structure has a light-guiding property, restricting the scattering effect, and improving the resolution. A second development is that of glass-ceramic storage phosphors, where PSL active crystals are embedded in a glass [38.35–37]; the combination of particle size, separation and refractive-index mismatch means that these composite materials are semitransparent and the problem of scattering of read-out light

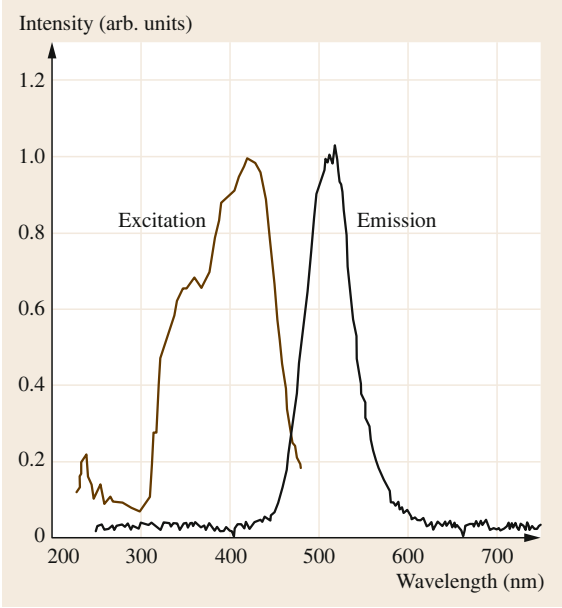


Fig. 38.9 Excitation and emission spectra of a commercial lighting strip based on Nemoto $SrAl_2O_4$:Eu/Dy material

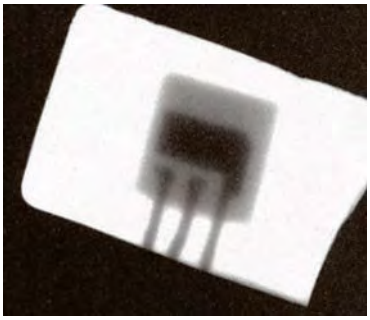


Fig. 38.10 PSL image of a BC549 transistor recorded on a glass-ceramic imaging plate

is reduced. Figure 38.10 shows an image of a BC549 transistor recorded in a glass-ceramic x-ray storage phosphor.

The quest for high resolution x-ray imaging has also led to the development of transparent glasses in which the image is recorded by a quite different luminescent ion technique. Some rare earth ions, particularly Eu and Sm can show a valence change under x-ray or UV irradiation. For example, Sm present in the divalent state in fluorophosphate and fluoroaluminate glasses has been shown to transform in part from the trivalent to the divalent state under x-irradiation [38.38]. The extra electron is sourced from the simultaneous creation of colour centres. The concentration of divalent samarium can be monitored by the Sm^{2+} photoluminescence spectrum, which is readily distinguished from Sm^{3+} as there is minimal overlap, and can be further accentuated by an appropriate choice of excitation wavelength. One-dimensional x-ray imaging based on the monitoring of the Sm^{2+} photoluminescence has been reported by *Okada* and *Morell* et al. [38.38, 39] using the methodology of confocal microscopy to achieve spatial resolutions of the order of ten microns. This particular application of one dimensional imaging was to beam profile monitoring in microbeam therapy, a proposed oncology treatment. *Edgar* et al. have shown [38.40] that two dimensional imaging can be achieved with the same materials using a simple camera technique, and demonstrated resolution of the 25 μm bonding wires on a semiconductor chip. However, the x-ray dose required for the fluorophosphate and fluoroaluminate glasses is too high for conventional medical radiography, and further work on alternative materials or dopants is required to extend the range of applications to the low dose regime.

38.6.5 Phosphors for Displays

Several new technologies have been developed to replace cathode ray tubes for domestic televisions, including plasma display panels (PDPs). In these units, each pixel is a sealed cell containing a mixture of Xe and Ne in a dielectric-shielded electrode structure (for a review, see *Boeuf* [38.41]). An alternating current (AC) voltage applied between the electrodes results in a glow discharge being set up in the gas, and a Xe dimer vacuum UV (VUV) emission predominantly between 147 and 190 nm occurs. (In comparison, the mercury discharge in a conventional fluorescent tube emits primarily at 254 nm.) The UV discharge excites red, blue, or green phosphors coated on one of the cells; each colour is activated by an adjacent electrode. The requirements for efficient output from these phosphors differ from conventional tubes since the latter were chosen on the basis of their luminescence efficiency at a 254 nm pump wavelength, and their resistance to degradation by the UV light and chemical attack by Hg^{+} ions. These requirements are evidently different for the PDP technology; in addition the phosphors must have a significant reflection coefficient in the visible to optimise the light output [38.42], and the surface quality is of greater significance due to the short penetration depth of the VUV. The phosphors which have been used so far include $BaMgAl_{10}O_{17}:Eu^{2+}$ (blue), $Zn_2SiO_4:Mn^{2+}$ (green), and $(Y, Gd)BO_3:Eu^{3+}$ and $Y_2O_3:Eu^{3+}$ (red). The blue phosphor is prone to degradation.

The widespread introduction of Xe excimer excitation in PDPs can be expected to stimulate applications in other lighting technologies. In this regard, the possibility of so-called *quantum cutting* is of much interest. This recognises that the energy of a VUV photon is equivalent to two or more visible photons, so that quantum efficiencies in excess of 100% can in principle be achieved. The difficulty lies in finding a luminescent ion system whose energy-level system provides for both efficient pumping and two-photon luminescence in the visible. One example which has been reported to have a quantum efficiency of up to $\approx 145\%$ is Pr^{3+} in YF_3 and other hosts [38.43, 44], where the excitation is through the allowed $4f^2 \rightarrow 4f^15d^1$ or host transitions. The system then decays to the $^1S_0$ excited state of the $4f^2$ configuration from which two-photon decay is possible through successive $^1S_0 \rightarrow {}^1I_6$ and $^3P_0 \rightarrow {}^3H_J$, ^3F_J transitions, as shown in Fig. 38.2. (The intermediate step from $^1I_6$ to $^3P_0$ is provided by a nonradiative transition.) A difficulty is that the photon for the first transition is in the UV region of the spectrum, and so it is necessary to incorporate a second luminescent ion pumped by this transition to convert the UV to visible output, and the visible quantum efficiency is necessarily reduced.

As at 2016, it would appear that PDP displays are losing ground in comparison with LCD and OLED based display technologies, which are discussed elsewhere in this handbook. In contrast, a rising technology is so-called QLEDs, or quantum dot OLEDs,

in which an OLED structure is modified to incorporate quantum dots within the electron-hole recombination zone [38.28]. The recombination energy migrates to the quantum dots, which can be customised for colour output by adjustments of size, semiconductor, and coating. A significant improvement in efficiency has been achieved by replacing the organic hole-transport layer with an inorganic one. The current application of QLEDs is primarily in backlighting for LCD displays, where the existing cold cathode fluorescent or white LED panels suffer from poor colour rendition.

38.6.6 Scintillators

Although semiconductor detectors of ionising radiation have made large inroads into the particle detection market, traditional scintillators are still widely used and are indispensable for some applications. The operating principle is that an incident gamma ray creates a large number of electrons and holes in the scintillating material directly or indirectly through the photoelectric effect, Compton scattering, or pair production. These energetic electrons and holes lose energy to the lattice on a picosecond time scale through thermalisation processes, and their energy approaches that of the band edges. They can then associate as free or bound excitons, and the energy of these electron-hole pairs is then localised on luminescent ions which have been deliberately doped into the material. The luminescent ion is promoted by this energy into an excited state, from which it can decay by photon emission. One gamma or x-ray photon typically creates tens of thousands of electron-hole pairs, and a similar number of visible region photons which are emitted in one aggregate pulse or *scintillation*. Charged particles such as protons produce electron–hole pairs through the Coulomb interaction with the band electrons. The scintillation or multiphoton bursts which signal the event is detected by a photomultiplier, and the height of the output pulse from the photomultiplier is proportional to the energy of the particle. A pulse-height analyser sorts the pulses according to energy, and so an energy spectrum can be obtained. Key figures of merit for a scintillator material include the light yield, i. e., the numbers of photons per MeV of particle energy, the radiative lifetime of the luminescent ion (since possible pulse overlap limits the maximum count rate which can be measured), and the weighted density ρZ_{eff}^4, which reflects the gamma sensitivity. (Here Z_{eff} is the effective atomic number.) The most widely used scintillator for many years has been NaI:Tl, but many different scintillators are being investigated, driven by the need for improved performance and lower cost for medical applications such as positron emission tomography (PET) and single-photon emission computed tomography (SPECT), and for large-scale elementary-particle facilities such as the Large Hadron Collider at the Centre Européen pour la Recherche Nucléaire (CERN). In the latter regard, the Crystal Clear collaborative project and other programs have resulted in several new materials such as $LaBr_3$, $LaCl_3$, Lu_2SiO_5, Gd_2SiO_5, $Lu_3Al_5O_{12}$ and $LuAlO_3$, all Ce-doped, $CeBr_3$, and undoped $Bi_4Ge_3O_{12}$ and $PbWO_4$, with typical performance figures of 10 000–70 000 photons/MeV and lifetimes of 10–50 ns. For gamma spectroscopy the pulse-height resolution is critical and $LaBr_3$:Ce has twice the resolution of NaI:Tl. The fastest scintillators are based on core-valence luminescence where a hole created in a core level recombines with an electron in the valence band. For example, BaF_2 shows this *cross-luminescence* effect with a lifetime as short as 600 ps. The effect is only shown by materials with a core-valence (CV) band energy gap less than the usual band gap, otherwise CV luminescence is absorbed. Neutron-sensitive scintillators can be made by incorporating an element which has a large cross section for thermal neutron capture such as Li, B, Gd, or Cd; the nuclear decay following neutron capture results in gamma, alpha, or beta emission which again results in electron–hole pair production and scintillation. Plastics are also useful for neutron detection because of the large neutron cross section of hydrogen.

Most scintillator materials are single crystal inorganics, typically halides or oxides, doped with rare earth ions, usually Ce or Eu. However, it has been shown in the past decade that high-performance scintillators can also be made in the form of transparent glass ceramics [38.25] which has advantages in terms of processing and manufacturing flexibility with regard to the geometrical form.

Rodnyi [38.45], *Lecoq* et al. [38.46], and *van Eijk* [38.47] discuss scintillator physics and performance in detail, whilst *Nikl* and *Yoshikawa* [38.48] present a recent update on inorganic single crystal scintillators.

38.7 Representative Phosphors

To conclude, we present in Table 38.2 a list of several luminescent materials of practical significance. The table is intended to be representative rather than comprehensive. It is noticeable from the table that just a few ions are responsible for a large number of applications, and primarily as oxides.

Table 38.2 Some luminescent materials of practical significance

Host	Dopants	Colour	Excitation	Application
$Bi_4Ge_3O_{12}$	–	Blue	Ionising radiation	Scintillator
$Y_3Al_5O_{12}$	Ce^{3+}	Yellow	Blue, violet	White LED
Gd_2SiO_5	Ce^{3+}	UV	Ionising radiation	Scintillator
$Lu_3Al_5O_{12}$	Ce^{3+}	Green	Ionising radiation	Scintillator
$LaBr_3$	Ce^{3+}	UV	Ionising radiation	Scintillator
BaFBr	Eu^{2+}	UV/blue	X-rays	X-ray imaging
CsBr	Eu^{2+}	Blue	X-rays	X-ray imaging
$BaMgAl_{10}O_{17}$	Eu^{2+}	Blue	UV	Fluorescent lamps, plasma displays
Gd_2O_2S	Tb^{3+}	Green	X-rays	Intensifying screens
Sr_3SiO_5	Eu^{2+}	Blue	UV	White LED
$SrGa_2S_4$	Eu^{2+}	Green	UV	White LED
$SrAl_2O_4$	Eu^{2+}, Dy^{3+}	Green	UV, violet	Persistent phosphor
$CaAl_2O_4$	Eu^{2+}, Nd^{3+}	Blue	UV, violet	Persistent phosphor
Y_2O_3	Eu^{3+}	Red	Electrons, UV	Plasma displays, fluorescent lamps
Sr_2SiO_4	Eu^{3+}	Yellow	UV	White LED
$(Y, Gd)BO_3$	Eu^{3+}	Red	UV	Plasma displays
SrY_2S_4	Eu^{3+}	Red	UV	White LED
LiF	Mg^{2+} and Ti^{4+}	UV/blue	Ionizing radiation	TL dosimetry
ZnS	Mn^{2+}	Yellow	Electric field	Panel displays
Zn_2SiO_4	Mn^{2+}	Green	UV	Plasma displays
$CeMgAl_{11}O_{19}$	Tb^{3+}	Green	UV	Fluorescent lamps

Acknowledgments. The author would like to acknowledge financial support from the New Zealand Ministry of Business, Innovation, and Employment, and the Foundation for Research Science and Technology.

References

38.1 D.J. Vij (Ed.): *Luminescence of Solids* (Plenum, New York 1998)
38.2 G. Blasse, B.C. Grabmeier: *Luminescent Materials* (Springer, Berlin, Heidelberg 1994)
38.3 B. Henderson, G.F. Imbusch: *Optical Spectroscopy of Inorganic Solids* (Oxford Univ. Press, Oxford 2006)
38.4 C. Ronda (Ed.): *Luminescence* (Wiley-VCH, Weinheim 2008)
38.5 A. Kitai (Ed.): *Luminescent Materials and Their Applications* (Wiley, Chichester 2008)
38.6 A. Lakshmanan: *Luminescence and Display Phosphors* (Nova Science, New York 2008)
38.7 G.H. Dieke: *Spectra and Energy Levels of Rare Earth Ions in Crystals* (Interscience, New York 1968)
38.8 W.T. Carnall, G.L. Goodman, K. Rajnak, R.S. Rana: J. Chem. Phys. **90**, 343 (1989)
38.9 G. Lui, B. Jacquier (Eds.): *Spectroscopic Properties of Rare Earths in Optical Materials* (Springer, Berlin 2005)
38.10 S. Huefner: *Optical Spectra of Transparent Rare Earth Compounds* (Academic Press, New York 1978)
38.11 A.A. Kaplyanski, R.M. McFarlane: *Spectroscopy of Crystals Containing Rare Earth Ions* (Elsevier, Amsterdam 1987)
38.12 E. Merzbacher: *Quantum Mechanics* (Wiley, New York 1970)
38.13 D. Curie: *Luminescence in Crystals* (Methuen, London 1962)
38.14 N.M. Avram, M.G. Brick: *Optical Properties of 3d-Ions in Crystals* (Springer, Berlin 2013)
38.15 C.W. Struck, W.H. Fonger: *Understanding Luminescence Spectra and Efficiency Using Wp and Related Functions* (Springer, Berlin 1991)
38.16 S.W.S. McKeever: *Thermoluminescence of Solids* (Cambridge Univ. Press, Cambridge 1985)
38.17 G. Kitis, J.M. Gomex-Ros, J.W.N. Tuyn: J. Phys. D **31**, 2636 (1998)
38.18 C. Furetta: *Handbook of Thermoluminescence*, 2nd edn. (World Scientific, Singapore 2010)

38.19 E.G. Yukihara, S.W.S. McKeever: *Optically Stimulated Luminescence: Fundamentals and Applications* (Wiley, Chichester 2011)
38.20 L. Botter-Jensen, S.W.S. McKeever, A.G. Wintle: *Optically Stimulated Luminescence Dosimetry* (Elsevier, Amsterdam 2003)
38.21 J. McKittrick, L.E. Shea-Rohwer: J. Am. Ceram. Soc. **97**, 1327 (2014)
38.22 S. Ye, F. Xiao, Y.X. Pan, Y.Y. Ma, Q.Y. Zhang: Mater. Sci. Eng. R **71**, 1 (2010)
38.23 Y. Tian: J. Solid State Light. **1**, 11 (2014)
38.24 R.J. Xie, N. Hirosaki: Sci. Technol. Adv. Mater. **8**, 588 (2007)
38.25 A. Edgar: Optical properties of glasses. In: *Optical Properties of Condensed Matter*, ed. by J. Singh (Wiley, Chichester 2006)
38.26 P.L. Kim, P.E. Jeon, Y.H. Park, J.C. Choi, L.P. Park: Appl. Phys. Lett. **85**, 3696 (2004)
38.27 J. Meyer, F. Tappe: Adv. Opt. Mater. **3**, 424 (2015)
38.28 V. Wood, V. Bulovic: Nano Rev. **1**, 5202 (2010)
38.29 K. van den Eeckhout, D. Poelman, P.F. Smet: Materials **3**, 2536 (2010)
38.30 K. van den Eeckhout, D. Poelman, P.F. Smet: Materials **6**, 2789 (2013)
38.31 P.F. Smet, J. Botterman, K. van den Eeckhout, K. Korthuit, D. Poelman: Opt. Mater. **36**, 1913 (2014)
38.32 S. Schweizer: Phys. Status Solidi **187**, 335 (2001)
38.33 J.A. Rowlands: Phys. Med. Biol. **47**, R123 (2002)
38.34 P. Leblans, D. Vandenbroucke, P. Willems: Materials **4**, 1034 (2011)
38.35 S. Schweizer, L. Hobbs, M. Secu, J.-M. Spaeth, A. Edgar, G.V.M. Williams: Appl. Phys. Lett. **83**, 449 (2003)
38.36 M. Secu, S. Schweizer, A. Edgar, G.V.M. Williams, U. Rieser: J. Phys. C **15**, 1 (2003)
38.37 A. Edgar, G.V.M. Williams, S. Schweizer, M. Secu, J.-M. Spaeth: Curr. Appl. Phys. **4**, 193 (2004)
38.38 G. Okada, B. Morell, C. Koughia, A. Edgar, C. Varoy, G. Belev, T. Wysokinski, D. Chapman, S. Kasap: Appl. Phys. Lett. **99**, 121105 (2011)
38.39 B. Morrell, G. Okada, S. Vahedi, C. Koughia, A. Edgar, C. Varoy, G. Belev: J. Appl. Phys. **115**, 063107 (2014)
38.40 A. Edgar, C.R. Varoy, C. Koughia, G. Okada, G. Belev, S. Kasap: J. Non-Cryst. Solids **377**, 124 (2012)
38.41 J.P. Boeuf: J. Phys. D **36**, R53 (2003)
38.42 H. Bechtel, T. Juestel, H. Glaeser, D.U. Wiechert: J. Soc. Inf. Disp. **10**, 63 (2002)
38.43 S. Kuck, I. Sokolska, M. Henke, M. Doring, T. Scheffler: J. Lumin. **102/103**, 176 (2003)
38.44 A.B. Vink, P. Dorenbos, C.W.E. Van Eijk: J. Solid State Chem. **181**, 308 (2003)
38.45 P.A. Rodnyi: *Physical Processes in Inorganic Scintillators* (CRC Press, Baton Rouge 1997)
38.46 P. Lecoq, A. Annenkov, A. Gektin, M. Korzhik, C.O. Pedrini: *Inorganic Scintillators for Detector Systems* (Springer, Berlin 2006)
38.47 C.W.E. Van Eijk: Nucl. Instrum. Methods Phys. Res. A **460**, 1 (2001)
38.48 M. Nikl, A. Yoshikawa: Adv. Opt. Mater. **4**, 463 (2015)

39. Nano-Engineered Tunable Photonic Crystals

Harry E. Ruda, Naomi Matsuura

Photonic crystals offer a well-recognized ability to control the propagation of modes of light in an analogous fashion to the way in which nanostructures have been harnessed to control electron-based phenomena. This has led to proposals and indeed demonstrations of a wide variety of photonic-crystal-based photonic devices with applications in areas including communications, computing and sensing, for example. In such applications, photonic crystals can offer both a unique performance advantage, as well as the potential for substantial miniaturization of photonic systems. However, as this review outlines, two-dimensional (2-D) and three-dimensional (3-D) structures for the spectral region covering frequencies from the ultraviolet (UV) to the near-infrared (near-IR) ($\approx 2\,\mu$m) are challenging to fabricate with appropriate precision, and in a cost-effective and also flexible way, using traditional methods. Naturally, a key concern is how amenable a given approach is to the intentional incorporation of selected defects into a particular structure. Beyond passive structures, attention turns to so-called active photonic crystals, in which the response of the photonic crystal to light can be externally changed or tuned. This capability has widespread potential in planar lightwave circuits for telecommunications, where it offers mechanisms for selective switching, for example. This review discusses alternative proposals for tuning of such photonic crystals.

39.1 **PC Overview** 1014
39.1.1 Introduction to PCs 1014
39.1.2 Nanoengineering of PC Architectures ... 1015
39.1.3 Materials Selection for PCs 1016

39.2 **Traditional Fabrication Methodologies for Static PCs** 1017
39.2.1 2-D PC Structures 1017
39.2.2 3-D PC Structures 1022

39.3 **Tunable PCs** 1027
39.3.1 Tuning the PC Response by Changing the Refractive Index of the Constituent Materials 1027
39.3.2 Tuning PC Response by Altering the Physical Structure of the PC 1028

39.4 **Summary and Conclusions** 1030

References 1031

Photonic crystals (PCs) are periodic, dielectric, composite structures in which the interfaces between the dielectric media behave as light-scattering centers. PCs consist of at least two component materials having different refractive indices, and which scatter light due to their refractive-index contrast. The one-, two-, or three-dimensional (1-D, 2-D, or 3-D) periodic arrangement of the scattering interfaces may, under certain conditions, prevent light with wavelengths comparable to the periodicity dimension of the PC from propagating through the structure. The band of forbidden wavelengths is commonly referred to as a *photonic band gap* (PBG). Thus, PCs are also commonly referred to as photonic-band-gap (PBG) structures.

PCs have great potential for providing new types of photonic devices. The continuing demand for photonic devices in the areas of communications, computing, and signal processing, using photons as information carriers, has made research into PCs an emerging field with considerable resources allocated to their technological development. PCs have been proposed to offer a means for controlling light propagation in submicron-scale volumes – the photon-based equivalent of a semiconductor chip – consisting of optical devices integrated together onto a single compact circuit. Proposed applications of PCs for the telecommunication sector include optical cavities, high-Q filters, mirrors, channel add/drop filters, superprisms and compact waveguides for use in so-called planar lightwave circuits (PLCs).

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_39

Practical applications of PCs generally require manmade structures, as photonic devices are designed primarily for light frequencies ranging from the ultraviolet to the near-infrared regime (i. e., $\approx 100\,\mathrm{nm}$ to $\approx 2\,\mu\mathrm{m}$ respectively) and PCs having these corresponding periodicities are not readily available in nature. 1-D PCs in this wavelength range may be easily fabricated using standard thin-film deposition processes. However, 2-D and 3-D PC structures are significantly more difficult to fabricate and remain among the more challenging nanometer-scale architectures to realize with cost-effective and flexible patterning using traditional fabrication methodologies. Recently, there has been considerable interest in PC-based devices that has driven advanced fabrication technologies to the point where techniques are now available to fabricate such complex structures reliably on the laboratory scale. In addition to traditional semiconductor nanostructure patterning methods based on advanced patterning/etching techniques developed by the semiconductor industry, novel synthesis methods have been identified for 2-D and 3-D periodic nanostructured PC arrays. There are several excellent reports reviewing these fabrication techniques in the literature [39.1–4] and this growing field has already been the subject of numerous recent reviews, special issues, and books in the area of theoretical calculations (both band-structure and application simulations), 2-D PC structures, 3-D PC structures, and opal-based structures [39.5].

Recently, there has been great interest in exploring the use of PCs for active applications in the field of telecommunications, such as in the area of PLCs (e.g., for optical switching). In such applications, the PC properties should be adjustable to create *tunable* photonic band gaps. This development increases the functionality of all present applications of PCs by allowing the devices in such applications to be adjustable, or tunable. We review recent developments in the engineering of tunable nanometer-scale architectures in 2-D and 3-D. This review aims to organize this ever-changing volume of information such that interested theorists can design structures that may be easily fabricated with certain materials, and such that technologists can try to meet existing fabrication gaps and issues with current systems.

The chapter begins with a brief introduction to PCs, followed by general criteria used to determine appropriate methodologies for the nanoengineering of tunable PCs. Finally, overviews of the most common fabrication methods for tunable 2-D and 3-D PCs will be given.

39.1 PC Overview

39.1.1 Introduction to PCs

The simplest PC structure is a multilayer film, periodic in 1-D, consisting of alternating layers of material with different refractive indices (Fig. 39.1). Theoretically, this 1-D PC can act as a perfect mirror for light with wavelengths within its photonic band gap, and for light incident normal to the multilayer surface. 1-D PCs are found in nature, as seen for example in the iridescent colors of abalone shells, butterfly wings and some crystalline minerals [39.6], and in manmade 1-D PCs (i. e., also known as Bragg gratings). The latter are widely used in a variety of optical devices, including dielectric mirrors, optical filters and in optical fiber technology.

The center frequency and size (i. e., frequency band) or so-called stop band of the PBG depends on the refractive index contrast (i. e., n_1/n_2, where n_1 and n_2 represent the refractive index of the first and second material respectively) of the component materials in the system. Figure 39.1 is an example of a 1-D PC, with a periodic arrangement of low-loss dielectric materials. This multilayer film is periodic in the z-direction and extends to infinity in the x- and y-directions. In 1-D, a photonic band gap occurs between every set of bands, at either the edge or at the center of the Brillouin zone – photonic band gaps will appear whenever n_1/n_2 is not equal to unity [39.7]. For such multilayer structures, corresponding photonic band gap diagrams show that the smaller the contrast, the smaller the band gaps [39.7]. In 1-D PCs, if light is not incident normal to the film surface, no photonic band gaps will exist. It is also important to note that at long wavelengths (i. e., at wavelengths much larger than the periodicity of the PC), the electromagnetic wave does not probe the fine structure of the crystal lattice and effectively sees the structure as a homogeneous dielectric medium.

The phenomena of light waves traveling in 1-D periodic media was generalized for light propagating in any direction in a crystal, periodic in all three dimensions, in 1987 when two independent researchers suggested that light propagation in 3-D could be controlled using 3-D PCs [39.8, 9]. By extending the periodicity of the 1-D PC to 2-D and 3-D, light within a defined frequency range may be reflected from any angle in a plane (in 2-D PBG structures) or at any angle (in 3-D PBG structures).

Since the periodicity of PCs prevents light of specific wavelengths (i. e., those within the photonic band

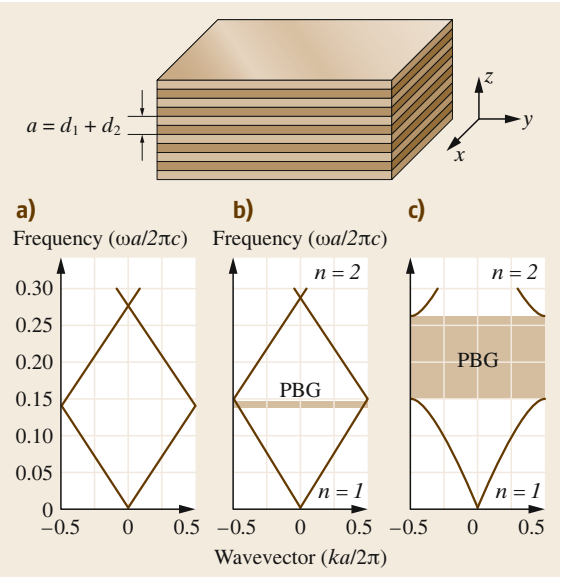


Fig. 39.1a–c Example of a 1-D photonic crystal. For a structure with $d = 0.5a$, the corresponding band gap diagrams are shown for: **(a)** GaAs bulk ($\varepsilon = 13$); **(b)** GaAs/GaAlAs multilayer ($\varepsilon_1 = 13$, $\varepsilon_2 = 12$); and **(c)** GaAs/air multilayer ($\varepsilon_1 = 13$, $\varepsilon_2 = 1$). (After [39.7])

gap) from propagating through them in a given direction, the intentional introduction of *defects* in these structures allows PCs to control and confine light. Propagation of light with wavelengths that were previously forbidden can now occur through such defect states located within the photonic band gap. Defects in such PCs are defined as regions having a different geometry (i. e., spacing and/or symmetry) and/or refractive-index contrast from that of the periodic structure. For example, in a 2-D PC consisting of a periodic array of dielectric columns separated by air spaces, a possible defect would include the removal of a series of columns in a line. Specific wavelengths of light forbidden from propagating through defect-free regions would then be able to propagate through the line defect, but not elsewhere. Indeed, by appropriately eliminating further columns, light may be directed to form optical devices, including, for example, a low-loss 90° bend in a 2-D waveguide, as shown by theoretical simulation [39.7] (Fig. 39.2).

Clearly, photons controlled and confined in small structures, with size on the order of the wavelength of light, using the extremely tight bend radii offered by PCs, would facilitate miniaturization and the fabrication of PLCs [39.7]. In addition, since the periodicity of the PC gives rise to the existence of band gaps that change the dispersion characteristics of light at given frequencies, defect-free PCs give rise to other

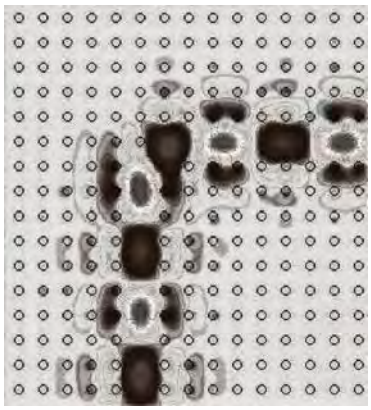


Fig. 39.2 Theoretical simulation of a low-loss 90° bend in a 2-D waveguide. (After [39.7])

interesting phenomena, including highly dispersive elements through the so-called superprism effect [39.10]. Possible designs of PC-based optical devices using such properties have been extensively explored [39.1, 11], generating much excitement in the field of optical telecommunications [39.6].

39.1.2 Nanoengineering of PC Architectures

Most of the promising applications of 2-D and 3-D PCs depend on the center frequency and frequency range of the photonic band gaps. A so-called *complete*, *full*, or *true* PBG is defined as one that extends throughout the entire Brillouin zone in the photonic band structure – that is, for all directions of light propagation for photons of an appropriate frequency. An incomplete band gap is commonly termed a *pseudogap* or a *stop band*, because it only occurs for reflection/transmission along a particular propagation direction. A complete gap occurs when stop-band frequencies overlap in all directions in 3-D. The center frequencies and stop-band locations of the PBGs critically depend on the unit-cell structure [39.7–9, 12, 13]. In particular, the PC properties depend on the symmetry of the structure (i. e., the unit-cell arrangements), the scattering element shape within the unit cell, the fill factor (i. e., the relative volume occupied by each material), the topology, and the refractive-index contrast (Fig. 39.3).

It has been shown that a triangular lattice with circular-cross-section scattering elements in 2-D, or a face-centered-cubic/diamond lattice with spherical scattering elements in 3-D, tend to produce larger PBGs [39.7]. Also, as discussed above, the dielectric contrast is an important determining factor for 2-D and 3-D PC structures. The lower the dimension of the structure, the more readily are PBG manifested, since overlap of the PBGs in different directions is more likely (something which is a certainty in 1-D structures). In 3-D structures, calculations have deter-

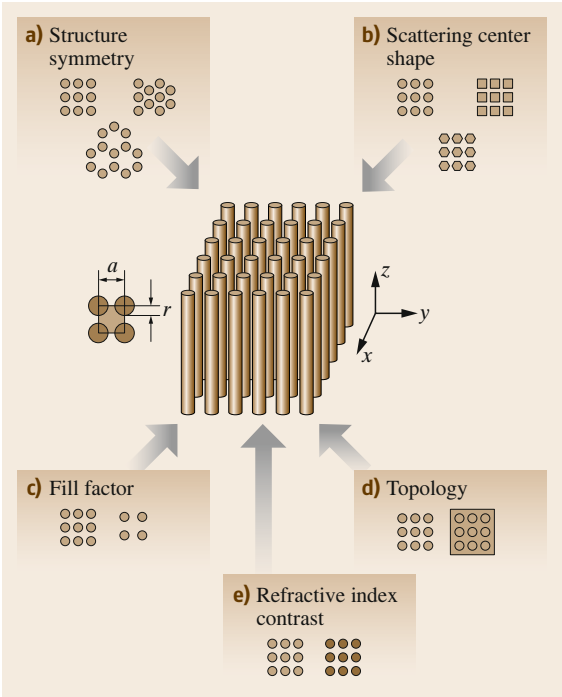


Fig. 39.3 PBG parameters that affect the center frequency and associated stop band (shown in *top view*, in *boxes*) relative to the 2-D periodic arrangement of one material (i. e., the rods, in *gray*) embedded in a second material (i. e., air), periodic in the *x*–*y* plane (shown in the *central figure*)

mined that the minimum dielectric contrast (n_1/n_2) required to obtain a full PBG is about 2.0 [39.2, 12, 14]. For full PBGs, the ideal structure typically consists of a dielectric-air combination, to obtain both the greatest dielectric contrast as well as to reduce losses associated with light propagation in optical materials other than air [39.7].

The relationship between the fill factor, structure symmetry, and scattering element shape on the size and location of the PBG is complex [39.7, 15–18] and will not be discussed in detail here. However, it is clear that the ability to adjust, or tune, one or more of these parameters and thus tune the PC properties, is a very exciting development for future PC device applications.

Real PC structures are of a finite size, and hence differ from the band structure of their infinite counterparts derived by invoking Block's theorem. Instead of having a continuum of allowed modes around the band edges, the finite structures present discrete modes. Developing appropriate models for these structures [39.19] has enabled an understanding of the lifetime of confined modes in specific structures to be achieved. Building on such work enabled engineered high-Q cavities ($\approx 10^6$) with small modal volumes (about $\approx 2(\lambda/n)^3$ for nanowire array cavities) [39.20, 21]. Important to understanding how to design such structures was the realization that for such 3-D PC structures, modal confinement within the PC can be improved by control over the out-of-plane mode profile. In particular, many groups focused on using modes below the light line to satisfy the total internal reflection condition at the interfaces for the top and bottom of the PC slab; however, in such a case, most of the electric field will then concentrate its energy in the dielectric, lowering the modes' frequency as compared with the light line, and decreasing the vertical scattering. To allow for more field energy in the interdielectric space, modes above the light line can provide a better choice. It was found that by a judicious choice of modes, one can identify modes lying above the light line that can still be totally internally reflected and provide for high Q [39.22].

39.1.3 Materials Selection for PCs

The physical architecture of a given PC is just one of the design considerations – another important one is the optical properties of the PC materials. In particular, the refractive index of a given material and its electronic band gap determine the performance and appropriate range of frequencies of PC devices fabricated from such a material.

For large PBGs, component materials need to satisfy two criteria; firstly, having a high refractive-index contrast, and secondly, having a high transparency in the frequency range of interest. This may be a challenge to satisfy at optical wavelengths. Suitable classes of PC materials include conventional semiconductors and ceramics, since at wavelengths longer than their absorption edge (or electronic band gap), they can have both high refractive indices and low absorption coefficients. In addition, these materials have other very useful electronic and optical properties that may complement the functions served by the presence of the PBG.

Another, often overlooked, consideration for PC material selection is the ability to translate the desired bulk single-crystal properties of well-known materials into nanometer-scale PC properties. Such structures typically have extremely large surface areas, and the microstructure of the constituent PC elements must be controlled during fabrication. Consequently, the final properties of the PC elements often vary from those of the bulk properties. This is extremely relevant when considering the functionality of a final PC device, such as those relying on electronic properties (e.g., lasing).

Finally, it is interesting to note that there is a strong correlation between the fabrication methods selected and the component materials. For example, top-down dry etching of semiconductors (e.g., Si, GaAs, InP, etc.)

into nanostructured 2-D arrays is well characterized and relatively commonplace, whereas the same cannot be said for ceramic materials. The opposite is generally true for chemical or bottom-up synthesis using sol-gel technology, in which ceramic PCs are relatively easy to fabricate when sol-gel-based infiltration techniques are used (e.g., in 3-D inverse opal fabrication). It is also clear that the fabrication technique primarily determines the PC structure that can be fabricated, which in turn determines the PC properties.

39.2 Traditional Fabrication Methodologies for Static PCs

Unlike for electrons, where there are natural length scales, for photons there is no such length scale [39.23]. This lack of an absolute length scale ensures that the physics of PCs is scalable: resizing the system rescales the energy in such a way that the spectrum, in units of c/a (where c is the speed of light and a is the lattice parameter of the PC), is independent of the system size [39.23]. Since $\omega/(2\pi c/a) = a/\lambda$, it is customary to use such dimensionless units to measure the photon energy [39.23]. Besides spatial scaling, there is also dielectric scaling – that is, there is no fundamental scale for ε. Two systems whose dielectric functions scale by a factor, $\varepsilon'(\boldsymbol{r}) = \varepsilon(\boldsymbol{r})/s^2$ have spectra that are scaled by the same factor: $\omega' = s\omega$, which means that increasing the dielectric function by a factor of four decreases the photon energy by a factor of two [39.23].

The apparent scalability of PCs is misleading because it suggests that PCs are truly size-independent, which in practice is not the case. As the size of the PC elements decrease, the effect of structural inhomogeneities become increasingly important, leading to inherent size limitations for particular architectures and for particular materials systems. This is most evident in changes in effective refractive index (e.g., due to the formation of intergranular phases, surface oxides, etc.) and functionality.

2-D systems are increasingly being used for microphotonic applications and consist of periodic repetitions of objects in a 2-D arrangement [39.23]. 3-D systems present a modulation in the third direction, such as a stack of spheres [39.23]. To date, 2-D and 3-D PCs have been fabricated using various approaches, including those based on electron-beam lithography, self-assembly and templating. The stringency of the requirements for a fabrication methodology, are dictated by the application of a given PC, in conjunction with associated requirements for material and architecture. Currently, there remains a great need for fabrication methods for mass production of nanometer-scale PCs, offering flexibility in material composition and design, while overcoming the low throughput of serial lithographic patterning, and the low reproducibility and material restrictions of self-assembly methods. Many excellent reports reviewing these fabrication techniques are available in the literature [39.1–4]. For readers unfamiliar with traditional PC fabrication techniques, the following sections review typical 2-D and 3-D PC fabrication methods, based on lithography, self-assembly and hybrid techniques.

39.2.1 2-D PC Structures

Introduction to 2-D PBG Structures

The advantage of the 2-D PC structure is that, due to its planar form, it is easier to fabricate using existing lithographic methods, it retains its original single-crystal properties, and it can be simply integrated into desired PLC designs. However, there are specific design limitations that must be considered for 2-D PC structures. The main problem that has been identified both theoretically and experimentally is diffraction losses in the third dimension (i. e., for light directed out of the plane of the 2-D PC) [39.1, 2].

Calculations often assume a perfect and infinite 2-D material. However, the aspect ratio in experimental structures fabricated using typical micromachining techniques (e.g., reactive-ion etching) tend to be very limited [39.2]. Thus, in order to confine light in the third dimension (i. e., the out-of-plane dimension), the structure must either be large compared to the beam size (i. e., many wavelengths deep), or must be sufficiently confined within a waveguide to experience the full interaction with the periodic lattice [39.1]. As a consequence, 2-D PC structures are now typically fabricated in three principal forms:

1. A 2-D array of air holes in a dielectric (or dielectric pillars of very high aspect ratio in air)
2. A slab waveguide consisting of a 2-D array of holes perforating a thin membrane (with the membrane thickness approximately equal to the hole diameter) and
3. A heterostructure waveguide (i. e., similar to those used for confining light and carriers in semiconductor laser heterostructures) with holes drilled through the heterostructure [39.24].

New fabrication methodologies need to be able to address the fabrication issues for each of these three

types of low-loss 2-D PC structures in order to implement these structures in practical device applications.

2-D Lithographic PC Fabrication Methods

The majority of semiconductor PCs are synthesized using the advanced lithographic and etching capabilities developed for the semiconductor microelectronic industry [39.1, 2, 4, 25]. One major benefit of being able to fabricate PCs from single crystals using top-down lithography is that the properties (e.g., nonlinear electro-optic properties) are typically superior to those for polycrystalline structures produced using bottom-up (e.g., template-based) processes [39.26]. Utilizing the former materials enables significant modulation of the optical properties suitable for developing tunable PC devices. Also, in top-down fabrication, the optical and dielectric properties of the bulk material may be characterized before PC fabrication, allowing them to be more effectively modeled after PC fabrication. Indeed, semiconductor-based PBG structures made using lithographic processes tend to be more reproducible than self-assembled PC structures. Most importantly, there is much greater flexibility in the design and implementation of lithographically fabricated structures, and hence to implementing functional devices. Finally, PC-based PLCs are ideal for optical system integration.

For high refractive-index contrast (i. e., larger photonic stop bands), the preferred PBG structure is a topologically patterned substrate (i. e., an air-solid PC), which may be fabricated using multistep submicron machining based on lithography. In this process, a radiation-sensitive resist is patterned to transfer structures to the bulk material. The patterned resist is developed (i. e., the properties of the exposed material are changed), followed by etching and/or metal thin-film deposition to define features in the bulk material. The majority of submicron machining processes involve exposure of the resist with either electromagnetic radiation (e.g., optical, UV or X-ray photons) or charged particles (electrons, low-energy heavy ions, high-energy light ions) [39.27]. Although resist exposure is essentially a surface micromachining technique and therefore planar, various wet and dry etching techniques have been successfully used to produce topologically varying microstructures in the axial direction [39.27].

Similar to all resist-based processes, the resist itself limits the ultimate minimum size and maximum density of nanostructured patterns [39.28, 29], and organic residues (from processed resist) are often sources of contamination [39.30]. Organic resists are chosen as a compromise between sufficient resilience for selected postprocessing procedures, and having sufficiently high sensitivity and contrast (i. e., a low energy threshold for resist exposure and total resist exposure occurring only above a certain energy threshold with no resist exposure occurring below, respectively) to the incident exposure [39.31] – this tends to be a difficult compromise to satisfy in practice for traditional resist materials.

Lithographic methods provide a simple and straightforward approach for the intentional introduction of defects into these 2-D systems (due to the simple surface patterning), and this top-down method may be the only way to retain the bulk-like semiconducting properties that are required for certain functionalities (e.g., lasing).

Electromagnetic Radiation Patterning. Generally, using electromagnetic radiation to micromachine structures requires masks to expose a resist material in a spatially selective manner, with subsequent development of the exposed resist to produce microstructures. Masks typically incorporate multiple repeat patterns, enabling multiple microstructures to be fabricated in a single exposure, which significantly reduces costs [39.27]. Masks tend to be less useful for patterning with charged particles, since the high energy deposited into the mask during exposure can produce undesirable mask instabilities due to heat expansion, stress and damage [39.27]. Charged-particle micromachining is therefore normally limited to direct-write processes, where a focused charged-particle beam is scanned over a material in a specific pattern to produce microstructures. Such a direct-write process has the advantage of not requiring a mask, but suffers from relatively low throughput as a serial write process, and hence loses its attraction for multiple-component production [39.27].

A main consideration of patterning approaches for PC fabrication is the minimum required feature size. Diffraction effects occurring for wavelengths of light below 100 nm impose a fundamental limitation on achievable structure sizes using optical lithography. Micromachined features smaller than about 250 nm are essentially beyond current readily available optical lithography systems [39.27], although recent advances (e.g., in UV lithography) will probably make this size regime achievable in the near future using a parallel patterning processes. Other techniques such as X-ray lithography that are capable of high-volume production by allowing parallel exposure for larger-area nanometer-scale patterning, are hampered by challenges including making suitable high-resolution masks, as well as the significant capital cost required to achieve the high photon flux suitable for mass production of nanometer-scale components [39.27].

Charged-Particle Beams. Limitations on the minimum feature size using traditional photolithography

have motivated a number of studies using serial patterning with charged-particle beams (e.g., electron-beam lithography (EBL) or, less commonly, focused ion-beam (FIB) lithography) to obtain the small feature sizes required in PC structures.

The advantages of EBL and FIB for defining sub-100 nm feature sizes in a maskless fashion are offset by the fact that high-volume production is commercially impractical using such serial writing techniques [39.27, 32, 33]. Conventional ion implantation, though widely available and robust, has thus far been incapable of patterning densely organized nanostructures due to ion scattering [39.31, 34] and mask limitations [39.35–39].

Another consideration for choosing lithographic techniques for sub-100 nm-scale feature fabrication is that all the steps in this multistep process have their own limitations. Indeed, the finite size of the molecular compound comprising the resist sets the ultimate limit on pattern resolution [39.28, 32]. Since all traditional submicron micromachining techniques except FIB require resist exposure, this is a major factor for determining resolution – for example, the ultimate resolution of EBL is not determined by the electron optical system, which is ≈ 0.1 nm, but by the resolution of the resist and by the subsequent fabrication process [39.40]. Also, electrons in the beam scatter very easily from electrons in the resist material, and both scattered and secondary electrons cause a lateral spread in the patterned material due to diffusion effects (i. e., proximity effects) [39.27]. Thus, the very high spatial resolution achievable at the surface using EBL deteriorates as the beam penetrates into the resist, making EBL unsuitable for high-resolution topographically patterned structures using resist alone [39.27]. The advantage of maskless processing for high-definition surface structures in resist and the inherently slower processing mean that EBL has been limited to mask-making and direct-writing on wafers for specialized applications such as fabrication of 2-D PCs for research purposes only [39.27].

Currently, the most commonly used method to fabricate 2-D PCs in the optical regime is EBL-based mask processing, combined with etching [39.1]. This approach has been used to pattern different semiconductor materials, including Si [39.4] and GaAs [39.1, 4, 7] (Fig. 39.4). Some detrimental effects arising from dry etching have been reported, including damage and ion channeling that increase nonradiative recombination in active microstructures [39.1]. Other post-mask procedures used to fabricate 2-D PC structures include vertical selective oxidation and lift-off techniques [39.1]. In vertical selective oxidation, an *all solid* 2-D PC structure may be fabricated, which maintains the physical integrity of the crystal and allows deposition of electrical contacts afterwards. However, the oxidation process

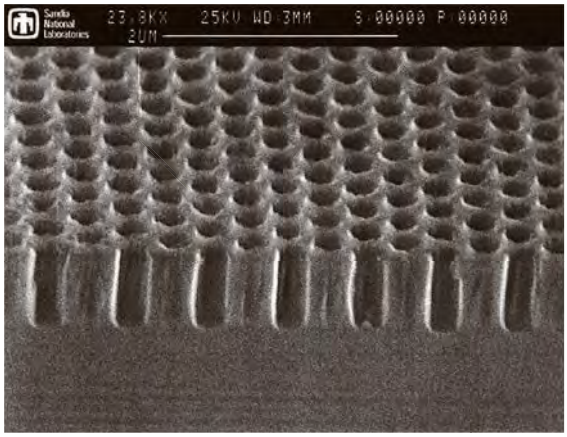


Fig. 39.4 Example of a 2-D PC structure in GaAs. (After [39.7]). The air hole radius is 122.5 nm, the lattice parameter is 295 nm, and the holes are 600 nm deep

usually also occurs laterally, while typical desired configurations for 2-D PC crystals are structured in the vertical direction. Even with improved vertically selective oxidation techniques, the depth-to-diameter aspect ratios of pillars are around 1 : 1, which is significantly less than the 10 : 1 ratios achieved using dry etching techniques [39.1]. Also, volume shrinkage of up to 13% [39.1] can lead to strain, the formation of microcracks and other structural problems. 2-D PC membranes have been fabricated using epitaxial lift-off/substrate removal, but this method is limited to materials with sufficient wet-etch selectivity [39.1].

Despite its extremely low throughput, focused ion beams are presently used for direct, resistless, nanometer-scale surface modification, micromachining, and ion implantation. This is primarily because such patterning cannot be achieved with any other method. However, FIB patterning is inherently slow and expensive, and requires sophisticated optics for ion focusing. Also, the patterning resolution and proximity of nanometer-scale features is greatly limited by the significant FIB tail distribution [39.30, 41] and the highly dose-dependent sputtering yield of FIBs [39.42]. The use of FIB patterning has not been widely used to fabricate PCs.

Light ions collimated by stencil masks can be used to expose resists in a parallel fashion in both ion-projection lithography and masked ion-beam lithography [39.31]. The resist is exposed in a similar manner to X-ray and EBL. Light ions are generally preferred over heavier ions in ion lithography because they have the greatest range in resists (thick resists can be used to give low defect densities [39.35]), but protons also scatter more than heavier ions at the end of the implant range, limiting resolution and nanometer-scale

Table 39.1 Summary of current submicron resist-based machining techniques. (After [39.31])

Technique	Potential sub-100 nm capability	Requires mask	Direct write	Facility availability	High-volume production capability	3-D capability
Optical lithography	No	Yes	No	Widespread industrial use	Yes	No[a]
X-ray lithography	Yes: depends on mask	Yes	No	Scarce and expensive	Yes	Yes
UV lithography	Yes: depends on mask	Yes	No	Scarce	Yes	Yes
Electron-beam writing	Yes: currently achieved	No	Yes	Medium availability	No	No
Low-energy ion beam writing	Yes: depends on focusing	No	Yes	Scarce	No	No[b]
High-energy ion beam writing	Yes: depends on focusing	No	Yes	Very scarce	No	Yes

[a] Optical lithography is a surface technique. However, new supplementary dry etching techniques such as plasma etching are being used to produce submicron 3-D structures from 2-D surface structures in silicon
[b] Low-energy ion-beam micromachining (FIB) relies on sputtering atoms from the surface. 3-D structures can be produced by continuous erosion at a given location although, in practice, this is far too slow for practical 3-D microcomponent production

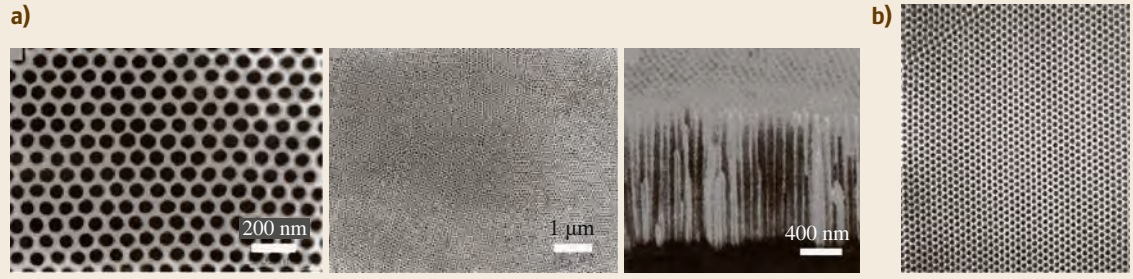


Fig. 39.5a,b Examples of 2-D self-organized PC structures of: **(a)** NCA [39.43], **(b)** NCG. (After [39.44])

feature proximity. However, the use of heavier ions to pattern, following the traditional approach used for semiconductor doping, can lead to mask erosion. This effect is negligible when light ions (e.g., hydrogen) are used, but can become pronounced for higher-dose implantation of heavier ions [39.31]. A summary of current lithography-based submicron micromachining techniques is given in Table 39.1.

2-D Self-Assembled PC Fabrication Methods

For 2-D PC structures, electrochemical methods (i. e., anodic etching and growth) and fiber-pulling have used to prepare extremely high-aspect-ratio ($\gg 50:1$) large-area nanometer-scale hole arrays (Fig. 39.5). These types of structures represent near-ideal infinite 2-D PC structures [39.24]. However, these methods are limited to specific materials (e.g., silicon [39.45] and alumina [39.43] for electrochemical methods, or glass [39.44] by fiber-pulling), and processes such as electrochemical etching are often limited in their capability to control fill factor, structure symmetry and scattering element shape. 2-D PC structures with long-range periodicity obtained by electrochemical etching often require surface prepatterning by serial lithography to obtain lattice dimensions in the visible to near-IR range.

Some examples of easily available, high-aspect-ratio nanoporous arrays include electrochemically grown porous alumina (nanochannel alumina, or NCA) [39.43], electrochemically grown silicon [39.46] and fiber-pulled nanochannel array glass (NCG) (Table 39.2) [39.47, 48].

Nanochannel Alumina. NCA may be fabricated economically with well-characterized properties. Pseudo-ordered regions can be easily fabricated using electrochemical etching, compared with porous silicon, which requires back-side illumination in addition to an applied

Table 39.2 Parameters of current high-aspect-ratio 2-D periodic structures. (After [39.46–48, 50])

Nanotemplate type	Alumina	Silicon	Glass
Pore size (nm)	$1-10^2$	$400-10^4$	> 10
Pore density (cm^{-2})	10^9-10^{12}	10^6-10^9	10^{10}
Pore arrangement	Pseudo-ordered[a]	Disordered[a]	Ordered
Thickness (nm)	10^3	10^3	10^3
Temperature tolerance	$< 450\,^{\circ}C$	$< 900\,^{\circ}C$	$< 600\,^{\circ}C$

[a] May be ordered using prepatterning processes [39.46, 49]

bias. The pore arrangement is completely disordered in silicon if no prepatterning with optical or EBL is used [39.50].

Aluminum is electrochemically oxidized to alumina (Al_2O_3) under positive polarization. For over a century, the growth of disordered pore arrangements has been observed and studied for selected electrolytes that weakly dissolve alumina. However, in 1995 it was first discovered that, after an extended period of anodization, self-ordered porous alumina films with pores arranged in a hexagonal pattern are obtained at the growth front [39.43]. Although knowledge of how to fabricate such NCA templates is now widespread, the formation mechanism of hexagonal NCA has still not been completely elucidated. Mechanical stress between neighboring pores due to the volume expansion of alumina with respect to the aluminum substrate has been proposed as a mechanism for the self-ordering [39.50].

Controlled single-domain porous arrays may be obtained by lithographically prepatterning the aluminum substrate, and applying the optimum potential for the corresponding interpore distance [39.51]. Recently, in addition to the hexagonal lattice, square and honeycomb lattices with square or triangular pore shapes have also been obtained by appropriate prepatterning [39.52].

Extremely high-aspect-ratio NCA templates can be fabricated with periodicities ranging from ≈ 80 to $\approx 600\,nm$. The domain sizes (i. e., areas with the same orientations) are controlled by the anodization conditions and increase with time to micrometer-scale sizes. The thickness of the NCA template can be controlled by the electrochemical etching time, from $\approx 50\,nm$ to $> 10\,\mu m$ thick.

Current Alternative 2-D PC Fabrication Methods. Templating of 2-D Structures. 2-D PCs have been fabricated by templating a secondary material into a PC backbone, that can either remain in the final structure as a composite material or be removed to form a dielectric/air PC structure. In many cases, a dielectric/air PC structure is preferred due to the higher dielectric contrast, but depending on the original PC backbone material, it may not be possible for the secondary materials to be removed after templating.

Over the last decade NCA templates have been commonly used to fabricate various nanostructured arrays by either dry etching or by growth through templates [39.49, 53, 54]. This method of using self-assembled templates for nanostructure patterning has been used to pattern physical topology, resulting in either nanostructured hole arrays (e.g., by dry etching through the templates), or growth of nanostructured *wire* arrays (e.g., by metalorganic chemical vapor deposition – MOCVD, molecular-beam epitaxy, and chemical-solution deposition). The effectiveness of NCA as a templating tool results from the fact that the alumina may be easily selectively chemically etched compared with semiconductor materials, and forms a robust mask for dry etching (compared to, for example, patterned resist materials).

Combining Self-Organized 2-D Structures with Ion Implantation. Ion-beam techniques may be used to pattern, for example, sub-100 nm resolution features on thin membranes, but they are unsuitable for deep subsurface patterning, particularly at the end of the implantation range. One reason for this is that stencil-type masks need to be used, due to the relatively short range of energetic ions compared to, for example, x-rays, which do not require stencil-type masks [39.28, 31]. In the past, high-aspect-ratio stencil masks for ion implantation were fabricated using techniques such as EBL, but such stencil masks were limited by the low pattern density achievable using such fabrication methods, requiring the use of self-complementary masks with multiple offset exposures to create printed patterns [39.38]. Also, the minimum resolution using nanoscale ion implantation through masks is so sensitive to the mask's wall angle [39.36, 37] that advanced photoresists and lithographic techniques need to be used to achieve such ideal masks.

Investigations have been reported on new ways to combine self-organized 2-D PC structures with conventional, broad-beam ion implantation to achieve nanoscale pattern transfer [39.55, 56]. Using self-organized templates as masks significantly simplifies

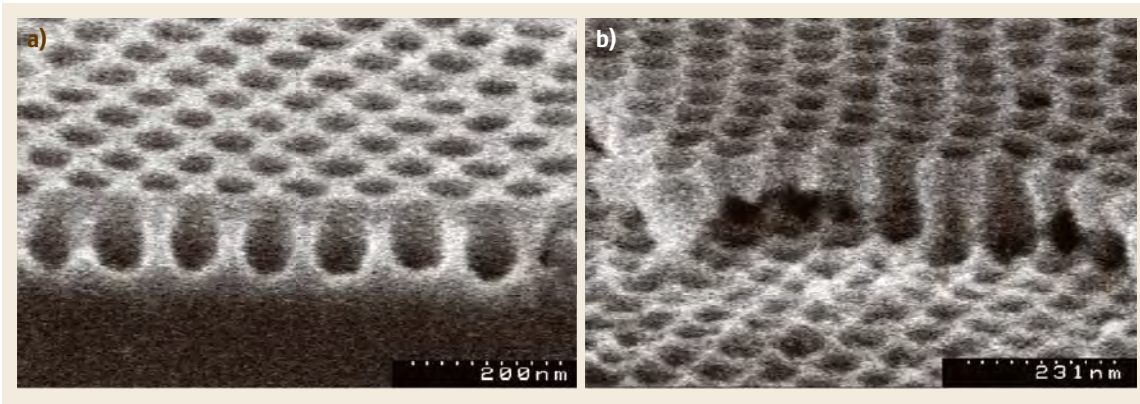


Fig. 39.6a,b Example of nanoscale features formed in a $SrTiO_3(100)$ substrate using conventional ion implantation through high-aspect-ratio masks combined with selective wet etching. (After [39.55, 56]). Scanning electron microscope images for: **(a)** tilt view of single implant and **(b)** multiple energy implants to create membrane structure

the process, as no resolution-limiting resist or no time-intensive serial patterning is necessary, and these ideal, stencil-type masks allow direct nanostructure array transfer with significantly less scattering than solid masks. The use of self-organized masks allows this process to be economically feasible for large-area nanostructure array patterning. This approach transfers the desired pattern from the self-organized template into a desired material that cannot be patterned on the same scale using self-organization.

Despite concerns about mask edge defects [39.57–59], extended radiation damage accumulation extending below the mask edges [39.34, 39] and possible mask distortion from stress and temperature [39.35], it has been shown that nanoscale features could be retained using conventional ion implantation combined with high-aspect-ratio masks to make PC-type structures [39.55, 56] (Fig. 39.6). Energetic, charged particles were accelerated into a substrate using broad-area parallel implantation through NCA templates to pattern 2-D nanostructured arrays of ions in both semiconductor (i. e., InP) and ceramic (i. e., $SrTiO_3$) single crystals [39.55, 56]. The ion-implanted volume was selectively etched away after implantation to form 2-D nanostructured air-hole arrays in the substrate material. The array pattern was successfully fabricated over the entire implant area ($\approx 1\,cm^2$), with resulting minimum feature sizes of $\approx 40\,nm$. Such a direct and parallel method, combining ion implantation with a nanoscale mask, can be immediately beneficial, particularly in applications where using FIB is impractically slow (e.g., for substrate nanostructure array prepatterning for selective or electrochemical etching). In addition, complex profiles (e.g., membrane formation) were shown to be possible using this technique. This method allows ions to be implanted into nanoscale areas, which may result in additional substrate modification, including, but not limited to, micromachining of implanted areas to fabricate topologically patterned nanostructured arrays.

The ability to process structures of different form and shape using ion-implantation techniques offers opportunities for nanomachining, and enables material processing on a nanometer scale – this is applicable to materials that cannot be structured using standard lithography (e.g., samples with pronounced topography [39.42]). This nanoscale patterning method can also be used to directly pattern substrates to create electrochemical initiation sites for fabrication of ordered porous Si and porous GaAs structures for light emission [39.60] and for 3-D PC fabrication in Si [39.61]. In addition, such array pattern transfer is applicable to prepatterning of self-assembled templates themselves, as a substitute for lithographic prepattering for the fabrication of ordered nanochannels in materials such as Si and Al_2O_3 [39.50].

39.2.2 3-D PC Structures

Introduction to 3-D PBG Structures

A main advantage of 3-D PBG structures over 2-D PBG structures is that the 3-D PBG structures can overcome the problem of diffraction losses in the third (out-of-plane) dimension [39.1, 2]. However, this advantage is countered by the significant challenges associated with fabricating 3-D PBG structures. Fabricated 3-D structures reported to date lack the reproducibility and design versatility required for functional devices [39.25].

3-D Lithographic Fabrication Methods

Fabrication of 3-D PCs using top-down micromachining techniques has been more difficult than for 2-D PCs.

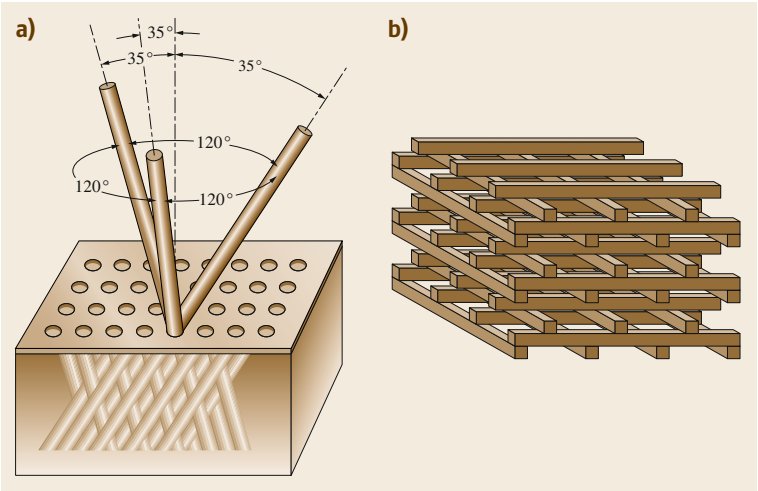


Fig. 39.7a,b Examples of 3-D photonic crystal structures: **(a)** *Yablonovite*, obtained by drilling holes in a homogeneous dielectric material (15 GHz); and **(b)** *wood-pile* structure, obtained by stacking micromachined Si rods (500 GHz). (After [39.2])

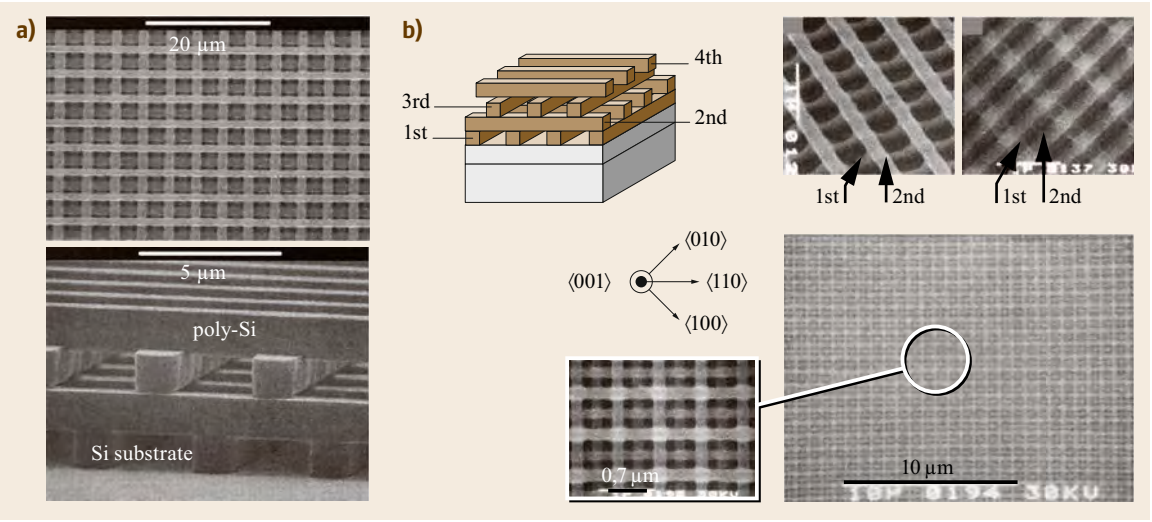


Fig. 39.8a,b Examples of 3-D PBG structures fabricated by micromachining/lithographic techniques: **(a)** advanced lithography. (After [39.62]). **(b)** wafer fusion. (After [39.63])

The first experimental demonstration of a full 3-D photonic band gap was performed in the microwave region (i. e., at 15 GHz, $\lambda \approx 2$ cm), due to the relative ease of fabrication of appropriate structures on the centimeter scale [39.64] (Fig. 39.7). Another case of a full PBG was demonstrated at 500 GHz ($\lambda \approx 600\,\mu$m) with a 3-D structure fabricated by assembling micromachined silicon dielectric rods [39.65] (Fig. 39.7) into log-pile structures [39.66, 67].

In a 3-D lattice, theoretical calculations have determined that the band gap position is generally about one half to one third of the periodicity scale [39.68]. This translates into patterning the dielectric materials with features of 100–300 nm in all three dimensions, with required alignment accuracies among the different layers exceeding tens of nanometers [39.68]. In addition, with air as the second dielectric, the empty volume required for large PBGs is so large ($\approx$ 70–80%) that the mechanical stability of the structure becomes very poor, adding another challenge for the etching procedure [39.2]. Often 3-D PC structures are so full of defects after such etching that the systems no longer exhibit significant PC behavior [39.2].

Due to difficulties in fabricating structures in the optical or near-infrared frequencies [39.2, 68, 69], only a few experimental systems have recently demonstrated the signature of a complete PBG. Lithographically derived 3-D structures require many complex fabrication steps, and it takes many months to make a single structure. Some examples of prototype 3-D PCs in-

clude log-pile structures fabricated using layer-by-layer patterning (i. e., repeated deposition and etching) of Si [39.62] (Fig. 39.8), and structures fabricated using wafer fusion combined with etching in GaAs and InP [39.63] (Fig. 39.8). In the former case, processing difficulties have limited the number of layers in such 3-D periodic structures (i. e., to below ten) and in the latter cases, to materials other than those commonly used in microelectronics [39.68]. Sequential deposition of dielectric films on a patterned substrate may be used to produce 3-D periodic structures in the *autocloning* process [39.70], which is considered to be a relatively cheap and flexible process. However, the range of structures that can be produced with this technique is limited. Also, the technique does not allow the component dielectrics to be continuously connected in 3-D [39.71].

Hole-drilling has also been implemented using x-ray lithography [39.72] and can be scaled to optical frequencies by directional etching through a mask on a semiconductor surface [39.73]. Other fabrication methods include direct laser writing, but its resolution is currently only suitable for infrared structures.

3-D Self-Assembly Methods: Inverse Opals

The technological challenges involved in extending microlithographic techniques to 3-D patterning have made self-assembly methods attractive for 3-D PC fabrication. The benefits of using this type of bottom-up assembly for 3-D fabrication include its low capital cost, and its parallel and rapid nature, which allows the ordered fabrication of 3-D PBG-type structures over macroscopic length scales [39.3, 68, 77, 78]. However, self-assembly methods for 3-D PC fabrication are often material-specific, and templated bottom-up PCs tend to have diluted properties compared to bulk single crystals, typically containing unintentional defects and lacking reproducibility.

Self-organized 3-D colloidal crystal structures in a face-centered-cubic lattice (i. e., artificial opals) can be fabricated through sedimentation of colloidal suspensions (i. e., typically spheres, but other structures, such as micelles formed from rod-coil block copolymers have been used). Previously, colloidal crystals were not considered appropriate for PCs with band gaps in the visible and IR ranges because the colloids (typically SiO_2 or polymer materials such as polystyrene (PS) [39.79, 80] or polymethyl-methacrylate (PMMA)) have refractive indices ≈ 1.5, which are too low to obtain a full PBG [39.81–83]. Such structures have weak refractive-index contrast, but theoretical calculations have shown that the final porosity of the artificial opal is too small to ever obtain a full PBG, even with very high refractive-index contrast [39.1, 3, 77].

These colloidal structures may also be infiltrated with a second material of higher refractive index to potentially improve the dielectric contrast [39.81, 84], while providing the required dimensional control and low solid fraction [39.78]. So-called *inverse* structures can be formed when the original colloidal structure is removed to leave a macroporous structure. Calculations have shown that these inverse structures give rise to larger PBGs for the same refractive-index contrast [39.85]. The porous materials obtained using this approach have also been referred to as *inverse opals* or *inverted opals* because they have an open, periodic 3-D framework complementary to that of the parent opal. Modeling suggests that a porous material consisting of an opaline lattice of interconnected air balls (embedded in an interconnected matrix with a higher refractive index) should produce a full PBG. Optimum photonic effects require that the volume fraction of the matrix material should be 20–30%. Template-based synthesis based on colloids is attractive because the periodicity of this system can be conveniently tuned and a wide variety of materials with relatively high refractive index can be used [39.77]. There have been many examples of inverse opals using materials such as carbon, titania, zirconia, Si and Ge reported in the literature [39.3, 74–78, 86–93] (Fig. 39.9).

The inverse structure is made by infiltrating the interstitial void spaces ($\approx 26\%$ by volume) between the colloidal spheres in the opaline array with a fluid precursor. The fluid may be a pure liquid (e.g., a liquid metal alkoxide or a molten metal), a solution (e.g., a salt solution), a vapor (e.g., using CVD – chemical vapor deposition), or a colloidal dispersion of nanocrystals (e.g., colloidal gold particles) [39.77, 78]. Alternately, suspensions containing nanoparticle precursors (with sizes of ≈ 1–$50\,\mathrm{nm}$) and monodisperse spheres can be coprecipitated into an ordered structure [39.77, 78]. The void spaces between the colloidal spheres has also been filled using a variety of materials by electrochemical deposition [39.77]. The fidelity of this procedure is mainly determined by van der Waals interactions, the wetting of the template surface, kinetic factors such as the filling of the void spaces in the template, and the volume shrinkage of precursors during solidification [39.77].

The choice of wall material depends on the desired optical functionality of the PC, as its properties depend on the transparency region of the dielectric material. The availability of a precursor that can infiltrate the voids between the colloidal spheres without significantly swelling or dissolving the template is critical. For example, high-temperature infiltration techniques exclude the use of polymer templates. The precursor must be carefully chosen to allow for shrinkage occur-

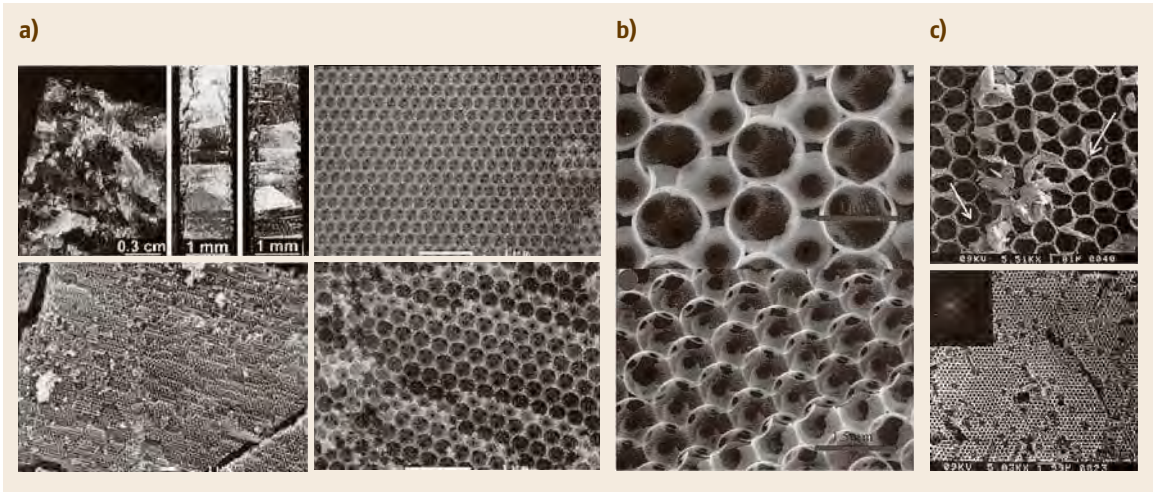


Fig. 39.9a–c Examples of 3-D inverse opal PBG structures: (a) graphite (using CVD infiltration). (After [39.74]), (b) Si (using CVD infiltration). (After [39.75]), (c) titania (using liquid-phase chemical reaction). (After [39.76])

ring during solidification and the final grain size, which affects the smoothness, density, effective refractive index, PC wall structure and the mechanical properties of the final 3-D PC [39.78]. The approach and conditions used for infiltration can affect the filling efficiencies (i. e., the effective wall density) which, in turn, affect the mechanical strength and reduces the average refractive index of the walls [39.78, 86]. Choice of template also depends on the method of solidification of the precursor and whether the template can be selectively removed from the wall material, either chemically or by heating during the final fabrication step.

Porous structures are produced by the removal of the colloidal templates, either by burning, extracting with a solvent, or, in the case of silica colloids, by dissolving with dilute hydrofluoric (HF) acid. The method of template removal depends on the colloid material and on the properties of the infiltrated precursor. Silica sphere template arrays are removed by dissolution in aqueous HF solutions [39.78, 92]. Most metal-oxide inverse opals have typically been synthesized using polymer spheres as templates [39.78]. Polymer sphere template arrays are often removed thermally, with the added benefit of chemically converting the sol-gel precursors in the infiltrated phase to the desired oxide product [39.78, 92]. If precursor solidification is feasible at low temperatures, or if the precursor has components that would be destroyed by thermal template removal, polymer spheres can also be removed chemically using organic solvents, such as toluene, or tetrahydrofuran (THF) and acetone mixtures, or by photodegradation [39.78, 92].

Fabrication of such structures is not trivial: some of the difficulties include incomplete infiltration, structural shrinkage after infiltration, and microstructural variations (affecting the refractive index). A major problem with chemical assembly is that residual disorder drastically reduces the PBG width and the intensity of the reflectance peaks [39.78, 89]. Such disorder can arise from several different sources. Firstly, there are defects in the initial opal template, such as stacking faults or small deviations in sphere size or shape. It is not clear whether the imperfections present in current inverse opals are small enough to observe a complete PBG [39.3]. While progress has been made in reducing such defects, numerical simulations predict that even a 2% deviation in sphere size (or lattice constant) can close a PBG in even high-refractive-index inverted opals [39.3, 78]. Secondly, significant disorder can arise from the infiltration process itself. Even the best samples to date have variations in the amount of guest material filling the lattice [39.3]. Since these inhomogeneities must be much smaller than the optical wavelength, the elimination of residual order remains a challenge with such methods [39.3].

Although 3-D inverse opals are fascinating research objects, they are not amenable for practical photonic applications because of difficulties in integrating them into PLCs. Thus, there is strong interest in fabricating 3-D inverse opals in pseudo-2-D form, or supported on a substrate as an inverse opal film, to be compatible with traditional optoelectronic elements for system integration [39.1, 90, 95–97]. Recently, ordered colloidal crystal films have been formed by convective assembly [39.90] by particle confinement in thin slits between two solid plate boundaries [39.77, 90] and by compacting particles in the dry state [39.90, 95, 97–100]. Recently, patterned substrates were successfully used to

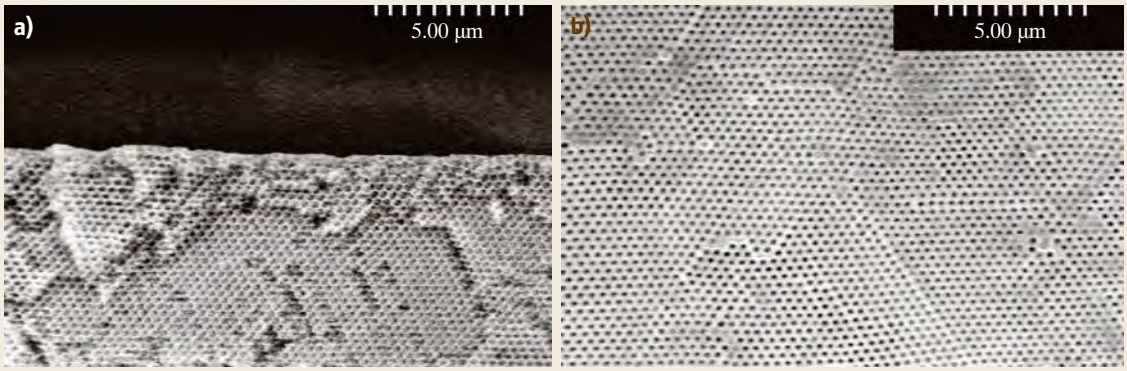


Fig. 39.10a,b Scanning electron microscope images of Pb-doped BST inverse opal thin film: **(a)** cross section; **(b)** top view. (After [39.94])

control slow sedimentation of colloidal particles to improve the long-range order and to induce growth along a specific crystal plane [39.78, 90]. However, growth of defect-free crystalline templates with desired symmetry and orientation remains one of the challenges of the field.

Liquid sol-gel precursors can be used to form a variety of oxide materials and are very convenient for infiltration into artificial opal templates [39.92, 101]. However, the typical high percentage of volume shrinkage (typically 15–30%) involved in the solidification of such sol-gel precursors during the precursor-to-oxide conversion process has often resulted in random cracks in the final inverse opal films [39.77, 78, 92], often preventing the fabrication of large, ordered, inverse opal film structures. Colloidal template films have been infiltrated by soaking the template into the precursor sol-gel solution [39.78, 89, 102], but achieving a dense, complete filling of the void spaces within the 3-D template film has proven to be challenging [39.101]. Another approach used for large-area inverse opal film fabrication involves the crystallization of colloids in a liquid medium containing nanoparticles of the desired material, which are subsequently packed into a dense, 3-D lattice [39.78, 101]. The main limitation of this method is that nanometer-sized particles are not always available for the desired wall material, and so only very specific materials may be made into inverse opals [39.92]. Another method of fabricating high-quality inverse opal films derived from liquid precursors is by infiltrating a sol-gel liquid precursor into a colloidal crystal template film by spin coating, which results in very uniform filling of the template [39.94] (Fig. 39.10). By comparison, solvent-extracted samples shrink about 10% less [39.92]. Thus, for organic polymers, it has been possible to fabricate such 3-D porous structures as free-standing thin films that are several square centimeters in area.

Alternative 3-D PBG Fabrication Methods

Alternative approaches to lithography or self-assembly for 3-D fabrication exist. Some notable examples include laser-guided stereolithography, in which structures can also be written directly layer-by-layer by using a tightly focused laser beam to cause chemical vapor deposition [39.103], and two-photon-induced deposition or polymerization [39.104, 105]. Currently, feature sizes that can be routinely achieved using these methods are on the order of a few micrometers due to focusing and diffraction effects.

A 3-D holographic patterning process using multiple laser beams has recently been investigated; a 3-D laser interference pattern exposes a photosensitive polymer precursor (i. e., a photoresist) rendering the exposed areas insoluble and unexposed areas are dissolved away to leave a 3-D PC formed of cross-linked polymer with air-filled voids [39.106]. This process allows flexible design of the structure of a unit cell and thus of the optical properties of the microstructured material. Of concern is that such structures have relatively weak refractive-index contrast, and the processing approach is quite demanding [39.68].

39.3 Tunable PCs

Although conventional PCs offer an ability to control light propagation or confinement through the introduction of defects, once such defects are introduced, the propagation or confinement of light in these structures is not controllable. Thus, discretionary switching of light, for example, or rerouting of optical signals, is not available with fixed defects in PCs. There are two approaches that have been pursued to tune the properties of PCs: these are by tuning the refractive index of the constituent materials, or by altering the physical structure of the PC. In the latter case, the emphasis has been principally on changing the lattice constant, although other approaches are also relevant (i. e., including for example, the fill factor, structure symmetry and scattering element shape). In the text below, we discuss the state of the art in both of these areas.

39.3.1 Tuning the PC Response by Changing the Refractive Index of the Constituent Materials

We discuss recent progress in using four approaches within this category – namely, tuning the PC response by using: (i) light, (ii) applied electric fields, (iii) temperature or electrical field in infiltrated PCs, and (iv) by changing the concentration of free carriers (using electric field or temperature) in semiconductor-based PCs.

PC Refractive-Index Tuning Using Light

One approach to modifying the behavior of PCs is to use intense illumination of the PC by one beam of light to change the optical properties of the crystal in a nonlinear fashion; this, in turn, can thus control the properties of the PC for another beam of light. An example of a nonlinear effect includes the flattening of the photon dispersion relation near a PBG, which relaxes the constraints on phase matching for second- and third-harmonic generation [39.107]. In addition, light location near defects can enhance a variety of third-order nonlinear processes in 1-D PCs [39.108].

A new approach for PBG tuning has been proposed based on photo-reversible control over molecular aggregation, based on using the photochromic effect in dyes [39.109]. This can cause a reversible change over the photonic stop band. Structures that were studied included opal films formed from 275 nm diameter silica spheres, infiltrated with photochromic dye. Two dyes were considered – namely 1,3-dihydro-1,3,3-trimethylspiro-[2*H*-indole-2,3′-[3*H*]-naphth [2,1-*b*] [1,4]oxazine] (SP) and *cis*-1,2-dicyano-1,2-bis(2,4,5-trimethyl-3-trienyl)ethane (CMTE). For the SP dye opal, a reversible 15 nm shift in the reflectance spectrum was observed following UV irradiation, which was ascribed to changes in the reflective index due to resonant absorption near the stop band. Smaller shifts (of about 3 nm) were observed for the CMTE dye opal. The recovery process on cessation of UV illumination was quite slow, taking about 38 s for the SP dye.

Nonlinear changes in refractive index have also been studied in PCs consisting of 220 nm diameter self-assembly polystyrene spheres infiltrated with water [39.110]. In these studies, the optical Kerr effect was used to shift the PBG. 40 GW/cm^2 [39.2] of peak pump power at 1.06 μm (35 ps pulses at 10-Hz repetition rate) was used to shift the PBG 13 nm. The large optical nonlinearity originates from the delocalization of conjugated π-electrons along the polymer chains, leading to a large third-order nonlinear optical susceptibility. The time response was measured as a function of the delay time between pump and probe, and confirmed a response time of several picoseconds.

Further to the discussions in Sect. 39.1.2 on high-Q nanowire-based PCs using linear optical field interactions, it was found that such structures have an additional benefit for nonlinear optical response. Their highly anisotropic optical response originates from the dielectric mismatch between the nanowire material and surrounding medium, introducing strong depolarising effects, which introduce anisotropy in the optical electric field in the nanowires, and make the probability of optical transitions strongly dependent on direction, and has led to a giant nonlinear response being observed [39.111, 112].

PC Refractive-Index Tuning Using an Applied Electric Field

There have been a number of studies focusing on using ferroelectric materials to form PCs [39.94, 102, 113, 114]. The application of an electric field to such materials can be used to change the refractive index and tune the PC optical response. A recently published review of electrically responsive PCs can be found in [39.115]. For example, lead lanthanum zirconate titanate (PLZT) inverse opal structures have been fabricated by infiltration using 350 nm diameter polystyrene sphere templates and annealing at 750 °C [39.113]. The films were formed on indium tin oxide (ITO) coated glass to enable electric-field-induced changes in reflectivity to be measured due to the electro-optic effect. Applying voltages of up to about 700 V across films of thickness of about 50 mm only achieved a few nm of peak shift, attributed to the very modest changes in refractive index from the applied field (i. e., from 2.405 to 2.435, as a re-

sult of the bias). It should be noted, however, that the intrinsic response of the electro-optic effect in these materials is known to be in the GHz range and hence highly suitable for rapid tuning. Other reports include the formation of inverse opal barium strontium titanate (BST) PCs using infiltration of polystyrene opals [39.94, 102]. BST is most interesting in that it provides a high-refractive-index material and a factor of at least two times higher breakdown field strength than PLZT, and hence offers a much wider range of applied fields for tuning. Reports of using other ferroelectric materials include a high-temperature infiltration process for the ferroelectric copolymer, poly(vinylidene fluoride-trifluoroethane), infiltrated into 3-D silica opals with sphere diameters of 180, 225 and 300 nm [39.114].

Refractive-Index Tuning of Infiltrated PCs

In this case the approach has been to consider modulation of the refractive index of a PC infiltrated with a tunable medium – in particular, the most popular approach has been to use liquid crystals to infiltrate porous 2-D and 3-D PC structures. Such liquid crystals can behave as ferroelectrics whose refractive index may be tuned using either an applied electric field, or by thermal tuning. Reported results for this approach, using 3-D inverse opal structures, have been restricted to infiltration of ferroelectric liquid-crystal material into a silicon/air PC [39.116]. Reported changes in the refractive index of the liquid crystal are 1.4−1.6 under an applied field [39.116]. However, since the ferroelectric liquid has a higher refractive index than air, the infiltration results in a significant decrease in the refractive index contrast. This means that the original full photonic band gap of the inverse opal silicon PC no longer exists, and the practical utility of the silicon structure as a PC is effectively lost. Theoretical simulations have shown that partial surface wetting of the internal inverse opal surface can retain the full photonic band in silicon [39.116], but it is questionable whether such a complex structure can ever be practically fabricated. Finally, the presence of ferroelectric liquid crystal surrenders one of the main advantages of the original concept for PC structures: that is, permitting light propagation in air [39.117]. Temperature tuning of the liquid-crystal material was shown to result in very small changes in refractive index (changes in $n < 0.01$ over a 70 °C change) and thus only provide minimal shifts in the transmittance through the PC over a large temperature range [39.118]. Aside from opalate PCs, nanowire-based PCs with heterostructured cavities with high Q ($\approx 4\times10^5$) have been shown to be highly tunable over a wide range by infiltration. As an example such structures' peak transmission could be continuously linearly tuned from $\approx$ 1604 to $\approx$ 1622 nm by changing the infiltrating medium refractive index from $\approx$ 1.392 to $\approx$ 1.442 [39.119]; note that the tuning range could be extended by careful control over the pillar array aspect ratio.

PC Refractive-Index Tuning by Altering the Concentration of Free Carriers (Using Electric Field or Temperature) in Semiconductor-Based PCs

An elegant way to rapidly tune the PBG of semiconductor-based PCs is to adjust the refractive index by modulating the free-carrier concentration using an ultrafast optical pulse [39.120]. Using this approach, reflectivity of a two-dimensional silicon-based honeycomb PCs with 412 nm air holes (100 μm in length) in a 500 nm periodic array, was studied with a pump-probe approach [39.120]. By varying the delay between the pump the speed of PBG tuning was measured to be about 0.5 ps. The reflectance relaxation (corresponding to the return of the PBG to its original position) occurred on a timescale of 10−100 ns, characteristic of recombination of excess electrons and holes. Although these results are very encouraging, this approach cannot suppress or correct for light-scattering losses caused by structural imperfections, which remain an important consideration for currently fabricated PCs.

Another elegant application of these ideas is in resonant absorption structures exposed to different levels of optical pumping, to change the response. In particular, lateral grating-based PCs have been demonstrated with arrays of nanometer sized silicon p-i-n detectors in which local optical field enhancement is provided by the periodic detectors and the interference within an air/SiO_2/Si cavity structure [39.121, 122].

39.3.2 Tuning PC Response by Altering the Physical Structure of the PC

The second approach that we discuss for tuning the response of a PC is based on changes to the physical structure of the PC. We discuss the following approaches for tuning using this approach: tuning using

1. Temperature
2. An applied magnetic field
3. Strain/deformation
4. Piezoelectric effects, and
5. Using micro-electromechanical systems (MEMS) (i. e., actuation).

Tuning PC Response Using Temperature

An example of this approach is a study of temperature tuning of PCs fabricated from self-assembled

polystyrene beads [39.123]. The PBG in these structures was fine-tuned by annealing samples at temperatures of 20–100 °C, resulting in a continuous blue shift of the stop band wavelength from 576 to 548 nm. New stop bands appeared in the UV transmission spectra when the sample was annealed above about 93 °C – the glass-transition temperature of the polystyrene beads.

Tuning PC Response Using Magnetism

An example of this approach is the use of an external applied magnetic field to adjust the spatial orientation of a PC [39.101]. This can find application in fabricating photonic devices such as tunable mirrors and diffractive display devices. These authors fabricated magnetic PCs by using monodisperse polystyrene beads self-assembled into a ferrofluid consisting of magnetite particles, with particle sizes below 15 nm [39.101]. On evaporation of the solvent, a cubic PC lattice was formed with the nanoparticles precipitating out into the interstices between the spherical polystyrene colloids. These authors then showed how the template could be selectively removed by calcination or wet etching to reveal an inverse opal of magnetite – such structures being proposed as suitable for developing magnetically tunable PCs.

Also, ordered magnetic particle arrays were fabricated through nanosphere deposition [39.124], which can in principle enable the realization of arbitrary material magnetically reconfigurable PCs.

Tuning PC Response Using Strain

The concept in this case is quite straightforward – deforming or straining the PC changes the lattice constant or arrangement of dielectric elements in the PC with concomitant change in the photonic band structure. Polymeric materials would appear to be most suited to this methodology, owing to their ability to sustain considerable strains. However, concerns of reversibility and speed of tuning would clearly need to be addressed. Theoretical predictions for the influence of deformation on such systems include a report on a new class of PC based on self-assembling cholesteric elastomers [39.125]. These elastomers are highly deformable when subjected to external stress. The high sensitivity of the photonic band structure to strain, and the opening of new PBGs have been discussed [39.125]. Charged colloidal crystals were also fixed in a poly(acrylamide) hydrogel matrix to fabricate PCs whose diffraction peaks were tuned by applying mechanical stress [39.126]. The PBG shifted linearly and reversibly over almost the entire visible spectral region (from 460 to 810 nm).

Modeling of the photonic band structure of 2-D silicon-based triangular PCs under mechanical deformation was also reported [39.127]. The structures considered consisted of a silicon matrix with air columns. The authors showed that while a 3% applied shear strain provides only minor modifications to the PBG, uniaxial tension can produce a considerable shift. Other modeling includes a study of how strain can be used to tune the anisotropic optical response of 2-D PCs in the long-wavelength limit [39.128]. These calculations showed that the decrease in dielectric constant per unit strain is larger in the direction of the strain than normal to it. Indeed, the calculated birefringence is larger than that of quartz. They suggest that strain-tuning of this birefringence has attractive applications in polarization-based optical devices.

To appreciate the sensitivity of such structures to mechanical tuning, it is instructive to refer to some recent work on PMMA inverse opal PC structures that were fabricated using silica opal templates [39.129]. Under the application of uniaxial deformation of these PCs, the authors found a blue shift of the stop band in the transmission spectrum – the peak wavelength of the stop band shifted from about 545 nm in the undeformed material to about 470 nm under a stretch ratio of about 1.6.

Another practical approach that has been applied to physically tuning PC structures is that of thermal annealing. One such study showed how the optical properties of colloidal PCs consisting of silica spheres can be tuned through thermal treatment [39.130]. This was attributed to both structural and physiochemical modification of the material on annealing. A shift in the minimum transmission from about 1000 nm (unannealed) to about 850 nm (for annealing at about 1000 °C) was demonstrated, or about a maximum shift in Bragg wavelength of $\approx 11\%$.

A quite novel application of strain tuning was recently reported [39.131]. The authors studied 2-D PCs consisting of arrays of coupled optical microcavities fabricated from vertical-cavity surface-emitting laser structures. The influence of strain, as manifested by shifts in the positions of neighboring rows of microcavities with respect to each other, corresponded to alternating square or quasihexagonal shear-strain patterns. For strains below a critical threshold value, the lasing photon mode-locked to the corresponding mode in the unstrained PC. At the critical strain, switching occurred between the square and hexagonal lattice modes.

Finally, there has been a proposal for using strain in a PC to tune the splitting of a degenerate photon state within the PBG, suitable for implementing tunable PC circuits [39.132]. The principle applied is analogous to the static Jahn–Teller effect in solids. These authors showed that this effect is tunable by using the symmetry and magnitude of the lattice distortion. Using this

effect, the design of an optical valve that controls the resonant coupling of photon modes at the corner of a T-junction waveguide structure has been discussed.

Tuning PC Response Using Piezoelectric Effects

In this section we discuss using piezoelectric effects to physically change PC structures and hence tune them. A proposal was made for using the piezoelectric effect to distort the original symmetry of a two-dimensional PC from a regular hexagonal lattice to a quasihexagonal lattice under an applied electric field [39.133]. The original bands decomposed into several strained bands, dependent on the magnitude and direction of the applied field. In the proposed structures, the application of $\approx 3\%$ shear strain is shown to be suitable for shifting 73% of the original PBG, which they refer to as the tunable-band-gap regime. An advantage of such an approach is that such structures are suggested to be capable of operation at speeds approaching MHz. Another report [39.134] discusses the design and implementation of a tunable silicon-based PBG microcavity in an optical waveguide, where tuning is accomplished using the piezoelectric effect to strain the PC; this was carried out using integrated piezoelectric microactuators. These authors report on a 1.54 nm shift in the cavity resonance at 1.56 μm for an applied strain of 0.04%.

There have also been reports of coupling piezoelectric-based actuators to PCs. One such report [39.135] discusses a poly(2-methoxyethyl acrylate)-based PC composite directly coupled to a piezoelectric actuator to study static and dynamic stop-band tuning characteristics; the stop band of this device could be tuned through a 172 nm tuning range, and could be modulated at up to 200 Hz.

Tuning PC Response Using Micro-Electromechanical Systems (MEMS) Actuation

There have been a number of interesting developments in this field including PC-based air-bridge devices consisting of suspended 1-D PC mirrors separated by a Fabry–Pérot cavity (gap) [39.136]. When such structures are mechanically perturbed, there can be a substantial shift in the PBG due to strain. The authors discuss how a suite of spectrally tunable devices can be envisioned based on such structures – these include modulators, optical filters, optical switches, wavelength division multiplexing (WDM), optical logical circuits, variable attenuators, power splitters and isolators. Indeed, the generalization of these concepts beyond 1-D was discussed in a recent patent [39.137] that covers tunable PC structures. This report [39.137], as well as others [39.137], consider the extension of these ideas to form families of micromachined devices. These authors [39.138] modeled and implemented a set of micromachined vertical resonator structures for 1.55 μm filters consisting of two PC (distributed Bragg reflector, DBR) mirrors separated by either an air gap or semiconductor heterostructure. Electromechanical tuning was used to adjust the separation between the mirrors and hence fine-tune the transmission spectrum. The mirror structures were implemented using strong-index-contrast InP/air DBRs giving an index contrast of 2.17, and weak-index-contrast (0.5) silicon nitride/silicon dioxide DBRs. In the former case, a tuning range of over 8% of the absolute wavelength was achieved; varying the intermembrane voltage up to 5 V gave a tuning range of over 110 nm. Similar planar structures are discussed in other papers [39.139] where the mirrors are formed from two slabs of PC separated by an adjustable air gap [39.137]. These structures have been shown to be able to perform as either flat-top reflection or all-pass transmission filters, by varying the distance between the slabs, for normally incident light. Unlike all previously reported all-pass reflection filters, based on Gires–Tournois interferometers using multiple dielectric stacks, their structure generates an all-pass transmission spectrum, significantly simplifying signal extraction and optical alignment – also the spectral response is polarization-independent owing to the 90° rotational symmetry of their structure. Some interesting recent work includes the coupling of a single photon source (based on an InAs quantum dot) embedded in a PC, coupled to a microfibre [39.140]. By adjusting the coupling strength of microfibre to microcavity mechanically, the spectral response was shown to be tunable over a 1.5 nm bandwidth. Such work has applicability for on-demand single-photon sources, of great importance for quantum information processing.

39.4 Summary and Conclusions

This chapter discussed the fabrication and properties of PCs, with emphasis on developing tunable structures. New fabrication methodologies and tuning schemes offer a glimpse of the far-reaching prospects for developing photonic devices which can, in a discretionary fashion, control the propagation of modes of light in an analogous fashion to the way in which nanostructures have been harnessed to control electron-based phenom-

ena. Analogous to the evolution of electronic systems, one can anticipate a path toward development of compact active integrated photonic systems, as envisioned based on the technology outlined in this review.

References

39.1 T.F. Krauss, R.M. de la Rue: Photonic crystals in the optical regime – Past, present, and future, Prog. Quantum Electron. **23**, 51 (1999)

39.2 V. Berger: From photonic band gaps to refractive index engineering, Opt. Mater. **11**, 131 (1999)

39.3 D.J. Norris, Y.A. Vlasov: Chemical approaches to three-dimensional semiconductor photonic crystals, Adv. Mater. **13**(6), 371 (2001)

39.4 V. Mizeikis, S. Juodkazis, A. Marcinkevicius, S. Matsuo, H. Misawa: Tailoring and characterization of photonic crystals, J. Photochem. Photobiol. C **2**, 35 (2001)

39.5 Photonics Nanostruct. **1**(1), 1 (2003): whole edition is dedicated to fundamentals and applications of photonic crystals

39.6 S.G. Johnson, J.D. Joannopoulos: *Photonic Crystals: Road from Theory to Practice* (Kluwer, Boston 2002)

39.7 J.D. Joannopoulos, R.D. Meade, J.N. Winn: *PCs: Moulding the Flow of Light* (Princeton Univ. Press, Princeton 1995)

39.8 S. John: Strong localization in certain disordered dielectric super-lattices, Phys. Rev. Lett. **58**(23), 2486 (1987)

39.9 E. Yablonovitch: Inhibited spontaneous emission in solid state physics and electronics, Phys. Rev. Lett. **58**(20), 2059 (1987)

39.10 H. Kosaka, T. Kawashima, A. Tomita, M. Notomi, T. Tamamura, T. Sato, S. Kwakami: Superprism phenomena in photonic crystals, Phys. Rev. B **58**(16), R10096 (1998)

39.11 V. Berger: Photonic crystals and photonic structures, Cur. Opin. Solid State Mater. Sci. **4**, 209 (1999)

39.12 K.M. Ho, C.T. Chan, C.M. Soukoulis: Existence of a photonic gap in periodic dielectric structures, Phys. Rev. Lett. **65**(25), 3152 (1990)

39.13 S. Satpathy, Z. Zhang, M.R. Salehpour: Theory of photon bands in three-dimensional periodic dielectric structures, Phys. Rev. Lett. **64**(11), 1239 (1990)

39.14 E. Yablonovitch: Photonic band-gap structures, J. Opt. Soc. Am. B **10**(2), 283 (1993)

39.15 Y. Xia: Photonic Crystals, Adv. Mater. **13**(6), 369 (2001)

39.16 C. Anderson, K. Giapis: Larger two-dimensional photonic band gaps, Phys. Rev. Lett. **77**, 2949 (1996)

39.17 R.D. Meade, A.M. Rappe, K.D. Brommer, J.D. Joannopoulos: Nature of the photonic band gap: Some insights from a field analysis, J. Opt. Soc. Am. B **10**(2), 328 (1993)

39.18 R.D. Meade, A.M. Rappe, K.D. Brommer, J.D. Joannopoulos, O.L. Alerhand: Accurate theoretical analysis of photonic band-gap materials, Phys. Rev. B **48**(11), 8434 (1993)

39.19 T. Xu, S. Yang, S.V. Nair, H.E. Ruda: Confined modes in finite-size photonic crystals, Phys. Rev. B **72**(11), 045126 (2005)

39.20 T. Xu, S. Yang, S.V. Nair, H.E. Ruda: Nanowire-array-based photonic crystal cavity by finite-difference time-domain calculations, Phys. Rev. B **75**(12), 125104 (2007)

39.21 T. Xu, M.S. Wheeler, H.E. Ruda, M. Mojahedi, S. Aitchison: The influence of material absorption on the quality factor of photonic crystal cavities, Opt. Express **17**(10), 8343 (2009)

39.22 T. Xu, M.S. Wheeler, S.V. Nair, H.E. Ruda, M. Mojahedi, S. Aitchison: Highly confined mode above the light line in a two-dimensional photonic crystal slab, Appl. Phys. Lett. **93**(24), 241105 (2008)

39.23 C. Lopez: Materials aspects of PCs, Adv. Mater. **15**(20), 1679 (2003)

39.24 C. Weisbuch, C.H. Benisty, S. Olivier, M. Rattier, C.J.M. Smith, T.F. Krauss: Advances in photonic crystals, Phys. Status Solidi **221**(93), 93 (2000)

39.25 C. Weisbuch, H. Benisty, M. Rattier, C.J.M. Smith, T.F. Krauss: Advances in 2D semiconductor PCs, Synth. Mater. **116**, 449 (2001)

39.26 S.L. Swartz: Topics in electronic ceramics, IEEE Trans. Electr. Insul. **25**(5), 935 (1990)

39.27 F. Watt: Focused high energy proton beam micromachining: A perspective view, Nucl. Instrum. Methods Phys. Res. **158**, 165 (1999)

39.28 D.K. Ferry, R.O. Grondin: *Physics of Submicron Devices* (Plenum, New York 1991)

39.29 J. Gierak, D. Mailly, G. Faini, J.L. Pelouard, P. Denk, F. Pardo, J.Y. Marzin, A. Septier, G. Schmmid, J. Ferre, R. Hydman, C. Chappert, J. Flicstein, B. Gayral, J.M. Gerard: Nano-fabrication with focused ion beams, Microelectron. Eng. **57/58**, 865 (2001)

39.30 K. Gamo: Nanofabrication by FIB, Microelectron. Eng. **32**, 159 (1996)

39.31 J. Melngailis, A.A. Mondelli, I.L. Berry III, R. Mohondro: A review of ion projection lithography, J. Vac. Sci. Technol. B **16**(3), 927 (1998)

39.32 P. Peercy: The drive to miniaturization, Nature **406**, 1023 (2000)

39.33 T. Ito, S. Okazaki: Pushing the limits of lithography, Nature **406**, 1027 (2000)

39.34 N. Peng, C. Jeynes, R.P. Webb, I.R. Chakarov, M.G. Blamire: Monte Carlo simulations of masked ion beam irradiation damage profiles in $YBa_2Cu_3O_{7-\delta}$ thin films, Nucl. Instrum. Methods Phys. Res. B **178**, 242 (2001)

39.35 J.L. Bartelt: Masked ion beam lithography: An emerging technology, Solid State Technol. **29**(5), 215 (1986)

39.36 D.P. Stumbo, J.C. Wolfe: Contrast of ion beam proximity printing with non-ideal masks, J. Vac. Sci. Technol. B **12**(6), 3539 (1994)

39.37 T. Devolder, C. Chappert, Y. Chen, E. Cambril, H. Launois, H. Bernas, J. Ferre, J.P. Jamet: Patterning of planar magnetic nanostructures by ion irradiation, J. Vac. Sci. Technol. B **17**(6), 3177 (1999)

39.38 P. Ruchhoeft, J.C. Wolfe, R. Bass: Ion beam aperture-array lithography, J. Vac. Sci. Technol. B **19**(6), 2529 (2001)

39.39 Y. Hsieh, Y. Hwang, J. Fu, Y. Tsou, Y. Peng, L. Chen: Dislocation multiplication inside contact holes, Microelectron. Reliab. **39**, 15 (1999)

39.40 A.N. Broers, A.C.F. Hoole, J.M. Ryan: Electron beam lithography – Resolution limits, Microelectron. Eng. **32**, 131 (1996)

39.41 J. Gierak, A. Septier, C. Vieu: Design and realization of a very high-resolution FIB nanofabrication instrument, Nucl. Instrum. Methods Phys. Res. A **427**, 91 (1999)

39.42 C. Lehrer, L. Frey, S. Petersen, H. Ryssel: Limitations of focused ion beam nano-machining, J. Vac. Sci. Technol. B **19**(6), 2533 (2001)

39.43 H. Masuda, K. Fukuda: Ordered metal nanohole arrays made by a two-step replication of honeycomb structures of anodic alumina, Science **268**, 1466 (1995)

39.44 R. Tonucci, B. Justus, A. Campillo, C. Ford: Nanochannel array glass, Science **258**, 783 (1992)

39.45 V. Lehmann, H. Foll: Formation mechanism and properties of electrochemically etched trenches in n-type silicon, J. Electrochem. Soc. **137**(2), 653 (1990)

39.46 A. Birner, R.B. Wehrspohn, U.M. Gosele, K. Busch: Silicon-based photonic crystals, Adv. Mater. **13**(6), 377 (2001)

39.47 J. Martin: Nanomaterials: A membrane-based synthetic approach, Science **266**, 1961 (1994)

39.48 J.I. Martin, J. Nogues, K. Liu, J.L. Vicent, I.K. Schuller: Ordered magnetic nanostructures: Fabrication and properties, J. Magn. Mater. **256**(1-3), 449 (2003)

39.49 H. Masuda, M. Ohya, H. Asoh, M. Nakao, M. Nohtomi, T. Tamamura: Photonic crystals using anodic porous alumina, Jpn. J. Appl. Phys. Pt. 2 **38**(12A), L1403 (1999)

39.50 R. Wehrspohn, J. Schilling: Electrochemically prepared pore arrays for photonic-crystal applications, MRS Bulletin **26**(8), 623 (2001)

39.51 A.P. Li, F. Muller, A.B.K. Nielsch, U. Gosele: Hexagonal pore arrays with a 50-420 nm interpore distance formed by self-organization in anodic alumina, J. Appl. Phys. **84**(11), 6023 (1998)

39.52 H. Masuda, H. Asoh, M. Watanabe, K. Nishio, M. Nakao, T. Tamamura: Square and triangular nanohole array architectures in anodic alumina, Adv. Mater. **13**, 189 (2001)

39.53 M. Nakao, S. Oku, T. Tamamura, K. Yasui, H. Masuda: GaAs and InP nanohole arrays fabricated by reactive beam etching using highly ordered alumina membranes, Jpn. J. Appl. Phys. Pt. 1 **38**(2B), 1052 (1999)

39.54 J. Liang, H. Chik, A. Yin, J. Xu: Two-dimensional lateral superlattices of nanostructures: Nonlithographic formation by anodic membrane template, J. Appl. Phys. **91**(4), 2544 (2002)

39.55 N. Matsuura, T.W. Simpson, C.P. McNorgan, I.V. Mitchell, X. Mei, P. Morales, H.E. Ruda: Nanometer-scale pattern transfer using ion implantation. In: *Three-Dimensional Nanoengineered Assemblies*, MRS Proceedings, Vol. 739, ed. by T.M. Orlando, L. Merhari, D.P. Taylor, K. Ikuta (Cambridge Univ. Press, Boston 2002)

39.56 N. Matsuura, T.W. Simpson, I.V. Mitchell, X. Mei, P. Morales, H.E. Ruda: Ultra-high density, nonlithographic, sub-100 nm pattern transfer by ion implantation an selective chemical etching, Appl. Phys. Lett. **81**(25), 4826 (2002)

39.57 E. Rimini: *Ion Implantation: Basics to Device Fabrication* (Springer, New York 1995)

39.58 G. Hobler: Monte Carlo simulation of two-dimensional implanted dopant distributions at mask edges, Nucl. Instrum. Methods Phys. Res. B **96**, 155 (1995)

39.59 M.M. Faye, C. Vieu, G.B. Assayag, P. Salles, A. Claverie: Lateral damage extension during masked ion implantation into GaAs, J. Appl. Phys. **80**(8), 4303 (1996)

39.60 P. Schmuki, L. Erickson: Direct micro-patterning of Si and GaAs using electrochemical development of focused ion beam implants, Appl. Phys. Lett. **73**, 2600 (1998)

39.61 K. Wang, A. Chelnokov, S. Rowson, P. Garouche, J.-M. Lourtioz: *Three-dimensional Yablonovite-like photonic crystals by focused ion beam etching of macroporous silicon*, *Mater. Res. Soc. Symp. Proc.* **637**, *E1.4.1* 2001)

39.62 S. Lin, J. Fleming, D. Hetherington, B. Smith, R. Biswas, K. Ho, M. Sigalas, W. Zubrzycki, S. Kurtz, J. Bur: A three-dimensional photonic crystals operating at infrared wavelengths, Nature **394**, 251 (1998)

39.63 S. Noda, K. Tomoda, N. Yamamoto, A. Chutinan: Full three-dimensional photonic bandgap crystals at near-infrared wavelengths, Science **289**, 604 (2000)

39.64 E. Yablonovitch, T. Gmitter, K. Leung: Photonic band structure: The face-centered-cubic case employing non-spherical atoms, Phys. Rev. Lett. **67**, 2295 (1991)

39.65 E. Ozbay, A. Abeyta, G. Tuttle, M. Tringides, R. Biswas, C. Chan, C. Soukoulis, K. Ho: Measurement of a three-dimensional photonic band gap in a crystal structure made of dielectric rods, Phys. Rev. B **50**, 1945 (1994)

39.66 K.M. Ho, C.T. Chan, C.M. Soukoulis, R. Biswas, M. Sigalas: Photonic band gaps in three dimensions: New layer-by-layer periodic structures, Solid State Commun. **89**(5), 413 (1994)

39.67 H.S. Sözüer, J.P. Dowling: Photonic band calculations for woodpile structures, J. Mod. Opt. **41**(2), 231 (1994)

39.68 Y. Xia, B. Gates, Z.-Y. Li: Self-assembly approaches to three-dimensional photonic crystals, Adv. Mater. **13**(6), 409 (2001)

39.69 A. Moroz: Three-dimensional complete photonic-band-gap structures in the visible range, Phys. Rev. Lett. **83**(25), 5274 (1999)

39.70 S. Kawakami: Fabrication of sub-micrometre 3D periodic structures composed of Si/SiO_2, Electron. Lett. **33**(4), 1260 (1997)

39.71 C.T. Chan, S. Datta, K.M. Ho, C.M. Soukoulis: A7 structure: A family of photonic crystals, Phys. Rev. B **50**(3), 1988 (1994)

39.72 G. Feiertag, W. Ehrfeld, H. Freimuth, H. Kolle, H. Lehr, M. Schmidt, M.M. Sigalas, C.M. Soukoulis, G. Kiriakidis, T. Pedersen, J. Kuhl, W. Koenig: Fabrication of photonic crystals by deep x-ray lithography, Appl. Phys. Lett. **71**(11), 1441 (1997)

39.73 C. Cheng, A. Scherer: Fabrication of photonic band-gap crystals, J. Vac. Sci. Technol. B **13**(6), 2153 (1995)

39.74 A.A. Zakhidov, R.H. Baughman, Z. Iqbal, C. Cui, I. Khayrullin, S.O. Dantas, J. Marti, V.G. Ralchenko: Carbon structures with three-dimensional periodicity at optical wavelengths, Science **282**, 897 (1998)

39.75 A. Blanco, E. Chomski, S. Grabtchak, M. Ibisate, S. John, S.W. Leonard, C. Lopez, F. Meseguer, H. Miguez, J.P. Mondia, G.A. Ozin, O. Toader, H.M. van Driel: Large-scale synthesis of a silicon photonic crystals with a complete three-dimensional bandgap near 1.5 micrometres, Nature **405**, 437 (2000)

39.76 J.E.G.J. Wijnhoven, W.L. Vos: Preparation of photonic crystals made of air spheres in titania, Science **281**, 802 (1998)

39.77 Y. Xia, B. Gates, Y. Yin, Y. Lu: Mono-dispersed colloidal spheres: Old materials with new applications, Adv. Mater. **12**(10), 693 (2000)

39.78 A. Stein: Sphere templating methods for periodic porous solids, Micropor. Mesopor. Mater. **44/45**, 227 (2001)

39.79 J. Martorell, N.M. Lawandy: Observation of inhibited spontaneous emission in a periodic dielectric structure, Phys. Rev. Lett. **65**(15), 1877 (1990)

39.80 I.I. Tarhan, G.H. Watson: Photonic band structure of FCC colloidal crystals, Phys. Rev. Lett. **76**(2–8), 315 (1996)

39.81 Y.A. Vlasov, M. Deutsch, D.J. Norris: Single-domain spectroscopy of self-assembled photonic crystals, Appl. Phys. Lett. **76**(12), 1627 (2000)

39.82 D.C. Reynolds, F. Lopez-Tejeira, D. Cassagne, F. Garcia-Vidal, C. Jouanin, J. Sanchez-Dehesa: Spectral properties of opal-based photonic crystals having a SiO_2 matrix, Phys. Rev. B **60**(16), 11422 (1999)

39.83 V.N. Bogomolov, S.V. Gaponenko, I.N. Germanenko, A.M. Kapitonov, E.P. Petrov, N.V. Gaponenko, A.V. Prokofiev, A.N. Ponyavina, N.I. Silvanovich, S.M. Samoilovich: Photonic band gap phenomenon and optical properties of artificial opals, Phys. Rev. E **55**(6), 7619 (1997)

39.84 S.G. Romanov, A.V. Fokin, R.M. De La Rue: Stopband structure in complementary three-dimensional opal-based photonic crystals, J. Phys. Condens. Matter **11**, 3593 (1999)

39.85 R. Biswas, M.M. Sigalas, G. Subramania, K.M. Ho: Photonic band gaps in colloidal systems, Phys. Rev. B **57**(9), 3701 (1998)

39.86 A. Richel, N.P. Johnson, D.W. McComb: Observation of Bragg reflection in photonic crystals synthesized from air spheres in a titania matrix, Appl. Phys. Lett. **76**(14), 1816 (2000)

39.87 B.T. Holland, C.F. Blanford, A. Stein: Synthesis of macroporous minerals with highly ordered three-dimensional arrays of spheroidal voids, Science **281**, 538 (1998)

39.88 M.S. Thijssen, R. Sprik, J.E.G.J. Wijnhoven, M. Megens, T. Narayanan, A. Lagendijk, W.L. Vos: Inhibited light propagation and broadband reflection in photonic air-sphere crystals, Phys. Rev. Lett. **83**(14), 2730 (1999)

39.89 F. Meseguer, A. Blanco, H. Miguez, F. Garcia-Santamaria, M. Ibisate, C. Lopez: Synthesis of inverse opals, Colloids Surf. A **202**, 281 (2002)

39.90 O.D. Velev, E.W. Kaler: Research news: Structured porous materials via colloidal crystal templating: From inorganic oxides to metals, Adv. Mater. **12**(7), 531 (2000)

39.91 O.D. Velev, A.M. Lenhoff: Colloidal crystals as templates for porous materials, Curr. Opin. Colloid Interface. Sci. **5**, 56 (2000)

39.92 F. Blanford, H. Yan, R.C. Schroden, M. Al-Daous, A. Stein: Gems of chemistry and physics: Macroporous metal oxides with 3D order, Adv. Mater. **13**(6), 401 (2001)

39.93 A.M. Kapitonov, N.V. Gaponenko, V.N. Bogomolov, A.V. Prokofiev, S.M. Samoilovich, S.V. Gaponenko: Photonic stop band in a three-dimensional SiO_2/TiO_2 lattice, Phys. Status Solidi (a) **165**(1), 119 (1998)

39.94 N. Matsuura, S. Yang, P. Sun, H.E. Ruda: Development of highly-ordered, ferroelectric inverse opal films using sol-gel infiltration, Appl. Phys. A **81**, 379 (2005)

39.95 S.M. Yang, H. Miguez, G.A. Ozin: Opal circuits of light – Planarized micro photonic crystals chips, Adv. Funct. Mater. **12**(6/7), 425431 (2002)

39.96 A. Polman, P. Wiltzius: Materials science aspects of PCs, MRS Bulletin **26**(8), 608 (2001)

39.97 V.L. Colvin: From opals to optics: Colloidal photonic crystals, MRS Bulletin **26**(8), 637 (2001)

39.98 S.H. Park, D. Qin, Y. Xia: Crystallization of mesoscale particles over large areas, Adv. Mater. **10**(3), 1028 (1998)

39.99 P. Jiang, J. Bertone, K. Hwang, V. Colvin: Single-crystal colloidal multi-layers of controlled thickness, Chem. Mater **11**, 2132 (1999)

39.100 Y.A. Vlasov, X.-Z. Bo, J.C. Sturm, D.J. Norris: On-chip natural assembly of silicon photonic bandgap crystals, Nature **414**, 289 (2001)

39.101 B. Gates, Y. Xia: Photonic crystals that can be addressed with an external magnetic field, Adv. Mater. **13**(21), 1605 (2001)

39.102 I. Soten, H. Miguez, S.M. Yang, S. Petrov, N. Coombs, N. Tetreault, N. Matsuura, H.E. Ruda, G.A. Ozin: Barium titanate inverted opals – Synthesis, characterization, and optical properties, Adv. Funct. Mater. **12**(1), 71 (2002)

39.103 M.C. Wanke, O. Lehmann, K. Muller, Q. Wen, M. Stuke: Laser rapid prototyping of photonic band-gap microstructures, Science **275**, 1284 (1997)

39.104 B.H. Cumpston, S.P. Ananthavel, S. Barlow, D.L. Dyer, J.E. Ehrlich, L.L. Erskine, A.A. Heikal, S.M. Kuebler, I.-Y.S. Lee, D. McCord-Maughon, J. Qin, H. Rockel, M. Rumi, X.-L. Wu, S.R. Marder, J.W. Perry: Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication, Nature **398**, 51 (1999)

39.105 H.-B. Sun, S. Matsuo, H. Misawa: Three-dimensional photonic crystals structures achieved with two-photon-absorption photo-polymerization of resin, Appl. Phys. Lett. **74**(6), 786 (1999)

39.106 M. Campbell, D. Sharp, M. Harrison, R. Denning, A. Turberfield: Fabrication of photonic crystals for the visible spectrum by holographic lithography, Nature **404**, 53 (2000)

39.107 J. Martorell, R. Vilaseca, R. Corbalan: Second harmonic generation in a photonic crystal, Appl. Phys. Lett. **70**(6), 702 (1997)

39.108 H. Inouye, Y. Kanemitsu: Direct observation of non-linear effects in a one dimensional photonic crystal, Appl. Phys. Lett. **82**(8), 1155 (2003)

39.109 Z.-Z. Gu, T. Iyoda, A. Fujishima, O. Sato: Photo reversible regulation of optical stop bands, Adv. Mater. **13**(7), 1295 (2001)

39.110 X. Hu, Q. Zhang, Y. Liu, B. Cheng, D. Zhang: Ultrafast three-dimensional tunable photonic crystal, Appl. Phys. Lett. **83**(13), 2518 (2003)

39.111 T.L. Spencer, R. Cisek, V. Barzda, U. Philipose, H.E. Ruda, A. Shik: Orientation dependent nonlinear optical effects in ZnSe nanowires, Appl. Phys. Lett. **94**(23), 233119 (2009)

39.112 R. Cisek, V. Barzda, H.E. Ruda, A. Shik: Nonlinear optical properties of semiconductor nanowires, IEEE J. Sel. Top. Quantum Electron. **17**(6), 915–921 (2011)

39.113 B. Li, J. Zou, X.J. Wang, X.H. Liu, J. Zi: Ferroelectric inverse opals with electrically tunable photonic band gap, Appl. Phys. Lett. **83**(23), 4704 (2003)

39.114 T.B. Xu, Z.Y. Cheng, Q.M. Zhang, R.H. Baughman, C. Cui, A.A. Zakhidov, J. Su: Fabrication and characterization of three dimensional periodic ferroelectric polymer-silica opal composites and inverse opals, J. Appl. Phys. **88**(1), 405 (2000)

39.115 L. Nucara, F. Greco, V. Mattoli: Electrically responsive photonic crystals: A review, J. Mater. Chem. C **3**(33), 8449 (2015)

39.116 S. John, K. Busch: Photonic bandgap formation and tunability in certain self-organizing systems, J. Lightwave Technol. **17**(11), 1931 (1999)

39.117 J.D. Joannopoulos: The almost-magical world of photonic crystals, Braz. J. Phys. **26**(1), 53 (1996)

39.118 X. Yoshino, Y. Kawagishi, M. Ozaki, A. Kose: Mechanical tuning of the optical properties of plastic opal as a photonic crystals, Jpn. J. Appl. Phys. Pt. 2 **38**(7A), L786 (1999)

39.119 T. Xu, N. Zhu, M.Y.C. Xu, L. Wosinski, J.S. Aitchison, H.E. Ruda: Pillar array microsensor, Opt. Express **18**(6), 5420 (2010)

39.120 S.W. Leonard, H.M. van Driel, J. Schilling, R.B. Wehrspohn: Ultrafast band edge tuning of a two dimensional silicon photonic crystal via free carrier injection, Phys. Rev. B **66**, 161102 (2002)

39.121 Y. Sun, E. Rusli, M. Yu, J. Salfi, C. Souza, H.E. Ruda, N. Singh, F.K. Lin, P. Lo, D.L. Kwong: Electrical characteristics and photocurrent spectral response of Si nanowires p-i-n junctions, Opt. Express **19**(6), 5464 (2011)

39.122 J. Salfi, C. Stewart, S.V. Nair, C.Y. Chen, S. Yongshun, E. Rusli, F.K. Lin, M. Yu, N. Singh, C.F. de Sousa, H.E. Ruda: Antenna-enhanced and polarisation sensitive photoresponse in arrays of silicon p-i-n nanowires, New J. Phys. **15**, 093029 (2013)

39.123 B. Gates, S.H. Park, Y. Xia: Tuning the photonic bandgap properties of crystalline arrays of polystyrene beads by annealing at elevated temperatures, Adv. Mater. **12**(9), 653 (2000)

39.124 Y. Song, W. Yin, Y.H. Wang, J.P. Zhang, Y. Wang, R. Wang, J. Han, W. Wang, S.V. Nair, H.E. Ruda: Magneto-plasmons in periodic nanoporous structures, Sci. Rep. **4**, 4991 (2014)

39.125 P. A. Bermel, M. Warner: Photonic bandgap structure of highly deformable self-assembling systems, Phys. Rev. E **65**(1), 010702 (2001)

39.126 Y. Iwayama, J. Yamanaka, Y. Takiguchi, M. Takasaka, K. Ito, T. Shinohara, T. Sawada, M. Yonese: Optically tunable gelled photonic crystal covering almost the entire visible light wavelength region, Langmuir **19**(4), 977 (2003)

39.127 S. Jun, Y.-S. Cho: Deformation-induced bandgap tuning of 2D silicon-based photonic crystals, Opt. Express **11**(21), 2769 (2003)

39.128 C.-S. Kee, K. Kim, H. Lim: Tuning of anisotropic optical properties of 2D dielectric photonic crystals, Physica B **338**, 153 (2003)

39.129 K. Sumioka, H. Kayashima, T. Tsutsui: Tuning the optical properties of inverse opal photonic crystals by deformation, Adv. Mater. **14**(18), 1284 (2002)

39.130 H. Miguez, F. Meseguer, C. Lopez, A. Blanco, J.S. Moya, J. Requena, A. Mifsud, V. Fornes: Control of the photonic crystal properties of fcc-packed sub-micrometer SiO_2 spheres by sintering, Adv. Mater. **10**(6), 480 (1998)

39.131 H. Pier, E. Kapon, M. Moser: Strain effects and phase transitions in photonic crystal resonator crystals, Nature **407**, 880 (2000)

39.132 N. Malkova, V. Gopalan: Strain tunable optical valves at T-junction waveguides in photonic crystals, Phys. Rev. B **68**, 245115 (2003)

39.133 S. Kim, V. Gopalan: Strain tunable photonic band gap crystals, Appl. Phys. Lett. **78**(20), 3015 (2001)

39.134 C.W. Wong, P.T. Rakich, S.G. Johnson, M. Qi, H.I. Smith, E.P. Ippen, L.C. Kimmerling, Y. Jeon, G. Barbastathis, S.-G. Kim: Strain-tunable silicon photonic bandgap microcavities in optical waveguides, Appl. Phys. Lett. **84**(8), 1242 (2004)

39.135 S.H. Foulger, P. Jiang, A. Lattam, D.W. Smith, J. Ballato, D.E. Dausch, S. Grego, B.R. Stoner: Photonic crystal composites with reversible high-frequency stop band shifts, Adv. Mater. **15**(9), 685 (2003)
39.136 S. Rajic, J.L. Corbeil, P.G. Datskos: Feasibility of tunable MEMS photonic crystal devices, Ultramicroscopy **97**, 473 (2003)
39.137 N. Matsuura, H. E. Ruda, B. G. Yacobi: Configurable photonic device, US Patent Ser., Vol. 6961501 B2 (2005)
39.138 H. Hiller, J. Daleiden, C. Prott, F. Römer, S. Irmer, V. Rangelov, A. Tarraf, S. Schüler, M. Strassner: Potential for a micromachined actuation of ultra-wide continuously tunable optoelectronic devices, Appl. Phys. B **75**, 3 (2002)
39.139 W. Suh, S. Fan: Mechanically switchable photonic crystal filter with either all-pass transmission or flat-top reflection characteristics, Opt. Lett. **28**(19), 1763 (2003)
39.140 C.M. Lee, H.J. Lim, C. Schneider, S. Maier, S. Hofling, M. Kamp, Y.H. Lee: Efficient single photon source based on micro-fibre coupled tunable microcavity, Sci. Rep. **5**, 14309 (2015)

40. Quantum Wells, Superlattices, and Band-Gap Engineering

Mark Fox, Radu Ispasoiu

This chapter reviews the principles of bandgap engineering and quantum confinement in semiconductors, with a particular emphasis on the optoelectronic properties of quantum wells. The chapter begins with a review of the fundamental principles of bandgap engineering and quantum confinement. It then describes the optical and electronic properties of semiconductor quantum wells and superlattices at a tutorial level, before describing the principal optoelectronic devices. The topics covered include edge-emitting lasers and light-emitting diodes (LEDs), resonant cavity LEDs and vertical-cavity surface-emitting lasers (VCSELs), quantum cascade lasers, quantum well solar cells, superlattice avalanche photodiodes, infrared detectors, and quantum well light modulators. The chapter concludes with a brief discussion of new research directions on quantum dot and nanowire structures.

40.1 **Principles of Bandgap Engineering and Quantum Confinement** 1038
40.1.1 Lattice Matching 1038
40.1.2 Quantum-Confined Structures 1039
40.2 **Optoelectronic Properties of Quantum-Confined Structures** 1040
40.2.1 Electronic States in Quantum Wells and Superlattices 1040
40.2.2 Interband Optical Transitions 1042
40.2.3 The Quantum-Confined Stark Effect 1044
40.2.4 Intersubband Transitions 1045
40.2.5 Vertical Transport 1046
40.2.6 Carrier Capture and Relaxation 1046
40.3 **Emitters** 1047
40.3.1 Interband Light-Emitting Diodes and Lasers 1047
40.3.2 Quantum Cascade Lasers 1049
40.4 **Detectors** 1050
40.4.1 Solar Cells 1050
40.4.2 Avalanche Photodiodes 1050
40.4.3 Infrared Detectors 1051
40.5 **Modulators** 1052
40.6 **Quantum Dots and Nanowires** 1053
40.7 **Conclusions** 1055
References 1055

The need for efficient light-emitting diodes (LEDs) and lasers operating over the whole of the visible spectrum and also the fiber optic windows at 1.3 and 1.55 μm drives research into new direct-gap semiconductors to act as the active materials. Since the emission wavelength of a semiconductor corresponds to its bandgap energy, the research focuses on engineering new materials that have their bandgaps at custom-designed energies. This science is called *bandgap engineering*.

In the early years of semiconductor optoelectronics, the bandgaps that could be achieved were largely determined by the physical properties of key III–V materials like GaAs and its alloys such as AlGaAs and InGaAs. Then in 1970 a major breakthrough occurred when *Esaki* and *Tsu* invented the semiconductor quantum well (QW) and superlattice (SL) [40.1]. They realized that developments in epitaxial crystal growth could open the door to new structures that exploit the principles of quantum confinement to engineer electronic states with custom-designed properties. They foresaw that these quantum-confined structures would be of interest both to research scientists who would be able to explore uncharted areas of fundamental physics and to engineers who would learn to use their unique properties for device applications. Their insight paved the way for a whole new breed of devices that are now routinely found in a host of everyday applications ranging from blu-ray players to traffic lights.

The emphasis of the chapter is on the optoelectronic properties of QW and SL structures. We begin by out-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_40

lining the basic principles of bandgap engineering and quantum confinement. We then discuss the electronic states in quantum-confined structures and the optical properties that follow from them. In Sects. 40.3–40.5, we explain the principles of the main optoelectronic devices that employ QWs and SLs, namely emitters, detectors, and modulators. Finally, we indicate a few interesting developments that involve semiconductor nanostructures with higher degrees of electronic confinement than QWs, namely nanowires and quantum dots (QDs). A number of texts cover these topics in more detail [40.2–5], and the interested reader is referred to these sources for a more comprehensive treatment. A description of the purely electronic properties of low-dimensional structures may be found in [40.6].

40.1 Principles of Bandgap Engineering and Quantum Confinement

40.1.1 Lattice Matching

The art of bandgap engineering relies heavily on developments in the science of crystal growth. The bulk crystals grown from the melt usually contain a large number of impurities and defects, and optoelectronic devices are therefore grown by epitaxial methods such as liquid phase epitaxy (LPE), molecular beam epitaxy (MBE), and metal-organic vapor phase epitaxy (MOVPE), which is also called metal-organic chemical vapor deposition (MOCVD) (Chap. 14). The basic principle of epitaxy is to grow thin layers of very high purity on top of a bulk crystal called the substrate. The system is said to be *lattice matched* when the lattice constants of the epitaxial layer and the substrate are identical. Lattice matching reduces the number of dislocations in the epitaxial layer, but it also imposes tight restrictions on the band gaps that can be engineered easily, because only a relatively small number of convenient substrate materials are available.

We can understand this point more clearly by referring to Fig. 40.1. This diagram plots the bandgap energy E_g of a number of important III–V semiconductors as a function of their lattice constant. The majority of optoelectronic devices for the red/near-infrared spectral regions are either grown on GaAs or InP substrates. The simplest case to consider is an epitaxial layer of GaAs grown on a GaAs substrate, which gives an emission wavelength of 873 nm (1.42 eV). This wavelength is perfectly acceptable for applications involving short-range transmission down optical fibers or for free-space infrared communication. However, for long distances we require emission at 1.3 or 1.55 μm, while for many other applications we require emission in the visible spectral region.

Let us first consider the preferred fiber-optic wavelengths of 1.3 and 1.55 μm. There are no binary semiconductors with bandgaps at these wavelengths, and so we have to use alloys to tune the bandgap by varying the composition (Chap. 30). A typical example is the ternary alloy $Ga_xIn_{1-x}As$, which is lattice matched to InP when $x = 47\%$, giving a bandgap of 0.75 eV (1.65 μm). $Ga_{0.47}In_{0.53}As$ photodiodes grown on InP substrates make excellent detectors for 1.3 or 1.55 μm radiation, but to make an emitter at this wavelength, we have to increase the band gap while maintaining the lattice-matching condition. This is achieved by incorporating a fourth element into the alloy – typically Al or P – which gives an extra design parameter that permits bandgap tuning while maintaining lattice matching. Thus, the quaternary alloys $Ga_{0.27}In_{0.73}As_{0.58}P_{0.42}$ and $Ga_{0.40}In_{0.60}As_{0.85}P_{0.18}$ give emission at 1.3 and 1.55 μm, respectively, and are both lattice matched to InP substrates.

Turning now to the visible spectral region, it is a convenient coincidence that the lattice constants of GaAs and AlAs are almost identical. This means that we can grow relatively thick layers of $Al_xGa_{1-x}As$ on GaAs substrates without introducing dislocations and other defects. The bandgap of $Al_xGa_{1-x}As$ at room temperature varies approximately quadratically with x

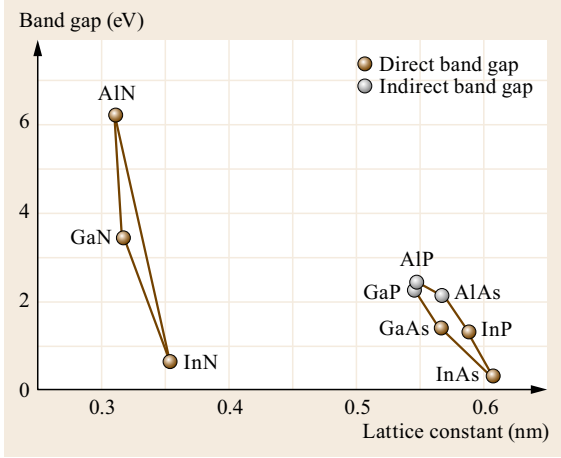


Fig. 40.1 Room-temperature bandgap of a number of important optoelectronic III–V materials versus their lattice constant

according to

$$E_g(x) = (1.42 + 1.087x + 0.438x^2)\,\text{eV}\,, \qquad (40.1)$$

but unfortunately the gap becomes indirect for $x > 43\%$. We can therefore engineer direct bandgaps from 1.42 to 1.97 eV, giving emission from 873 nm in the near infrared to 630 nm in the red spectral range. Much work has been done on quaternary alloys such as AlGaInP (Chap. 30), but it has not been possible to make blue- and green-emitting devices based on GaAs substrates to date, due to the tendency for arsenic and phosphorous compounds to become indirect as the bandgap increases.

The approach for the blue/green end of the spectrum preferred at present is to use nitride-based compounds (Chap. 31). Early work on nitrides established that their large direct gaps made them highly promising candidates for use as blue/green emitters [40.7]. However, it was not until the 1990s that this potential was fully realized. The rapid progress followed two key developments, namely the activation of the p-type dopants and the successful growth of strained $In_xGa_{1-x}N$ QWs that did not satisfy the lattice-matching condition [40.8]. The second point goes against the conventional wisdom of bandgap engineering and highlights the extra degrees of freedom afforded by quantum-confined structures, as will now be discussed.

40.1.2 Quantum-Confined Structures

A quantum-confined structure is one in which the motion of the electrons (and/or holes) is confined in one or more directions by potential barriers. The general scheme for classifying quantum-confined structures is given in Table 40.1. In this chapter, we are concerned primarily with *QWs*, although we will briefly refer to *quantum wires* and *quantum dots* in Sect. 40.6. Quantum size effects become important when the thickness of the layer becomes comparable with the de Broglie wavelength of the electrons or holes. If we consider the free thermal motion of a particle of mass m in the z-direction, the de Broglie wavelength at a temperature T is approximately given by

$$\lambda_{\text{deB}} \sim \frac{h}{\sqrt{mk_BT}}\,. \qquad (40.2)$$

For an electron in GaAs with an effective mass of $0.067\,m_0$, we find $\lambda_{\text{deB}} \approx 42$ nm at 300 K. This implies that we need structures of thickness ≈ 10 nm in order to observe quantum confinement effects at room temperature. Layers of this thickness are routinely grown by the MBE or MOVPE techniques described in Chap. 14.

Figure 40.2 shows a schematic diagram of a GaAs/AlGaAs QW. The quantum confinement is provided by the discontinuity in the bandgap at the interfaces, which leads to a spatial variation of the conduction and valence bands as shown in the lower half of the figure. The Al concentration is typically chosen to be around 30%, which gives a bandgap discontinuity of 0.36 eV according to (40.1). This splits the conduction and valence bands roughly at 2 : 1, so that electrons see a confining barrier of 0.24 eV and the holes see 0.12 eV.

If the GaAs layers are thin enough according to the criterion given above, the motion of the electrons and holes will be quantized in the growth (z) direction, giving rise to a series of discrete energy levels, as indicated by the dashed lines inside the QW in Fig. 40.2. The motion in the other two directions (i. e., the xy plane) is still free, and so we have quasi two-dimensional () behavior.

Table 40.1 Classification of quantum-confined structures. In the case of QWs, the confinement direction is usually taken as the z-axis

Structure	Confined directions	Free directions (dimensionality)
Quantum well	1(z)	2 (x,y)
Quantum wire	2	1
Quantum dot (or box)	3	none

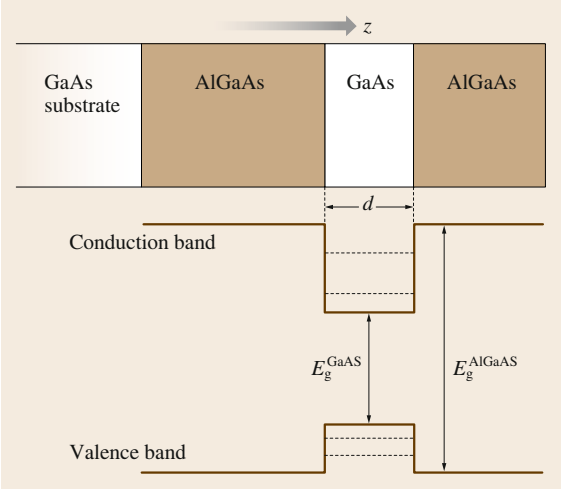


Fig. 40.2 Schematic diagram of the growth layers and resulting band diagram for a GaAs/AlGaAs QW of thickness d. The quantized levels in the QW are indicated by the *dashed lines*. Note that in real structures a GaAs buffer layer is usually grown immediately above the GaAs substrate

The quantization of the motion in the z-direction has three main consequences. Firstly, the quantization energy shifts the effective band edge to higher energy, which provides an extra degree of freedom in the art of bandgap engineering. Secondly, the confinement keeps the electrons and holes closer together and hence increases the radiative recombination probability. Finally, the density of states becomes independent of energy, in contrast to 3-D materials, where the density of states is proportional to $E^{1/2}$. Many of the useful properties of the QWs follow from these three properties.

The arrangement of the bands as shown in Fig. 40.2 in which both the electrons and holes are confined in the QW is called *type I band alignment*. Other types of band alignments are possible in which only one of the carrier types is confined within a specific region (*type II band alignment*). Furthermore, the flexibility of the MBE and MOVPE growth techniques easily allows the growth of SL structures containing many repeated QWs with thin barriers separating them, as shown in Fig. 40.3. SLs behave like artificial one-dimensional periodic crystals, in which the periodicity is designed into the structure by the repetition of the QWs. The electronic states of SLs form delocalized *minibands* as the wave functions in neighboring wells couple together through the thin barrier that separates them. Structures containing a smaller number of repeated wells or with thick barriers that prevent coupling between adjacent wells are simply called *multiple quantum well* (MQW) structures.

In the next section, we shall describe in more detail the electronic properties of QWs and SLs. Before doing so, it is worth highlighting two practical considerations that are important additional factors that contribute to their usefulness. The first is that the bandgap tunability can be achieved without using alloys as the active material, which is desirable because alloys inevitably contain more defects than simple compounds like GaAs. The second factor is that the QWs can be grown as strained layers on top of a lattice with a different unit cell size. A typical example is the $In_xGa_{1-x}N$/GaN QWs mentioned above. These layers do not satisfy the lattice-matching condition, but as long as the total $In_xGa_{1-x}N$ thickness is less than the critical value, there is an energy barrier to the formation of dislocations (Sect. 40.2.1, *Strained Quantum Wells*). In practice, this allows considerable extra flexibility in the bandgap engineering that can be achieved.

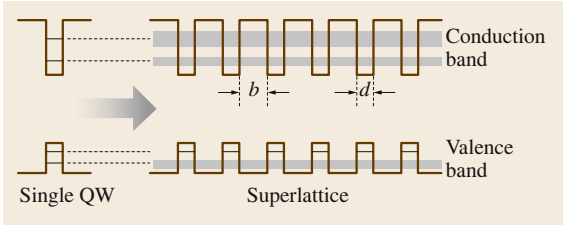


Fig. 40.3 Schematic diagram of an SL, showing the formation of minbands from the energy levels of the corresponding single QW. The structure forms an artificial one-dimensional crystal with period $(d + b)$, where d and b represent the thickness of the QW and barrier regions respectively. The width of the minibands depends on the strength of the coupling through the barriers. It is frequently the case that the lowest hole states do not couple strongly, and hence remain localized within their respective wells, as shown in the figure

40.2 Optoelectronic Properties of Quantum-Confined Structures

40.2.1 Electronic States in Quantum Wells and Superlattices

Quantum Wells

The electronic states of QWs can be understood by solving the Schrödinger equation for the electrons and holes in the potential wells created by the band discontinuities. The simplest approach is the infinite well model shown in Fig. 40.4a. The Schrödinger equation in the well is

$$-\frac{\hbar^2}{2m_w^*}\frac{d^2\psi(z)}{dz^2} = E\psi(z)\,, \tag{40.3}$$

where m_w^* is the effective mass in the well and z is the growth direction. Since the potential barriers are infinite, there can be no penetration into the barriers, and we must therefore have $\psi(z) = 0$ at the interfaces. If we choose our origin such that the QW runs from $z = 0$ to $z = d$, d being the width of the well, the normalized wave functions take the form [40.9]

$$\psi_n(z) = \sqrt{\frac{2}{d}}\sin k_n z\,, \tag{40.4}$$

where $k_n = (n\pi/d)$ and the quantum number n is an integer (≥ 1). The energy E_n is given by

$$E_n = \frac{\hbar^2 k_n^2}{2m_w^*} = \frac{\hbar^2}{2m_w^*}\left(\frac{n\pi}{d}\right)^2\,. \tag{40.5}$$

The wave functions of the $n = 1$ and $n = 2$ levels are sketched in Fig. 40.4a.

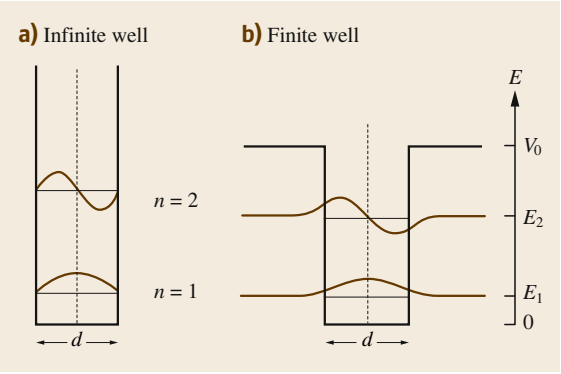


Fig. 40.4a,b Confined states in a QW of width d. **(a)** A perfect QW with infinite barriers. **(b)** A finite well with barriers of height V_0. The wave functions for the $n = 1$ and $n = 2$ levels are sketched for both types of well

Although the infinite well model is very simplified, it nonetheless provides a good starting point for understanding the general effects of quantum confinement. Equation (40.5) shows us that the energy is inversely proportional to d^2, implying that narrow wells have larger confinement energies. Furthermore, the confinement energy is inversely proportional to the effective mass, which means that lighter particles experience larger effects. This also means that the heavy- and light-hole states have different energies, in contrast to bulk semiconductors in which the two types of holes are degenerate at the top of the valence band.

Now, let us consider the more realistic finite well model as shown in Fig. 40.4b. The Schrödinger equation in the well is unchanged, but in the barrier regions we now have

$$-\frac{\hbar^2}{2m_b^*}\frac{d^2\psi(z)}{dz^2} + V_0\psi(z) = E\psi(z)\,, \tag{40.6}$$

where V_0 is the potential barrier and m_b^* is the effective mass in the barrier. The boundary conditions require that the wave function and particle flux $(1/m^*)d\psi/dz$ must be continuous at the interface. This gives a series of even and odd parity solutions that satisfy

$$\tan\left(\frac{kd}{2}\right) = \frac{m_w^*\kappa}{m_b^* k} \tag{40.7a}$$

and

$$\tan\left(\frac{kd}{2}\right) = -\frac{m_b^* k}{m_w^*\kappa} \tag{40.7b}$$

respectively. k is the wave vector in the well, given by

$$\frac{\hbar^2 k^2}{2m_w^*} = E_n\,, \tag{40.8}$$

while κ is the exponential decay constant in the barrier given by

$$\frac{\hbar^2\kappa^2}{2m_b^*} = V_0 - E_n\,. \tag{40.9}$$

Solutions to (40.7a) and (40.7b) are easily found by simple numerical techniques [40.9, 10]. As with the infinite well, the eigenstates are labeled by the quantum number n and have parities of $(-1)^{n+1}$ with respect to the axis of symmetry about the center of the well. The wave functions are approximately sinusoidal inside the well, but decay exponentially in the barriers, as illustrated in Fig. 40.4b. The eigenenergies are smaller than those of the infinite well due to the reduced confinement of the wave functions that follows from their spreading into the barriers. There are only a limited number of solutions, but there is always at least one, no matter how small V_0 might be.

As an example, we consider a typical $GaAs/Al_{0.3}Ga_{0.7}As$ QW with $d = 10\,nm$. The confinement energy is 245 meV for the electrons and 125 meV for the holes. The infinite well model predicts $E_1 = 56\,meV$ and $E_2 = 224\,meV$ for the electrons, whereas (40.7a) and (40.7b) give $E_1 = 30\,meV$ and $E_2 = 113\,meV$. For the heavy (light) holes, the infinite well model predicts 11 meV (40 meV) and 44 meV (160 meV) for the first two bound states, instead of the more accurate values of 7 meV (21 meV) and 29 meV (78 meV) calculated from the finite well model. Note that the separation of the electron levels is greater than k_BT at 300 K, so that that the quantization effects will be readily observable at room temperature.

Strained Quantum Wells

Even more degrees of freedom for tailoring the electronic states can be achieved by epitaxially stacking semiconductor layers with different lattice constants to form *strained QWs*. Examples include $In_xGa_{1-x}As$ on GaAs, $In_xGa_{1-x}N$ on GaN, and $Si_{1-x}Ge_x$ on Si. Large biaxial strain develops within the xy plane of a QW grown on a substrate with a different lattice constant. In order to avoid the build-up of misfit dislocations at the interfaces, the strained layers need to be thinner than a certain critical dimension. For example, a defect-free strained $In_xGa_{1-x}As$-on-GaAs layer requires a thickness less than around 10 nm when $x = 0.2$. Since the band structure is related to the lattice constant, the strain alters several key band parameters. It is due to some of these effects that strained QW structures have become widely exploited in optoelectronic devices.

The most significant effect of the strain in GaAs-based materials is to alter the bandgap and remove the valence band degeneracy near the Γ valley. The split-

ting of the valence band is a consequence of the lattice distortion that reduces the crystal symmetry from cubic to tetragonal [40.11]. There are essentially two types of strain. Compressive strain occurs when the active layer has a larger lattice constant than the substrate, for example, in $In_xGa_{1-x}As$ on GaAs. In this case, the bandgap increases and the effective mass of the highest hole band decreases, while that of the next valence band increases. The opposite case is that of tensile strain, which occurs when the active layer has a smaller lattice constant than the substrate, such as $Si_{1-x}Ge_x$ on Si. The ordering of the valence bands is opposite to the case of compressive strain, and the overall bandgap is reduced.

Superlattices

The analytical derivation of the allowed energy values in an SL is similar to that for a single QW, with the appropriate change of the boundary conditions imposed by the SL periodicity. The mathematical description of an SL is similar to a one-dimensional crystal lattice, which allows us to borrow the formalism of the band theory of solids, including the well-known Kronig–Penney model [40.9]. Within this model, the electron envelope wave function $\psi(z)$ can be expressed as the superposition of Bloch waves propagating along the z-axis. For an SL with a barrier height V_0, the allowed energy is calculated numerically as a solution of the transcendental equation involving the Bloch wave vector

$$\cos(ka) = \cos(kd)\cos(\kappa' b) - \frac{k^2+\kappa'^2}{2k\kappa'}\sin(kd)\sin(\kappa' b)\,, \quad E > V_0\,, \tag{40.10a}$$

$$\cos(ka) = \cos(kd)\cos(\kappa b) - \frac{k^2-\kappa^2}{2k\kappa}\sin(kd)\sin(\kappa b)\,, \quad E < V_0\,, \tag{40.10b}$$

where $a \equiv (b+d)$ is the period, and k and κ are given by (40.8) and (40.9), respectively. The decay constant κ' is given by

$$E - V_0 = \frac{\hbar^2\kappa'^2}{2m_b^*}\,. \tag{40.11}$$

The electronic states in SLs can be understood in a more qualitative way by referring to Fig. 40.3 and making use of the analogy with the tight-binding model of band formation in solids. Isolated atoms have discrete energy levels that are localized on the individual atom sites. When the atoms are brought close together, the energy levels broaden into bands, and the overlapping wave functions develop into extended states. In the same way, repeated QW structures with large values of the barrier thickness b (i. e., MQWs) have discrete levels with wave functions localized within the wells. As the barrier thickness is reduced, the wave functions of adjacent wells begin to overlap and the discrete levels broaden into *minibands*. The wave functions in the minibands are delocalized throughout the whole SL. The width of the miniband depends on the cross-well coupling, which is determined by the barrier thickness and the decay constant κ (40.9). In general, the higher lying states give rise to broader minibands because κ decreases with E_n. Also, the heavy-hole minibands are narrower than the electron minibands, because the cross-well coupling decreases with increasing effective mass.

40.2.2 Interband Optical Transitions

Absorption

The optical transitions in QWs take place between electronic states that are confined in the z-direction but free in the xy plane. The transition rate can be calculated from Fermi's golden rule, which states that the probability for optical transitions from the initial state $|i\rangle$ at energy E_i to the final state $|f\rangle$ at energy E_f is given by

$$W(i \to f) = \frac{2\pi}{\hbar}|\langle f|e\boldsymbol{r}\cdot\mathbf{E}|i\rangle|^2 g(\hbar\omega)\,, \tag{40.12}$$

where $e\boldsymbol{r}$ is the electric dipole of the electron, $\mathbf{E}$ is the electric field of the light wave, and $g(\hbar\omega)$ is the joint density of states at photon energy $\hbar\omega$. Conservation of energy requires that $E_f = (E_i + \hbar\omega)$ for absorption, and $E_f = (E_i - \hbar\omega)$ for emission.

Let us consider a transition from a confined hole state in the valence band with the quantum number n to a confined electron state in the conduction band with the quantum number n'. We apply Bloch's theorem to write the wave functions in the following form

$$|i\rangle = u_v(\boldsymbol{r})\exp(i\boldsymbol{k}_{xy}\cdot r_{xy})\psi_{hn}(z)$$
$$|f\rangle = u_c(\boldsymbol{r})\exp(i\boldsymbol{k}_{xy}\cdot r_{xy})\psi_{en'}(z)\,, \tag{40.13}$$

where $u_v(\boldsymbol{r})$ and $u_c(\boldsymbol{r})$ are the envelope functions for the valence and conduction bands, respectively, $\boldsymbol{k}_{xy}$ is the in-plane wave vector for the free motion in the xy plane, r_{xy} being the xy component of the position vector, and $\psi_{hn}(z)$ and $\psi_{en'}(z)$ are the wave functions for the confined hole and electron states in the z-direction. The photon wave vector is very small, and so conservation of momentum requires that the in-plane wave vectors of the electron and hole are the same.

On inserting these wave functions into (40.12), we find that the transition rate is proportional to both the

square of the overlap of the wave functions and the joint density of states [40.12]

$$W \propto |\langle \psi_{en'}(z)|\psi_{hn}(z)\rangle|^2 g(\hbar\omega)\,. \tag{40.14}$$

The wave functions of infinite wells are orthogonal unless $n = n'$, which gives a selection rule of $\Delta n = 0$. For finite wells, the $\Delta n = 0$ selection rule is only approximately obeyed, although transitions between states of different parity (i. e., Δn odd) are strictly forbidden. The joint density of states is independent of energy due to the quasi-2-D nature of the QW.

Figure 40.5a illustrates the first two strong transitions in a typical QW. These are the $\Delta n = 0$ transitions between the first and second hole and electron levels. The threshold energy for these transitions is equal to

$$\hbar\omega = E_g + E_{hn} + E_{en}\,. \tag{40.15}$$

The lowest value is thus equal to $(E_g + E_{h1} + E_{e1})$, which shows that the optical bandgap is shifted by the sum of the electron and hole confinement energies. Once the photon energy exceeds the threshold set by (40.15), a continuous band of absorption occurs with the absorption coefficient independent of energy due to the constant 2-D density of states of the QW. The difference between the absorption of an ideal QW with infinite barriers and the equivalent bulk semiconductor is illustrated in Fig. 40.5b. In the QW, we find a series of steps with constant absorption coefficients, whereas in the bulk the absorption varies as $(\hbar\omega - E_g)^{1/2}$ for $\hbar\omega > E_g$. Thus, the transition from 3-D to 2-D alters the shape of the absorption curve, and also causes an effective shift in the bandgap by $(E_{h1} + E_{e1})$.

Up to this point, we have neglected the Coulomb interaction between the electrons and holes that are involved in the transition. This attraction leads to the formation of bound electron–hole pairs called excitons. The exciton states of a QW can be modeled as 2-D hydrogen atoms in a material with a relative dielectric constant ε_r. In this case, the binding energy E^X is given by [40.13]

$$E^x(\nu) = \frac{\mu}{m_0}\frac{1}{\varepsilon_r^2}\frac{1}{\left(\nu - \frac{1}{2}\right)^2}R_H\,, \tag{40.16}$$

where ν is an integer ≥ 1, m_0 is the electron mass, μ is the reduced mass of the electron–hole pair, and R_H is the Rydberg constant for hydrogen (13.6 eV). This contrasts with the standard formula for 3-D semiconductors in which E^X varies as $1/\nu^2$ rather than $1/(\nu - 1/2)^2$, and implies that the binding energy of the ground-state exciton is four times larger in 2-D than 3-D. This allows excitonic effects to be observed at room temperature in QWs, whereas they are only usually observed at low temperatures in bulk semiconductors.

Figure 40.6 compares the band edge absorption of a GaAs MQW sample with that of bulk GaAs at room temperature [40.14]. The MQW sample contained 77 GaAs QWs of thickness 10 nm with thick $Al_{0.28}Ga_{0.72}As$ barriers separating them. The shift of the band edge of the MQW to higher energy is clearly observed, together with the series of steps due to each $\Delta n = 0$ transition. The sharp lines are due to excitons,

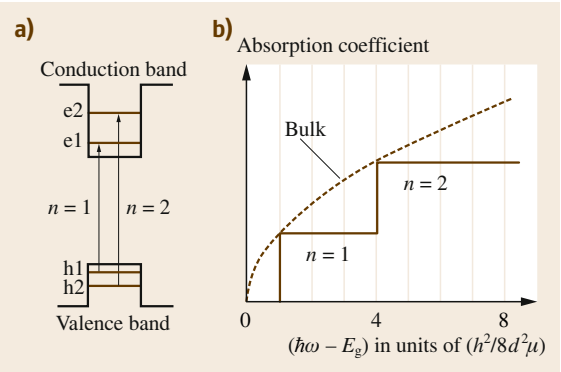


Fig. 40.5a,b Interband optical transitions between confined states in a QW. **(a)** Schematic diagram showing the $\Delta n = 0$ transitions between the $n = 1$ and $n = 2$ subbands. **(b)** Absorption spectrum for an infinite QW of thickness d with a reduced electron–hole mass μ in the absence of excitonic effects. The absorption spectrum of the equivalent bulk semiconductor in shown by the *dashed line* for comparison

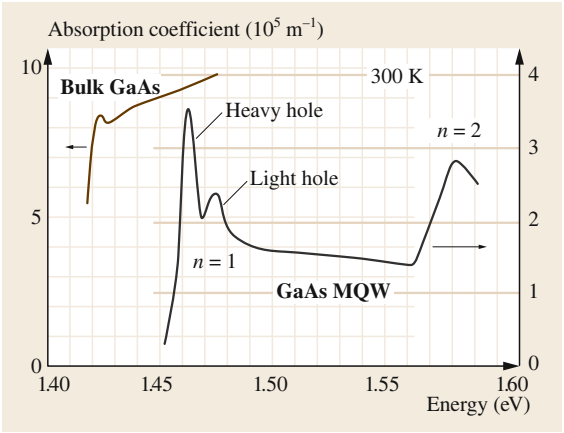


Fig. 40.6 Absorption spectrum of a 77-period GaAs/$Al_{0.28}Ga_{0.72}As$ MQW structure with 10 nm QWs at room temperature. The absorption spectrum of bulk GaAs is included for comparison. (After [40.14], with permission from AIP)

which occur at energies given by

$$\hbar\omega = E_g + E_{hn} + E_{en} - E^X \,. \tag{40.17}$$

Equation (40.16) predicts that E^X should be around 17 meV for the ground-state exciton of an ideal GaAs QW, compared to 4.2 meV for the bulk. The actual QW exciton binding energies are somewhat smaller due to the tunneling of the electrons and holes into the barriers, with typical values of around 10 meV. However, this is still substantially larger than the bulk value and explains why the exciton lines are so much better resolved for the QW than the bulk. The absorption spectrum of the QW above the exciton lines is approximately flat due to the constant density of states in 2-D, which contrasts with the rising absorption of the bulk due to the parabolic 3-D density of states. Separate transitions are observed for the heavy and light holes. This follows from their different effective masses, and can also be viewed as a consequence of the lower symmetry of the QW compared to the bulk.

Emission

Emissive transitions occur when electrons excited in the conduction band drop down to the valence band, and recombine with holes. The optical intensity $I(\hbar\omega)$ is proportional to the transition rate given by (40.12) multiplied by the probability that the initial state is occupied and the final state is empty

$$I(\hbar\omega) \propto W(\mathrm{c} \rightarrow \mathrm{v}) f_c (1 - f_v) \,, \tag{40.18}$$

where f_c and f_v are the Fermi–Dirac distribution functions in the conduction and valence bands respectively. In thermal equilibrium, the occupancy of the states is largest at the bottom of the bands and decays exponentially at higher energies. Hence, the luminescence spectrum of a typical GaAs QW at room temperature consists of a peak of width $\approx k_B T$ at the effective bandgap of $(E_g + E_{hh1} + E_{e1})$. At lower temperatures, the spectral width is affected by inhomogeneous broadening due to unavoidable fluctuations in the well thickness. Furthermore, in QWs employing alloy semiconductors, the microscopic fluctuations in the composition can lead to additional inhomogeneous broadening. This is particularly true of InGaN/GaN QWs, where indium compositional fluctuations produce substantial inhomogeneous broadening even at room temperature.

The intensity of the luminescence peak in QWs is usually much larger than that of bulk materials because the electron–hole overlap is increased by the confinement. This leads to faster radiative recombination, which then wins out over competing nonradiative decay mechanisms and leads to stronger emission. This enhanced emission intensity is one of the main reasons why QWs are now so widely used in diode lasers and LEDs.

40.2.3 The Quantum-Confined Stark Effect

The *quantum-confined Stark effect* (QCSE) describes the response of the confined electron and hole states in QWs to a strong DC electric field applied in the growth (z) direction. The field is usually applied by growing the QWs inside a p–n junction, and then applying reverse bias to the diode. The magnitude of the electric field F is given by

$$F = \frac{V^{\text{built-in}} - V^{\text{bias}}}{L_i} \,, \tag{40.19}$$

where $V^{\text{built-in}}$ is the built-in voltage of the diode, V^{bias} is the bias voltage, and L_i is the total thickness of the intrinsic region. $V^{\text{built-in}}$ is approximately equal to the bandgap voltage of the doped regions (≈ 1.5 V for a GaAs diode).

Figure 40.7 gives a schematic band diagram of a QW with a strong DC electric field applied. The field tilts the potential and distorts the wave functions as the electrons tend to move toward the anode and the holes toward the cathode. This has two important consequences for the optical properties. Firstly, the lowest transition shifts to lower energies due to the electrostatic interaction between the electric dipole induced by the field and the field itself. At low fields, the dipole is proportional to F, and the red shift is thus proportional to F^2 (the quadratic Stark effect). At higher fields, the dipole saturates at a value ultimately limited by the size of the well, and the Stark shift is linear with F. Secondly, the parity selection rule no longer applies due to the breaking of the inversion symmetry about the center of the well. This means that *forbidden* optical transitions with Δn being equal to an odd number become allowed. At the same time, the $\Delta n = 0$ transitions gradually weaken with increasing field as the distortion to the wave functions reduces the electron–hole overlap.

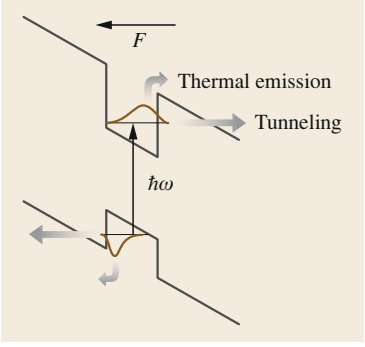


Fig. 40.7 Schematic band diagram and wave functions for a QW in a DC electric field F

Figure 40.8 shows the normalized room temperature photocurrent spectra of a GaAs/$Al_{0.3}Ga_{0.7}As$ MQW p–i–n diode containing 9.0 nm QWs at 0 and −10 V applied bias. These two bias values correspond to field strengths of around 15 and 115 kVcm^{-1}, respectively. The photocurrent spectrum closely resembles the absorption spectrum, due to the field-induced escape of the photo-excited carriers in the QWs into the external circuit (Sect. 40.2.5). The figure clearly shows the Stark shift of the absorption edge at the higher field strength, with a red shift of around 20 meV (≈ 12 nm) at the −10 V bias for the hh1 → e1 transition. The intensity of the line weakens somewhat due to the reduction in the electron–hole overlap, and there is lifetime broadening caused by the field-assisted tunneling. Several parity-forbidden transitions are clearly observed. The two most obvious ones are identified with arrows, and correspond to the hh2 → e1 and hh1 → e2 transitions, respectively.

A striking feature in Fig. 40.8 is that the exciton lines are still resolved even at very high field strengths. In bulk GaAs, the excitons ionize at around 5 kVcm^{-1} [40.12], but in QWs the barriers inhibit the field ionization, and excitonic features can be preserved even up to ≈ 300 kVcm^{-1} [40.15]. The ability to control the absorption spectrum by the QCSE is the principle behind a number of important modulator devices, as will be discussed in Sect. 40.5.

In the case of an SL, such as that illustrated in Fig. 40.3, a strong perpendicular electric field can break the minibands into discrete energy levels local to each QW, due to the bandgap tilting effect represented in Fig. 40.7. The possibility of using an electric field to modify the minibands of an SL is yet another remarkable ability of bandgap engineering to achieve control over the electronic properties by directly using fundamental principles of quantum mechanics.

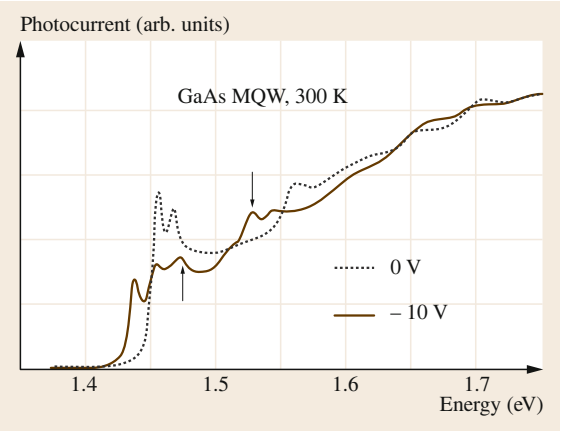


Fig. 40.8 Room-temperature photocurrent spectra for a GaAs/$Al_{0.3}Ga_{0.7}As$ MQW p–i–n diode with a 1 µm-thick i-region at the zero bias and −10 V. The QW thickness was 9.0 nm. The *arrows* identify transitions that are forbidden at zero field. (Adapted from [40.15], © 1991 IEEE)

40.2.4 Intersubband Transitions

The engineered band structure of QWs leads to the possibility of *intersubband* (ISB) *transitions* which take place between confined states within the conduction or valence bands, as illustrated schematically in Fig. 40.9. The transitions typically occur in the infrared spectral region. For example, the $e1 \leftrightarrow e2$ ISB transition in a 10 nm GaAs/AlGaAs QW occurs around 15 µm. For ISB absorption transitions, we must first dope the conduction band so that there is a large population of electrons in the e1 level as shown in Fig. 40.9a. This is typically achieved by n-type doping of the barriers, which produces a large electron density as the extrinsic electrons drop from the barriers to the confined states in the QW. Undoped wells are used for ISB emission transitions, and electrons must first be injected into excited QW states, as shown in Fig. 40.9b.

The basic properties of ISB transitions can be understood by the extension of the principles outlined in Sect. 40.2.2. The main difference is that the envelope functions for the initial and final states are the same, since both states lie in the same band. The transition rate for a conduction band ISB transition then turns out to be given by

$$W^{\mathrm{ISB}} \propto |\langle \psi_{en'}(z)|z|\psi_{en}(z)\rangle|^2 g(\hbar\omega)\,, \qquad (40.20)$$

where n and n' are the quantum numbers for the initial and final confined levels. The z operator within the Dirac bracket arises from the electric–dipole interaction and indicates that the electric field of the light wave must be parallel to the growth direction. Furthermore, the odd parity of the z operator implies that the wave functions must have different parities, and hence that Δn must be an odd number. The use of ISB transitions in infrared emitters and detectors will be discussed in Sects. 40.3.2 and 40.4.3.

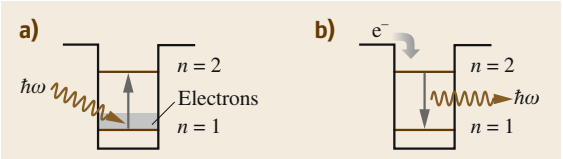


Fig. 40.9 **(a)** Intersubband absorption in an n-type QW. **(b)** Intersubband emission following electron injection to a confined upper level in the conduction band

40.2.5 Vertical Transport

Quantum Wells

Vertical transport refers to the processes by which electrons and holes move in the growth direction. Issues relating to vertical transport are important for the efficiency and frequency response of most QW optoelectronic devices. The transport is generally classified as either *bipolar*, when both electrons and holes are involved, or *unipolar*, when only one type of carrier (usually electrons) is involved. In this section, we will concentrate primarily on bipolar transport in QW detectors and QCSE modulators. Bipolar transport in light-emitting devices is discussed in Sects. 40.2.6 and 40.3.1, while unipolar transport in quantum cascade (QC) lasers and avalanche photodiodes (APDs) is discussed in Sects 40.3.2 and 40.4.3.

In QW detectors and QCSE modulators, the diodes are operated in the reverse bias. This produces a strong DC electric field and tilts the bands as shown in Fig. 40.7. Electrons and holes generated in the QWs by absorption of photons can escape into the external circuit by the tunneling and/or thermal emission, as illustrated schematically in Fig. 40.7.

The physics of tunneling in QWs is essentially the same as that of the α-decay in nuclear physics. The confined particle oscillates within the well and attempts to escape every time it hits the barrier. The escape rate is proportional to the attempt frequency ν_0 and the transmission probability of the barrier. For the simplest case of a rectangular barrier of thickness b, the escape time τ_T is given by

$$\frac{1}{\tau_T} = \nu_0 \exp(-2\kappa b) \,, \tag{40.21}$$

where κ is the tunneling decay constant given by (40.9). The factor of 2 in the exponential arises from the dependence of the transmission probability on $|\psi(z)|^2$. The situation in a biased QW is more complicated due to the nonrectangular shape of the barriers. However, Equation (40.21) allows the basic trends to be understood. To obtain fast tunneling, we need thin barriers and small κ. The second requirement is achieved by keeping m_b^* as small as possible and by working with a small confining potential V_0. The tunneling rate increases with increasing field, because the average barrier height decreases.

The thermal emission of electrons over a confining potential is an old problem that was originally applied to the heated cathodes in vacuum tubes. It has been shown that the thermal current fits well to the classical Richardson formula

$$J_E \propto T^{\frac{1}{2}} \exp\left(-\frac{e\Phi}{k_B T}\right) , \tag{40.22}$$

with the work function Φ replaced by $(V(F) - E_n)$, $V(F)$ being the height of the barrier that must be overcome at field strength F [40.16]. The emission rate is dominated by the Boltzmann factor that represents the probability that the carriers have enough thermal kinetic energy to escape over the top of the barrier. At low fields, $V(F) \approx (V_0 - E_n)$, but as the field increases, $V(F)$ decreases as the barriers tilt over. Hence, the emission rate (like the tunneling rate) increases with increasing field. The only material-dependent parameter that enters the Boltzmann factor is the barrier height. Since this is insensitive both to the effective masses and to the barrier thickness, the thermal emission rate can dominate over the tunneling rate in some conditions, especially at room temperature in samples with thick barriers. For example, the fastest escape mechanism in GaAs/$Al_{0.3}Ga_{0.7}As$ QWs at room temperature can be the thermal emission of holes, which have a much smaller barrier to overcome than the electrons [40.17].

Superlattices

The artificial periodicity of SL structures gives rise to additional vertical transport effects related to the phenomenon of Bloch oscillations. It is well known that an electron in a periodic structure is expected to oscillate when a DC electric field is applied. This effect has never been observed in natural crystals because the oscillation period (equal to h/eFa, where a is the unit cell size) is much longer than the scattering times of the electrons. In an SL, by contrast, the unit cell size is equal to $(d + b)$ (Fig. 40.3) and the oscillation period can be made much shorter. This allows the electron to perform several oscillations before being scattered. The oscillatory motion of the electrons in an SL was first observed by two groups in 1992 [40.18, 19]. The following year, another group directly detected the radiation emitted by the oscillating electron wave packet [40.20]. The subject has since developed greatly, and a comprehensive review can be found in [40.21].

40.2.6 Carrier Capture and Relaxation

In a QW light-emitting device, the emission occurs after carriers injected from the contacts are transported to the active region and then captured by the QWs. The capture and subsequent relaxation of the carriers is thus of crucial importance. Let us consider the band-edge profile of a typical QW diode laser active region, as illustrated in Fig. 40.10. The active region is embedded between larger bandgap cladding layers designed to prevent thermally assisted carrier leakage outside the active region. Electrons and holes are injected from the n- and p-doped cladding layers under forward bias and

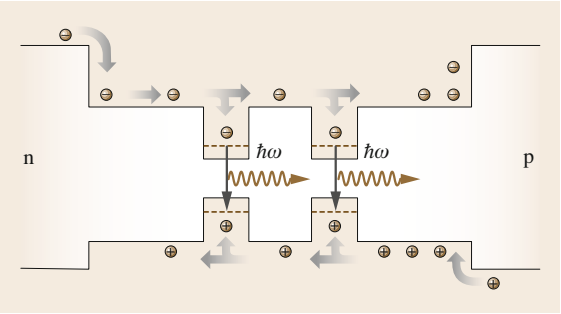


Fig. 40.10 Schematic representation of the drift of injected carriers and their subsequent capture by QWs. Light emission occurs when electrons and holes are captured in the same QW and then recombine with each other

light emission follows after four distinct processes have taken place:

1. Relaxation of carriers from the cladding layers to the confinement barriers (CB)
2. Carrier transport across the CB layers, by diffusion and drift
3. Carrier capture into the QWs
4. Carrier relaxation to the fundamental confined levels.

The carrier relaxation to the CB layers occurs mainly by longitudinal optical (LO) phonon emission. The CB layer transport is governed by a classical electron fluid model. The holes are heavier and less mobile than the electrons, and hence the ambipolar transport is dominated by the holes. Carrier nonuniformities, such as carrier pile up at the p-side CB region due to the lower mobility of the holes, are taken into account in the design of the barrier layers. The carrier capture in the QWs is governed by the phonon-scattering-limited carrier mean free path. It is observed experimentally that the capture time oscillates with the QW width. Detailed modeling reveals that this is related to a resonance between the LO phonon energy and the energy difference between the barrier states and the confined states within the well [40.22]. As another design guideline, the QW widths must be larger or at least equal to the phonon-scattering-limited carrier mean free path at the operating temperature in order to speed up the carrier capture. Finally, the relaxation of carriers to the lowest subband occurs on a subpicosecond time scale if the intersubband energy separation is larger than the LO phonon energy. Carrier–carrier scattering can also contribute to an ultrafast thermalization of carriers, on a femtosecond time scale at the high carrier densities present inside laser diodes. Many of these processes have been studied in detail by ultrafast laser spectroscopy [40.23].

Carrier capture and escape are complementary vertical transport mechanisms in MQW structures. In the design of vertical transport-based MQW devices, one process must often be sped up at the expense of making the other as slow as possible. For example, in order to enhance the performance of QW laser diodes, the carrier confinement capability of the MQW active region must be optimized in terms of minimizing the ratio between the carrier capture and escape times [40.24].

40.3 Emitters

40.3.1 Interband Light-Emitting Diodes and Lasers

QWs have found widespread use in light-emitting diode (LED) and laser diode applications for a number of years now. As discussed in Sects. 40.1.2 and 40.2.2, there are three main reasons why this is so:

1. The ability to control the quantum confinement energy provides an extra degree of freedom to engineer the emission wavelength.
2. The change of the density of states and the enhancement of the electron–hole overlap leads to superior performance.
3. The ability to grow strained layers of high optical quality greatly increases the variety of material combinations that can be employed, thus providing much greater flexibility in the design of the active regions.

Much of the early work concentrated on lattice-matched combinations such as GaAs/AlGaAs on GaAs substrates. GaAs/AlGaAs QW lasers operating around 800 nm have now become industry standard for applications in local area networks. Furthermore, the development of high power arrays has opened up new applications for pumping solid-state lasers such as Nd:YAG. Other types of lattice-matched combinations can be used to shift the wavelength into the visible spectral region and also further into the infrared. QWs based on the quaternary alloy $(Al_yGa_{1-y})_xIn_{1-x}P$, are used for red-emitting laser pointers [40.25], while

$Ga_{0.47}In_{0.53}As$ QWs and its variants incorporating Al are typically used for the important telecommunication wavelengths of 1300 and 1550 nm.

The development of strained-layer QW lasers has greatly expanded the range of material combinations that are available. The initial work tended to focus on $In_xGa_{1-x}As/GaAs$ QWs grown on GaAs substrates. The incorporation of indium into the QW shifts the band edge to lower energy, thereby giving emission in the wavelength range 900–1100 nm. An important technological driving force has been the need for powerful sources at 980 nm to pump erbium-doped fiber amplifiers [40.26]. Furthermore, as mentioned in Sect. 40.2.1, *Strained Quantum Wells* the strain alters the band structure and this can have other beneficial effects on the device performance. For example, the compressive strain in the $In_xGa_{1-x}As/GaAs$ QW system has been exploited in greatly reducing the threshold current density. This property is related to the reduced effective mass of the holes and hence the reduced density of states. An extensive account of the effects of strain on semiconductor layers and the performance of diode lasers is given in [40.27, 28].

At the other end of the spectral range, a spectacular development has been the $In_xGa_{1-x}N/GaN$ QWs grown on sapphire substrates. These highly strained QWs are now routinely used in ultrabright blue and green LEDs [40.29], while commercial laser diodes operating around 400 nm [40.8] are used in blu-ray technology and laser projectors (see Chap. 31 for a review of nitride materials). One current research area focuses on developing lasers to fill the *green gap* that lies between the blue spectral regions covered by standard nitride lasers and the red spectral region accessible to GaAs-based compounds. Conventional nitride devices are grown on *c*-plane GaN buffer layers deposited on 0001-oriented sapphire substrates. The crystal is piezoelectric in this direction, and so the QWs experience a large electric field because of the strain associated with the lattice mismatch between $In_xGa_{1-x}N$ and GaN. This in turn causes a decrease in the radiative efficiency due to the reduced overlap of the electrons and holes through the QCSE (Sect. 40.2.3). In order to reach longer wavelengths, the fraction of indium has to be increased, and this therefore leads to a drop in the efficiency, and hence an increase in the threshold current with increasing wavelength. One possible solution to this problem is to grow the devices on semipolar or nonpolar substrates, which have reduced or zero piezoelectricity, respectively [40.30–32]. The use of semi-polar or nonpolar materials may also have benefits in other applications. Nitride LEDs lie at the heart of many white-light LEDs [40.33], but their efficiency droops off at the high current levels required for their use in solid-state lighting. The reason for this efficiency droop is much debated [40.34], but the use of semipolar material has been reported to reduce the effect [40.35].

A major application of QWs is in *vertical-cavity surface-emitting lasers* (VCSELs). These lasers emit from the top of the semiconductor surface, and have several advantages over the more conventional edge emitters. Arrays are readily fabricated, which facilitates their manufacture. No facets are required, which avoids complicated processing procedures. The beam is circular which enhances the coupling efficiency into optical fibers, while their small size leads to very low threshold currents. For these reasons the development of VCSELs has been very rapid, and many local area fiber networks operating around 850 nm currently employ VCSEL devices. This would not have been possible without the high gain coefficients that are inherent to the QW structures.

Figure 40.11 gives a schematic diagram of a typical GaAs-based VCSEL. The device contains an active QW region inserted between two DBR mirrors consisting of AlGaAs quarter wave stacks made of alternating high and low refractive index layers. The structure is grown on an n-type GaAs substrate, and the mirrors are appropriately doped n- or p-type to form a p-n junction. Electrons and holes are injected into the active region under a forward bias, where they are captured by the QWs and produce gain at the lasing wavelength λ. The thickness of the layer containing the QWs is chosen so that the light at the design wavelength is reflected back and forth from the DBR mirrors through the gain medium and adds up constructively, forming

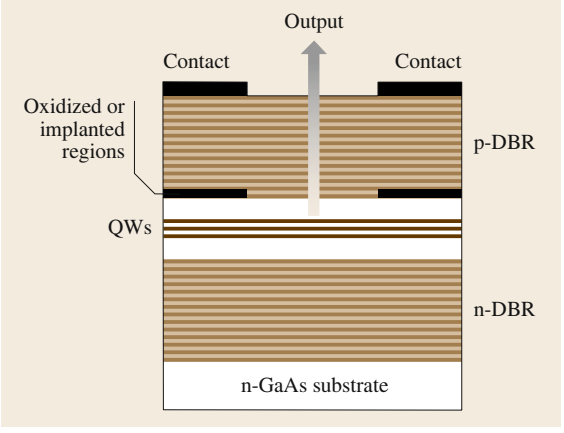


Fig. 40.11 Schematic diagram of a vertical-cavity surface-emitting laser (VCSEL). The QWs that comprise the gain medium are placed at the center of the cavity formed between two distributed Bragg reflector (DBR) mirrors. Oxidized or proton-implanted regions provide the lateral confinement for both the current and the optical mode

a laser resonator. The boundary conditions usually give field peaks close to the edges of the spacer layer, so that a thickness of $\approx \lambda/n_0$, gives an antinode at the center, where the QWs are located. Oxidized or proton-implanted regions provide lateral confinement of both the current and the optical mode. Reviews of the design and applications properties of VCSELs may be found in [40.36, 37].

The conventional VCSEL structures grown on GaAs substrates operate best in the wavelength range between 700 and 1100 nm. Much work is currently focussed on extending the range of operation to the telecommunication wavelengths of 1300 and 1550 nm, and also to shorter wavelengths in the visible spectral range. Unfortunately, it is hard to grow DBR mirrors with sufficient reflectivity on InP substrates due to the low refractive index contrast of the materials, which leads to intense research on new materials such as dilute nitrides, antimonides, and bismides that are compatible with GaAs substrates and hence GaAs/AlAs DBR mirrors [40.38–40]. Alternative approaches based on QDs will be mentioned in Sect. 40.6.

Resonator structures such as the VCSEL as shown in Fig. 40.11 can be operated below threshold as resonant cavity LEDs (RCLEDs). The presence of the cavity reduces the emission linewidth and hence increases the intensity at the peak wavelength [40.41]. Furthermore, the narrower linewidth leads to an increase in the bandwidth of the fiber communication system due to the reduced chromatic dispersion [40.42]. A review of the progress in RCLEDs is given in [40.43].

40.3.2 Quantum Cascade Lasers

The principles of infrared emission by ISB transitions were described in Sect. 40.2.4. Electrons must first be injected into an upper confined electron level as shown in Fig. 40.9b. Radiative transitions to lower confined states with different parities can then occur. ISB emission is usually very weak, as the radiative transitions have to compete with a very rapid nonradiative decay by phonon emission (Sect. 40.2.6). However, when the electron density in the upper level is large enough, population inversion can occur, giving rapid stimulated emission and subsequent laser operation. This is the operating concept of the *QC laser* first demonstrated in 1994 [40.44]. The laser operated at 4.2 μm at temperatures up to 90 K. Although the threshold current for the original device was high, the progress in the field had been very rapid. A comprehensive account of the present state of the art is given in [40.45].

The QW structures used in QC lasers are very complicated, and often contain hundreds of different layers. Figure 40.12 illustrates a relatively simple design based

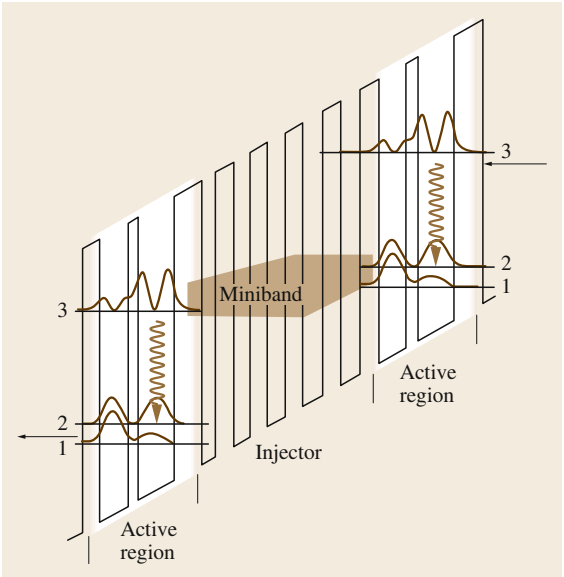


Fig. 40.12 Conduction band diagram for two active regions of an InGaAs/AlInAs QC laser, together with the intermediate miniband injector region. The levels in each active region are labeled according to their quantum number n, and the corresponding wave function probability densities are indicated. Laser transitions are indicated by the wavy arrows, while electron tunneling processes are indicated by the straight arrows. (After [40.46], with permission from IOP)

on lattice-matched $In_{0.47}Ga_{0.53}As/Al_{0.48}In_{0.52}As$ QWs grown on an InP substrate [40.46]. The diagram shows two active regions and the miniband injector region that separates them. A typical operational laser might contain 20–30 such repeat units. The population inversion is achieved by resonant tunneling between the $n = 1$ ground state of one active region and the $n = 3$ upper laser level of the next one. The basic principles of this process were enunciated as early as 1971 [40.47], but it took more than 20 years to demonstrate the ideas in the laboratory. The active regions contain asymmetric coupled QWs, and the laser transition takes place between the $n = 3$ and $n = 2$ states of the coupled system. The separation of the $n = 2$ and $n = 1$ levels is carefully designed to coincide with the LO-phonon energy, so that very rapid relaxation from the $n = 2$ level to the $n = 1$ level occurs, and the system behaves as a four-level laser. This latter point is crucial, since the lifetime of the upper laser level is very short (typically ≈ 1 ps), and population inversion is only possible when the lifetime of the lower laser level is shorter than that of the upper one. The lasing wavelength can be varied by the detailed design of the coupled QW active region. The transition energy for the design as shown in Fig. 40.12

is 0.207 eV, giving emission at 6.0 μm. Further details may be found in [40.46].

Present research on QC lasers focuses on practical systems for chemical sensing and medical imaging, and also on extending their performance at both the short (< 5 μm) and long (> 50 μm) wavelength ends. At short wavelengths, new material systems based on antimonides offer much promise [40.48], while at the long wavelengths, there is much interest in developing sources for THz imaging [40.49, 50].

40.4 Detectors

Photodetectors for the visible and near-infrared spectral regions are generally made from bulk silicon or III–V alloys such as GaInAs. Since these devices work very well, the main applications for QW photodetectors are in the infrared spectral region and for especially demanding applications such as APDs and solar cells. These three applications are discussed separately below, starting with solar cells.

40.4.1 Solar Cells

The power generated by a solar cell is determined by the product of the photocurrent and the voltage across the diode. In conventional single-junction cells, both of these parameters are determined by the bandgap of the semiconductor used. Large photocurrents are favored by small-gap materials, because semiconductors only absorb photons with energies greater than the bandgap, and small-gap materials therefore absorb a larger fraction of the solar spectrum. However, the largest open circuit voltage that can be generated in a p-n device is the built-in voltage, which increases with the bandgap of the semiconductor. This compromise sets a theoretical upper limit to the efficiency of single-junction devices at 31%, with the best silicon devices achieving efficiencies of around 25% [40.51].

QW devices can give better performance than the bulk counterparts because they permit separate optimization of the current and voltage generating factors [40.52]. This is because the built-in voltage is primarily determined by the bandgap of the barrier regions, whereas the absorption edge is determined by the band gap of the QWs. This gives rise to improvements in the efficiency compared to silicon devices, with 28.3% achieved for a GaAsP/InGaAs strain-balanced single-junction QW device in a concentrator cell under 535 sun illumination [40.53]. This enhanced efficiency under high-sun illumination is significant, because the performance of other types of solar cells tends to deteriorate in the higher operating temperatures found in concentrators.

The benefits of QWs become more apparent when they are used in tandem solar cells that incorporate multiple junctions to improve the efficiency. The idea here is to form a solar cell with a series of junctions that are separately optimized to capture energy from different parts of the solar spectrum. The top junction has the widest gap and is designed to absorb the highest energy (shortest wavelength) photons from the sun. This top junction is transparent to lower energy photons, which can then be absorbed efficiently in lower junctions with smaller gaps, making it possible to capture a larger part of the solar spectrum efficiently. In this way, efficiencies surpassing the 31% limit of single-junction cells have been achieved, and the versatility provided by QWs can provide significant benefits in the design of the device. The increased cost of the III–V epitaxy can then become worthwhile in space applications where the use of the highest efficiency devices is paramount to minimize the size and weight of the solar panels. A comprehensive review of the present state of the art in QW solar cells may be found in [40.54], and new research directions on nanowire and QD devices are discussed briefly in Sect. 40.6.

40.4.2 Avalanche Photodiodes

APDs are the detectors of choice for many applications in telecommunications, single-photon counting and imaging. For telecommunication applications, APDs are generally operated in linear gain mode, where the output photocurrent is proportional to the reverse bias applied below breakdown. On the other hand, single-photon detection and high speed imaging applications require APD operation in *Geiger-mode* – when the reverse bias voltage is pulsed above breakdown [40.55].

A variety of semiconductor materials has been researched for optimizing APD design. The material of choice in terms of fabrication process robustness and cost-effectiveness is obviously silicon – but the drawback is the limitation to visible light detection. For example, focal plane arrays of silicon single-photon avalanche diodes (SPADs) have been in development for high-speed image sensor design and fabrication [40.56]. In order to extend the photoresponse into the near infrared, SiGe/Si multiple QW APDs have been developed, based on CMOS process-

ing, and demonstrated to be exhibiting photo-response in the spectral range 800–1500 nm [40.57].

Commercially available III–V semiconductor APDs are typically engineered with different bandgap materials in the absorption and multiplication regions [40.55]. The absorption layer has a relatively narrow bandgap (such as $In_{0.53}Ga_{0.47}As$) to allow for larger absorption, whereas the multiplication region has a wider bandgap (such as InP or $In_{0.52}Al_{0.48}As$) to reduce dark current at the high electric fields required. The avalanche multiplication mechanism plays a critical role in determining the APD photodetection gain, dark current, and gain–bandwidth product. It is the design of the thickness of the multiplication region that makes the fundamental difference between optimizing APDs for linear mode versus Geiger mode. To increase the gain–bandwidth product and reduce noise in linear mode operation, the multiplication region must be thinned down [40.58], whereas for Geiger-mode photon counting the SPAD's multiplication region must be thicker [40.59]. There are many design trade-offs involved, the most critical being the onset of band-to-band tunneling and dark current increase with the thinning of the multiplication region [40.60]. To circumvent this problem, it has been demonstrated that the dark current can be reduced by incorporating adequately designed SLs into the multiplication layer to form an SL-APD as shown in Fig. 40.13 [40.61]. Essentially, the miniband formation (Fig. 40.3) in the SL corresponds to a larger effective bandgap and thus reduces the probability of band-to-band tunneling at the high electric fields required for carrier impact ionization and subsequent carrier multiplication.

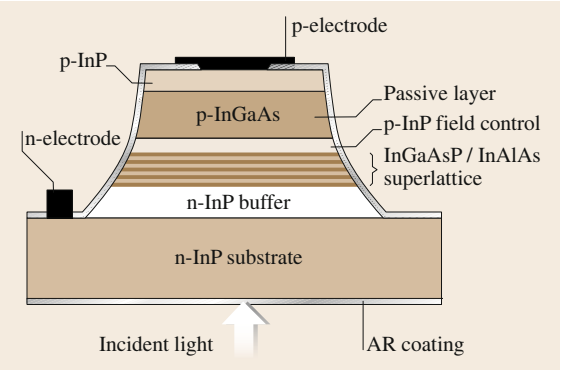


Fig. 40.13 Schematic representation of an InGaAs/InP/InGaAsP/InAlAs superlattice avalanche photodiode (SL-APD). Light is absorbed in the bulk InGaAs layer and the resulting photocurrent is multiplied by the avalanche process in the InGaAsP/InAlAs superlattice region. The spatial periodicity of the SL reduces the dark current

It has been proposed that the excess noise factor can be further reduced by designing multiplication layers with SL [40.62] or *staircase* [40.63] structures to enhance the ionization rate of one carrier type relative to the other due to their different band edge discontinuities [40.64]. However, this proposal is still under active theoretical and experimental scrutiny, because most of the electrons populate the higher energy satellite valleys (X and L) at the high electric fields required for the avalanche gain [40.65]. Monte Carlo simulations [40.66], backed up by experimental evidence [40.67], have shown that the ratio of the electron and hole ionization rates, commonly used as a figure of merit in bulk multiplication layers [40.68], is not critically affected by the bandgap discontinuities. In fact, these models emphasize that it is the spatial modulation of the impact ionization probability, associated with the periodic bandgap discontinuity, which leads to a reduction in the multiplication noise.

40.4.3 Infrared Detectors

The principles of infrared absorption by ISB transitions were described in Sect. 40.2.4. Infrared detectors are required for applications in defence, night vision, astronomy, thermal mapping, gas-sensing, and pollution monitoring. *Quantum well intersubband photodetectors* (QWIPs) are designed so that the energy separation of the confined levels is matched to the chosen wavelength. A major advantage of QWIPs over the conventional approach employing narrow-gap semiconductors is the use of mature GaAs-based technologies, and the avoidance of toxic metals such as mercury and cadmium. Furthermore, the detection efficiency should in principle be high due to the large oscillator strength that follows from the parallel curvature of the in-plane dispersions for states within the same bands [40.69]. However, technical challenges arise from the requirement that the electric field of the light must be polarized along the growth (z) direction. This means that QWIPs do not work at normal incidence, unless special steps are taken to introduce a component of the light field along the z-direction, for example, by using gratings, random scattering or microlenses [40.70]. A major area of current research focuses on developing arrays of QWIPs for applications in imaging [40.71, 72].

Much of the early work on QW infrared detectors was based on the established GaAs/AlGaAs and GaAs/InGaAs material systems. In recent years, however, interest has shifted to other QW materials, especially to type-II InAs/GaSb SL (T2SL) devices [40.73, 74]. In T2SLs, the absorption occurs between hole states at the top of the GaSb valence band and electron miniband states within the InAs conduction band. This

means that they are not actually intersubband devices per se, and hence are not subject to the z-polarization selection rule for ISB absorption. They are therefore sensitive to photons impinging at normal incidence, making the design of the detector arrays much simpler. In Sect. 40.6, we briefly consider another class of III–V nanostructures that also absorb infrared photons at normal incidence based on QDs.

The combination of resonant intersubband (ISB) photodetection and avalanche multiplication (Sect. 40.4.2 above) has been studied to explore the possibility of designing *unipolar avalanche photodiodes* (UAPDs) [40.75]. UAPDs rely on impact ionization involving only one type of carrier in order to achieve gain in photoconductive detectors for mid- and far-infrared light. The unipolar impact ionization occurs when the kinetic energy of the primary carrier exceeds an activation energy E_a defined as the transition energy between the QW ground state and the QW top state (Fig. 40.14). A single QW then releases an extra electron each time it is subject to an impact with an incoming electron.

The impact ionization probability for this process is given by the product between the carrier capture probability and the carrier escape probability under a mechanism of carrier–carrier scattering in the QW. The electron transport occurs through a sequence of escape, drift, and kinetic energy gain under the applied electric field, combined with relaxation, capture and QW ionization. Ultimately, the multiplication gain in a UAPD is governed by the QW capture probability, the number of QWs, and the field uniformity over the QW sequence. The unipolar nature of the multiplication process must be preserved in order to avoid field nonuniformities stemming from spatial charge variation caused by bipolar carrier transport across the multiplication region.

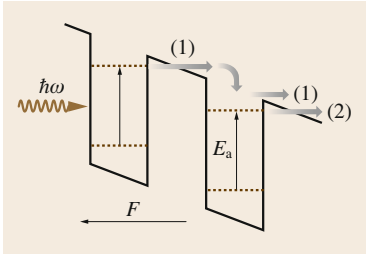


Fig. 40.14 Unipolar carrier multiplication in a multiple QW structure at field strength F: (1) is the primary electron resulting from photodetection, while (2) is the secondary electron resulting from the impact ionisation of the QW by the primary electron

Interest in a purely unipolar multiplication mechanism was originally motivated by the possibility of reduced noise in comparison to bipolar APDs, where band-to-band transitions lead to gain fluctuations manifested as excess noise [40.68]. For this purpose, the QWs in an UAPD structure are typically tailored such that the intersubband activation energy E_a is smaller than the interband impact ionization activation energy that would be responsible for bipolar avalanche multiplication. However, further investigation has shown that unipolar avalanche multiplication is also accompanied by an excess noise factor, such that the noise gain exceeds the photoconductive gain [40.76], thus limiting the practical applications of UAPDs.

40.5 Modulators

In Sect. 40.2.3, we noted that the optical properties of QW diodes are strongly modified by the application of voltages through the quantum-confined Stark effect (QCSE). Referring to Fig. 40.8, we see that at wavelengths below the heavy-hole exciton at 0 V, the absorption increases as the voltage is applied, which provides a mechanism for the modulation of light. For example, the amount of light transmitted at wavelengths close to the bandedge would change with the voltage applied. Moreover, since changes of absorption are accompanied by changes of the refractive index through the Kramers–Kronig relationship, it is possible to make QCSE phase modulators as well [40.77].

Figure 40.15 shows a schematic design of a QW *electroabsorption modulator* (EAM) device incorporated in a waveguide geometry on the same chip as a semiconductor laser. The QCSE modulator transmits the incoming laser light when no voltage is applied and absorbs the beam when the MQW stack is suitably biased. For operation at high speeds, the combination of a continuous laser and a high-speed modulator offers better control of the phase chirp of the pulses than the direct modulation of the laser output itself [40.78]. In particular, the chirp factor is expected to become negative if the photo-generated carriers can be swept out fast enough in the EAM, which is desirable for long-distance data transmission through optical fibers [40.79].

The growth of silicon photonics in recent years has led to a strong interest in QW modulators that can be grown on silicon substrates. Silicon is transparent at

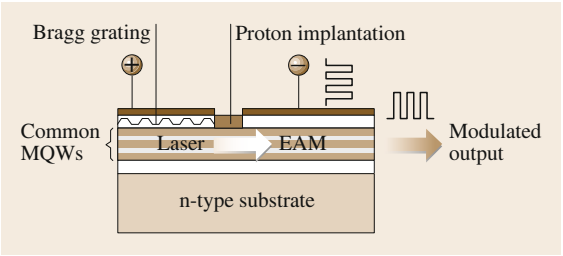


Fig. 40.15 Schematic diagram of an integrated quantum-well waveguide electroabsorption modulator (EAM) and distributed feedback (DFB) laser. The laser is forward biased while the EAM is reverse biased. The p-contacts of the two electrically independent devices are separated by proton implantation. The light emitted by the laser is guided through the EAM region, resulting in a modulated output beam when data pulses are applied to the EAM ►

the technologically important wavelengths of 1300 and 1550 nm, and this enables the development of sophisticated integrated photonic circuits that are compatible with CMOS electronics. Off-chip III–V lasers are coupled into the circuits and are then encoded with data by on-chip modulators. Hybrid devices composed of III–V QCSE modulators grown on silicon substrates have achieved impressive results [40.80]. However, attractive alternative approaches based on germanium QWs have developed following the observation of a strong QCSE around 1400 nm in a Ge/GeSi MQW structure grown on silicon in 2005 [40.81]. Operation at 1550 nm was achieved a year later [40.82], and high-speed devices operating at 23 GHz have now been demonstrated [40.83].

40.6 Quantum Dots and Nanowires

The main emphasis of this chapter has been on outlining the benefits of using QWs and SLs in modern photonic devices. The chapter would not be complete, however, without a brief discussion of semiconductor nanostructures with higher degrees of electronic confinement, namely quantum wires and quantum dots (Table 40.1).

Quantum wires are nanostructures in which the electrons are confined in two dimensions, and free in the third. They can therefore be considered as one-dimensional materials. Laser operation from a GaAs quantum wire formed in a V-groove structure was demonstrated in as early as 1989 [40.86]. However, the direct development of that work was difficult due to the complexity of making quantum wires by the V-groove technique. By contrast, in recent years, there has been very rapid progress in the development of semiconductor nanowires by growth on patterned substrates. Figure 40.16 illustrates the two main approaches [40.84]. In the first method shown in Fig. 40.16a, size-selected metal nanoparticles are deposited on the substrate to act as catalysts for the nanowire growth; the wires then form under the particles by the process of vapor–liquid–solid (VLS) growth. The second method shown in Fig. 40.16b does not make use of catalysts; instead, an array of nanoapertures is patterned into a mask on top of the substrate by e-beam lithography, and the nanowires grow from the holes by the process of selective-area vapor-phase epitaxy (SA-VPE). By these, and similar methods, it is possible to grow large arrays of nanowires for a variety of applications in electronics and photonics. For example, Figs. 40.16c and 40.16d illustrate an array of GaAs nanowires grown by MOVPE on a square array of holes patterned into a SiO_2 mask deposited on a GaAs substrate [40.85]. One particularly interesting potential application is in the development of high-efficiency nanowire-array photovoltaic devices [40.87]. Another major development has been the observation of exotic particles called Majorana fermions in transport studies of InSb nanowires coupled to superconducting electrodes [40.88]. Reviews on the progress of the growth and applications of nanowires may be found in [40.89, 90].

The final member of the family of quantum-confined structures is the QD. These are semiconductor nanostructures in which the electrons are confined in all three directions. This gives rise to zero-dimensional (0-D) physics, in which the continuous electronic bands that are characteristic of bulk semiconductors are broken in discrete quantized states. QDs for photonic applications are usually formed by either epitaxial or colloidal growth. Epitaxial QDs form spontaneously during MBE growth in the Stranski–Krastanow regime, while colloidal QDs are grown in a solution from precursors. Figure 40.17 shows an electron microscope image of an InAs QD grown on a GaAs crystal by the *Stranski–Krastanow* technique [40.91]. The dots are formed because of the very large mismatch between the lattice constants of the InAs and the underlying GaAs. The strain that would be produced in a uniform layer is so large that it is energetically favorable to form small clusters. This then leads to the formation of islands of InAs with nanoscale dimensions, which can then be encapsulated within an optoelectronic structure by the overgrowth of further GaAs layers.

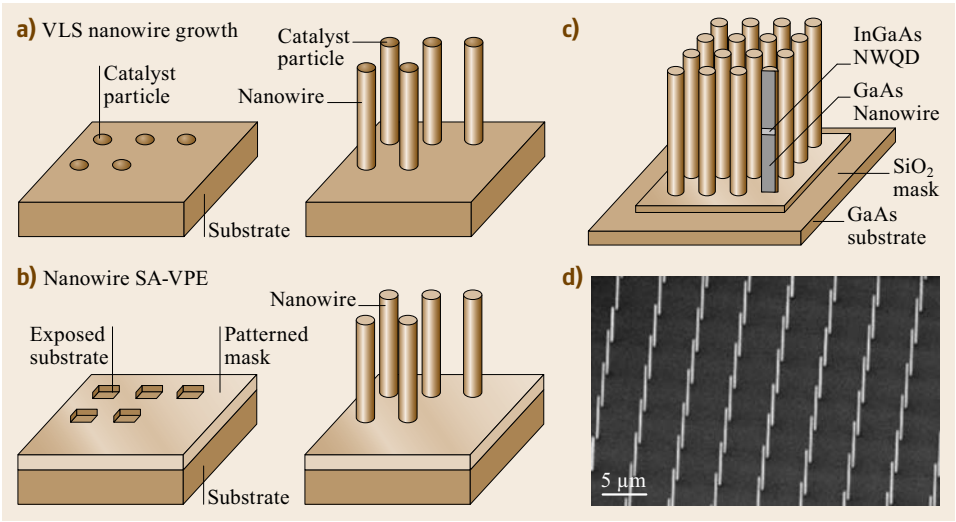


Fig. 40.16a–d Schematic illustration of nanowire formation by **(a)** vapor–liquid–solid (VLS) growth catalysed by metal nanoparticles, and **(b)** catalyst-free selective–area vapor-phase epitaxy (SA-VPE) on nanoapertures. **(c)** Schematic diagram of an array of GaAs nanowires with embedded InGaAs nanowire quantum dots (NWQDs), with SEM image in **(d)**. The wires were grown by method **(b)** using a square array of holes patterned into a SiO_2 mask on the GaAs (111)B substrate by e-beam lithography. (Parts **(a)** and **(b)** after [40.84], with permission from the Electrochemical Society. Parts **(c)** and **(d)** reprinted with permission from [40.85], © 2013 ACS)

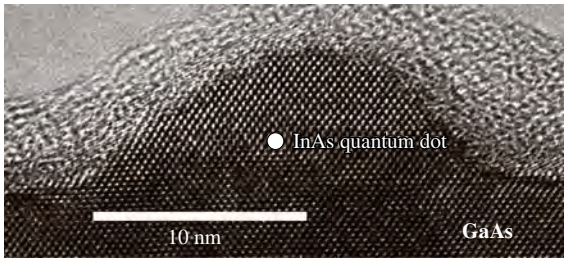


Fig. 40.17 Transmission electron microscope image of an uncapped InAs quantum dot grown on GaAs by the Stranski–Krastanow technique. (Reprinted with permission from [40.91], © 2000 APS)

Colloidal QDs can now be obtained commercially from several companies in both aqueous and organic suspensions. The ready availability of these dots makes them highly useful in a variety of applications. One strong area of application is in bio-imaging [40.92, 93]. Here, the dots are attached to a functionalized biomolecule by a chemical bond and introduced into the biosystem to be studied. The tagged biomolecule is then imaged as it moves through the system by exciting fluorescence from the dots. The high brightness and photostability of the dots compared to organic dyes make them particularly well suited for this application. Another active research area is in the development of high brightness organic LEDs [40.94]. Here, the photostability of the dots is again a key advantage compared to organic materials, while their color purity and brightness add further benefit. A third main area of research focuses on photovoltaic devices [40.95, 96]. The low cost of solution-processed material together with their high photostability makes them attractive for use in single-junction cells, with efficiencies approaching 9% reported [40.97], while the wide range of bandgaps that can be achieved by tuning the size and composition of the dots makes them appealing for use in multijunction cells.

The growth of QDs by epitaxy (i. e., MBE or MOVPE) is more expensive than the chemical synthesis of colloidal dots, but has the advantage of incorporating the dots into III–V wafers that can be processed into advanced optoelectronic devices. It has been known for many years that QDs offer significant benefits over QWs in semiconductor lasers – notably that the intrinsic gain is higher [40.98] and that the threshold current is less sensitive to temperature [40.99] – and this has prompted intense research into the development of QD lasers [40.100–102]. As mentioned in Sect. 40.3.1, the production of VCSELs at 1300 and 1550 nm has proven to be difficult for InP-based QW structures due to the low refractive index contrast of the materials that form the DBR mirrors. The use of InAs/GaAs QDs as the active region circumvents this problem and allows the benefits of mature GaAs-based VCSEL technology to be exploited [40.103]. At the same time, significant steps toward QD integra-

tion with silicon foundry process technology have been taken by demonstrating high-performance QD lasers on silicon substrates [40.104, 105]. Another very active area of research is on QD infrared photodetectors (QDIPs). In contrast to QW devices (Sect. 40.4.3), QDIPs are sensitive at normal incidence due to the relaxation of the ISB selection rules for electrons with 3-D confinement [40.106, 107]. This obviates the need for the complicated photon-coupling methods, leading to rapid progress in the deployment of QDs in infrared detectors [40.108–110]. A third intense area of research focuses on exploiting the atom-like nature of the 3-D-confined electronic states for applications in the field of quantum information processing. The emission of antibunched photons from QDs was first demonstrated in 2000 [40.111], and this has led to rapid progress in the development of single-photon sources for use in quantum cryptography and quantum computing [40.112]. At the same time, the use of excitons or carrier spin states in QDs as quantum bits for the processing of quantum information offers much promise [40.113].

40.7 Conclusions

Semiconductor QWs are excellent examples of quantum mechanics in action. The reduced dimensionality has led to major advances in both the understanding of 2-D physics and the applied science of optoelectronics. In some cases, the QWs have enhanced the performance of conventional devices (e.g., LEDs and edge-emitting lasers), and in others, they have led to radically new devices (e.g., VCSELs, quantum cascade lasers, QCSE modulators). At the present time, the main commercial use for QW optoelectronic devices is in LEDs, laser diodes, and QCSE modulators. It remains to be seen whether some of the other devices described here (QW solar cells, SL-APDs, QWIPs) will come to commercial fruition, and whether systems of lower dimensionality (e.g., QDIPs) will eventually replace QWs in the same way that QWs have replaced bulk devices.

References

40.1 L. Esaki, R. Tsu: IBM J. Res. Dev. **14**, 61 (1970)
40.2 C. Weisbuch, B. Vinter: *Quantum Semiconductor Structures* (Academic, San Diego 1991)
40.3 G. Bastard: *Wave Mechanics Applied to Semiconductor Heterostructures* (Wiley, New York 1992)
40.4 S.O. Kasap: *Optoelectronics and Photonics: Principles and Practices*, 2nd edn. (Prentice Hall, Upper Saddle River 2012)
40.5 P. Harrison: *Quantum Wells, Wires and Dots: Theoretical and Computational Physics of Semiconductor Nanostructures*, 3rd edn. (Wiley, Chichester 2009)
40.6 M.J. Kelly: *Low-Dimensional Semiconductors* (Clarendon, Oxford 1995)
40.7 P.J. Dean: III–V Compound semiconductors. In: *Electroluminescence*, ed. by J.I. Pankove (Springer, Berlin, Heidleberg 1977) p. 63
40.8 S. Nakamura, S. Pearton, G. Fasol: *The Blue Laser Diode*, 2nd edn. (Springer, Berlin, Heidelberg 2000)
40.9 S. Gasiorowicz: *Quantum Physics*, 3rd edn. (Wiley, New York 2003)
40.10 D.A.B. Miller: *Quantum Mechanics for Scientists and Engineers* (Cambridge Univ. Press, Cambridge 2008)
40.11 E.P. O'Reilly: Semicond. Sci. Technol. **4**, 121 (1989)
40.12 M. Fox: *Optical Properties of Solids*, 2nd edn. (Clarendon, Oxford 2010)
40.13 M. Shinada, S. Sugano: J. Phys. Soc. Japan **21**, 1936 (1966)
40.14 D.A.B. Miller, D.S. Chemla, D.J. Eilenberger, P.W. Smith, A.C. Gossard, W.T. Tsang: Appl. Phys. Lett. **41**, 679 (1982)
40.15 A.M. Fox, D.A.B. Miller, G. Livescu, J.E. Cunningham, W.Y. Jan: IEEE J. Quantum Electron. **27**, 2281 (1991)
40.16 H. Schneider, K. von Klitzing: Phys. Rev. B. **38**, 6160 (1988)
40.17 A.M. Fox, R.G. Ispasoiu, C.T. Foxon, J.E. Cunningham, W.Y. Jan: Appl. Phys. Lett. **63**, 2917 (1993)
40.18 J. Feldmann, K. Leo, J. Shah, D.A.B. Miller, J.E. Cunningham, T. Meier, G. von Plessen, A. Schulze, P. Thomas, S. Schmitt-Rink: Phys. Rev. B **46**, 7252 (1992)
40.19 K. Leo, P. Haring Bolivar, F. Brüuggemann, R. Schwedler, K. Köhler: Solid State Commun. **84**, 943 (1992)
40.20 C. Waschke, H.G. Roskos, R. Schwedler, K. Leo, H. Kurz, K. Köhler: Phys. Rev. Lett. **70**, 3319 (1993)
40.21 K. Leo: *High-Field Transport in Semiconductor Superlattices* (Springer, Berlin, Heidelberg 2003)
40.22 P.W.M. Blom, C. Smit, J.E.M. Haverkort, J.H. Wolter: Phys. Rev. B **47**, 2072 (1993)
40.23 J. Shah: *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*, 2nd edn. (Springer, Berlin, Heidelberg 1999)

40.24 R.G. Ispasoiu, A.M. Fox, D. Botez: IEEE J. Quantum Electron. **36**, 858 (2000)
40.25 P. Blood: Visible-emitting quantum well lasers. In: *Semiconductor Quantum Optoelectronics*, ed. by A. Miller, M. Ebrahimzadeh, D.M. Finlayson (Institute of Physics, Bristol 1999) p. 193
40.26 N. Chand, S.N.G. Chu, N.K. Dutta, J. Lopata, M. Geva, A.V. Syrbu, A.Z. Mereutza, V.P. Yakovlev: IEEE J. Quantum Electron. **30**, 424 (1994)
40.27 E.P. O'Reilly, A.R. Adams: IEEE J. Quantum Electron. **30**, 366 (1994)
40.28 A.R. Adams: IEEE J. Sel. Top. Quantum Electron. **17**, 1364 (2011)
40.29 T.-Y. Seong, J. Han, H. Amano, H. Morkoc (Eds.): *III-Nitride Based Light Emitting Diodes and Applications* (Spinger, Berlin, Heidelberg 2013)
40.30 M.T. Hardy, D.F. Feezell, S.P. DenBaars, S. Nakamura: Mater. Today **14**, 408 (2011)
40.31 S. Pearton (Ed.): *GaN and ZnO-Based Materials and Devices* (Springer, Berlin, Heidelberg 2012)
40.32 S. Nakamura, M.R. Krames: Proc. IEEE **101**, 2211 (2013)
40.33 E.F. Schubert, J.K. Kim, H. Luo, J.-Q. Xi: Rep. Prog. Phys. **69**, 3069 (2006)
40.34 J. Piprek: Physica Stat. Sol. A **207**, 2217 (2010)
40.35 D.F. Feezell, J.S. Speck, S.P. DenBaars, S. Nakamura: J. Display Technol. **9**, 190 (2013)
40.36 H.E. Li, K. Iga (Eds.): *Vertical-Cavity Surface-Emitting Laser Devices* (Springer, Berlin, Heidelberg 2003)
40.37 R. Michalzik (Ed.): *VCSELs* (Springer, Berlin, Heidelberg 2013)
40.38 K. Iga: Jap. J. Appl. Phys. **47**, 1 (2008)
40.39 J.S. Harris, T. O'Sullivan, T. Sarmiento, M.M. Lee, S. Vo: Semicond. Sci. Technol. **26**, 014010 (2011)
40.40 S.J. Sweeney, S.R. Jin: J. Appl. Phys. **113**, 043110 (2013)
40.41 E.F. Schubert, Y.-H. Wang, A.Y. Cho, L.-W. Tu, G.J. Zydzik: Appl. Phys. Lett. **60**, 921 (1992)
40.42 N.E. Hunt, E.F. Schubert, R.F. Kopf, D.L. Sivco, A.Y. Cho, G.J. Zydzik: Appl. Phys. Lett. **63**, 2600 (1993)
40.43 D. Delbeke, R. Bockstaele, P. Bienstman, R. Baets, H. Benisty: IEEE J. Sel. Top. Quantum Electron. **8**, 189 (2002)
40.44 J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho: Science **264**, 553 (1994)
40.45 J. Faist: *Quantum Cascade Lasers* (Oxford Univ. Press, Oxford 2013)
40.46 C. Gmachl, F. Capasso, D.L. Sivco, A.Y. Cho: Rep. Prog. Phys. **64**, 1533 (2001)
40.47 R.F. Kazarinov, R.A. Suris: Sov. Phys. Semicond. **5**, 707 (1971)
40.48 D.G. Revin, J.P. Commin, S.Y. Zhang, A.B. Krysa, K. Kennedy, J.W. Cockburn: IEEE J. Select. Top. Quantum Electron. **17**, 1417 (2011)
40.49 B.S. Williams: Nat. Photon. **1**, 517 (2007)
40.50 C. Sirtori, S. Barbieri, R. Colombelli: Nat. Photon. **7**, 691 (2013)
40.51 M.A. Green, K. Emery, Y. Hishikawa, W. Warta, E.D. Dunlop: Prog. Photovolt: Res. Appl. **20**, 12 (2012)
40.52 K. Barnham, I. Ballard, J. Barnes, J. Connolly, P. Griffin, B. Kluftinger, J. Nelson, E. Tsui, A. Zachariou: App. Surface Sci. **113/114**, 722 (1997)
40.53 N.J. Ekins-Daukes, K.-H. Lee, L. Hirst, A. Chan, M. Führer, J. Adams, B. Browne, K.W.J. Barnham, P. Stavrinou, J. Connolly, J.S. Roberts, B. Stevens, R. Airey, K. Kennedy: J. Phys. D: Appl. Phys. **46**, 264007 (2013)
40.54 J. Nelson, N. Ekins-Daukes: Quantum well solar cells. In: *Clean Electricity from Photovoltaics*, 2nd edn., ed. by M.D. Archer, M. Green (Imperial College, London 2014) p. 453
40.55 M.A. Itzler, R. Ben-Michael, C.-F. Hsu, K. Slomkowski, A. Tosi, S. Cova, F. Zappa, R. Ispasoiu: J. Mod. Opt. **54**, 283 (2007)
40.56 S. Burri, Y. Maruyama, X. Michalet, F. Regazzoni, C. Bruschini, E. Charbon: Opt. Express **22**, 17573 (2014)
40.57 P.-H. Sun, S.-T. Chang, Y.-C. Chen, H. Lin: Solid-State Electron. **54**, 1216 (2010)
40.58 J.P.R. David, C.H. Tan: IEEE J. Sel. Topics Quant. Electron. **14**, 998 (2008)
40.59 J.S. Vickers: US Patent US6720588B2 (2004)
40.60 G. Karve, S. Wang, F. Ma, X. Li, J.C. Campbell, R.G. Ispasoiu, D.S. Bethune, W.P. Risk, G.S. Kinsey, J.C. Boisvert, T.D. Isshiki, R. Sudarsanan: Appl. Phys. Lett. **86**, 063505 (2005)
40.61 A. Suzuki, A. Yamada, T. Yokotsuka, K. Idota, Y. Ohiki: Jpn. J. Appl. Phys. **41**, 1182 (2002)
40.62 F. Capasso, W.T. Tsang, A.L. Hutchinson, G.F. Williams: Appl. Phys. Lett. **40**, 38 (1982)
40.63 G. Ripamonti, F. Capasso, A.L. Hutchinson, D.J. Muehlner, J.F. Walker, R.J. Malek: Nucl. Instrum. Meth. Phys. Res. A **288**, 99 (1990)
40.64 R. Chin, N. Holonyak, G.E. Stillman, J.Y. Tang, K. Hess: Electron. Lett. **16**, 467 (1980)
40.65 C.K. Chia, J.P.R. David, G.J. Rees, P.N. Robson, S.A. Plimmer, R. Grey: Appl. Phys. Lett. **71**, 3877 (1997)
40.66 F. Ma, X. Li, S. Wang, K.A. Anselm, X.G. Zheng, A.L. Holmes, J.C. Campbell: J. Appl. Phys. **92**, 4791 (2002)
40.67 P. Yuan, S. Wang, X. Sun, X.G. Zheng, A.L. Holmes, J.C. Campbell: IEEE Photon. Technol. Lett. **12**, 1370 (2000)
40.68 M.A. Saleh, M.M. Hayat, P.P. Sotirelis, A.L. Holmes, J.C. Campbell, B.E.A. Saleh, M.C. Teich: IEEE Trans. Elect. Dev. **48**, 2722 (2001)
40.69 L.C. West, S.J. Eglash: Appl. Phys. Lett. **46**, 1156 (1985)
40.70 B.F. Levine: J. Appl. Phys. **74**, R1 (1993)
40.71 A. Rogalski, J. Antoszewski, L. Faraone: J. Appl. Phys. **105**, 091101 (2009)
40.72 S.D. Gunapala, D.Z. Ting, C.J. Hill, J. Nguyen, A. Soibel, S.B. Rafol, S.A. Keo, J.M. Mumolo, M.C. Lee, J.K. Liu, B. Yang, A. Liao: IR Phys. Technol. **54**, 155 (2011)
40.73 M. Razeghi, B.-M. Nguyen: Rep. Prog. Phys. **77**, 082401 (2014)
40.74 E.A. Plis: Adv. Electron. **2014**, 246769 (2014)
40.75 B.F. Levine, K.K. Choi, C.G. Bethea, J. Walker, R.J. Malik: Appl. Phys. Lett. **51**, 934 (1987)

40.76 H. Schneider: Appl. Phys. Lett. **82**, 4376 (2003)
40.77 J.S. Weiner, D.A.B. Miller, D.S. Chemla: Appl. Phys. Lett. **50**, 842 (1987)
40.78 Y. Miyazaki, H. Tada, T. Aoyagi, T. Nishimura, Y. Mitsui: IEEE J. Quant. Electron. **38**, 1075 (2002)
40.79 G. Agrawal: *Fiber Optic Communication Systems*, 4th edn. (Wiley, New York 2010)
40.80 Y. Tang, J.D. Peters, J.E. Bowers: Optics Express **20**, 11529 (2012)
40.81 Y.-H. Kuo, Y.K. Lee, Y. Ge, S. Ren, J.E. Roth, T.I. Kamins, D.A.B. Miller, J.S. Harris: Nature **437**, 1334 (2005)
40.82 Y.-H. Kuo, Y.K. Lee, Y. Ge, S. Ren, J.E. Roth, T.I. Kamins, D.A.B. Miller, J.S. Harris: IEEE J. Sel. Top. Quantum Electron. **12**, 1503 (2006)
40.83 P. Chaisakul, D. Marris-Morini, M.-S. Rouifed, G. Isella, D. Chrastina, J. Frigerio, X. Le Roux, S. Edmond, J.-R. Coudevylle, L. Vivien: Optics Express **20**, 3219 (2012)
40.84 M. Cantoro, G. Brammertz, O. Richard, H. Bender, F. Clemente, M. Leys, S. Degroote, M. Caymax, M. Heyns, S. De Gendt: J. Electrochem. Soc. **156**, H860 (2009)
40.85 M.N. Makhonin, A.P. Foster, A.B. Krysa, P.W. Fry, D.G. Davies, T. Grange, T. Walther, M.S. Skolnick, L.R. Wilson: Nano Lett. **13**, 861 (2013)
40.86 E. Kapon, D.M. Hwang, R. Bhat: Phys. Rev. Lett. **63**, 430 (1989)
40.87 R.R. LaPierre, A.C.E. Chia, S.J. Gibson, C.M. Haapamaki, J. Boulanger, R. Yee, P. Kuyanov, J. Zhang, N. Tajik, N. Jewell, K.M.A. Rahman: Phys. Stat. Sol. RRL **7**, 815 (2013)
40.88 V. Mourik, K. Zuo, S.M. Frolov, S.R. Plissard, E.P.A.M. Bakkers, L.P. Kouwenhoven: Science **336**, 1003 (2012)
40.89 H.J. Joyce, Q. Gao, H.H. Tan, C. Jagadish, Y. Kim, J. Zou, L.M. Smith, H.E. Jackson, J.M. Yarrison-Rice, P. Parkinson, M.B. Johnston: Prog. Quantum Electron. **35**, 23 (2011)
40.90 N.P. Dasgupta, J. Sun, C. Liu, S. Brittman, S.C. Andrews, J. Lim, H. Gao, R. Yan, P. Yang: Adv. Mat. **26**, 2137 (2014)
40.91 P.W. Fry, I.E. Itskevich, D.J. Mowbray, M.S. Skolnick, J.J. Finley, J.A. Barker, E.P. O'Reilly, L.R. Wilson, I.A. Larkin, P.A. Maksym, M. Hopkinson, M. Al-Khafaji, J.P.R. David, A.G. Cullis, G. Hill, J.C. Clark: Phys. Rev. Lett. **84**, 733 (2000)
40.92 B.A. Kairdolf, A.M. Smith, H.T. Stokes, M.D. Wang, A.N. Young, S. Nie: Annu. Rev. Anal. Chem. **6**, 143 (2013)
40.93 E. Petryayeva, W.R. Algar, I.L. Medintz: Appl. Spectroscopy **67**, 215 (2013)
40.94 Y. Shirasaki, G.J. Supran, M.G. Bawendi, V. Bulović: Nat. Photon. **7**, 13 (2013)
40.95 P.V. Kamat: J. Phys. Chem. C **112**, 18737 (2008)
40.96 G. Konstantatos, E.H. Sargent (Eds.): *Colloidal Quantum Dot Optoelectronics and Photovoltaics* (Cambridge Univ. Press, Cambridge 2013)
40.97 C.-H.M. Chuang, P.R. Brown, V. Bulović, M.G. Bawendi: Nat. Mat. **13**, 796 (2014)
40.98 M. Asada, Y. Miyamoto, Y. Suematsu: IEEE J. Quantum Electron. **22**, 1915 (1986)
40.99 Y. Arakawa, H. Sakaki: Appl. Phys. Lett. **40**, 939 (1982)
40.100 P. Bhattacharya, Z. Mi, J. Yang, D. Basu, D. Saha: J. Cryst. Growth **311**, 1625 (2009)
40.101 J.J. Coleman, J.D. Young, A. Garg: J. Lightwave Technol. **29**, 499 (2011)
40.102 N.N. Ledentsov: Semicond. Sci. Technol. **26**, 014001 (2011)
40.103 H.C. Yu, J.S. Wang, Y.K. Su, S.J. Chang, F.I. Lai, Y.H. Chang, H.C. Kuo, C.P. Sung, H.P.D. Yang, K.F. Lin, J.M. Wang, J.Y. Chi, R.S. Hsiao, S. Mikhrin: IEEE Photon. Technol. Lett. **18**, 418 (2006)
40.104 D.A. Lee, Q. Jiang, M. Tang, Y. Zhang, J.A. Seeds, H. Liu: IEEE J. Select. Top. Quantum Electron. **19**, 1901107 (2013)
40.105 A.Y. Liu, C. Zhang, J. Norman, A. Snyder, D. Lubyshev, J.M. Fastenau, A.W.K. Liu, A.C. Gossard, J.E. Bowers: Appl. Phys. Lett. **104**, 041104 (2014)
40.106 H.C. Liu, M. Gao, J. McCaffrey, Z.R. Wasilewski, S. Fafard: Appl. Phys. Lett. **78**, 79 (2001)
40.107 A.D. Stiff, S. Krishna, P. Bhattacharya, S.W. Kennerly: IEEE J. Quantum Electron. **37**, 1412 (2001)
40.108 S. Krishna, S.D. Gunapala, S.V. Bandara, C. Hill, D.Z. Ting: Poc. IEEE **95**, 1838 (2007)
40.109 P. Martyniuk, A. Rogalski: Prog. Quantum Electron. **32**, 89 (2008)
40.110 A.V. Barve, J. Montaya, Y. Sharma, T. Rotter, J. Shao, W.-Y. Jang, S. Meesala, S.J. Lee, S. Krishna: IR Phys. Technol. **54**, 215 (2011)
40.111 P. Michler, A. Kiraz, C. Becher, W.V. Schoenfeld, P.M. Petroff, L. Zhang, E. Hu, A. Imamoglu: Science **290**, 2282 (2000)
40.112 S. Buckley, K. Rivoire, J. Vučković: Rep. Prog. Phys. **75**, 126503 (2012)
40.113 F. Henneberger, O. Benson (Eds.): *Semiconductor Quantum Bits* (Pan Stanford, Singapore 2008)

41. Glasses for Photonic Integration

Ray DeCorby, Mehrdad Irannejad

Inorganic glasses form the backbone of dielectric materials in optics and photonic applications. In addition to offering a range of transparency windows, intrinsic physical properties of glass materials also offer flexibility of processing for the realization of fibers, films, and shaped optical elements. Traditionally, the main role of glass has been as an optically passive material. However, a significant attribute of glass as an optical medium is to host dopants such as nanoparticles or active ions in a unique chemical environment, which can be engineered via processing, by controlling the geometry to develop active and passive photonic devices, such as laser, amplification, storage media, optical switching, frequency conversion, and sensor media.

For photonic integration, many of the attributes of glasses are particularly compelling. Glasses allow numerous options for thin film deposition and integration on arbitrary platforms. The possibility of controlling the thermal and flow properties, e.g., expansion coefficient, crystallization, and viscosity during film deposition of a glass may allow realization of extremely low-loss microphotonic waveguides, photonic crystals, and microcavities. The metastable nature of glass can enable the direct patterning of photonic elements by energetic beams.

This chapter provides an overview of these unique properties of glasses, from the perspectives of the technology options they afford and the practical limitations they present. Further, an overview is provided of the main families of glassy inorganic films studied for integrated optics. Finally, the main features of rare-earth doped glasses are reviewed, with an emphasis on their potential for implementation in compact integrated light sources and amplifiers.

41.1 **Main Attributes of Glasses as Photonic Materials** 1061
41.1.1 The Glass Transition and Control of Viscosity 1062
41.1.2 Metastability 1064
41.1.3 Glass as Host Material 1068

41.2 **Glasses for Integrated Optics** 1068
41.2.1 Low-Index Glassy Films 1068
41.2.2 Medium-Index Glassy Films 1069
41.2.3 High-Index Glassy Films 1070

41.3 **Laser Glasses for Integrated Light Sources** 1071
41.3.1 Advantages of Glass-Based Light Sources 1071
41.3.2 Alternative Glass Hosts 1072
41.3.3 Progress Towards Integrated Light Sources in Glass 1075
41.3.4 Integration of Dissimilar Materials Using Laser Processing 1075

41.4 **Summary** 1077

References 1077

Inorganic glasses have played a central role in optical science and technology, and more generally within the electrical and electronic engineering disciplines [41.1]. Amongst optical materials, the unique advantages of glasses are well known [41.2, 3]. Glasses can be worked relatively easily into various forms, such as bulk lenses, fibers, and thin films. A remarkable example of geometrical control is shown by the manufacture of low-loss silica fibers, in which a hair-thin glass fiber of $\approx 125\,\mu m$ (with tightly controlled geometrical and material properties) is drawn from a heated glass preform at rates as high as 20 m/s or greater [41.4]. Glass can be manufactured with excellent homogeneity and without grain boundaries, so that scattering of light is acceptably low. While perfect crystals typically exhibit even lower levels of scattering, they are difficult to realize in large sizes and with arbitrary shapes. Conversely, optical losses in polycrystalline materials are generally

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_41

excessive due to scattering from grain boundaries and other crystal defects. Finally, the standard glass compositions (oxide, halide, and chalcogenide glasses and their hybrids) provide transparency windows from the UV to the mid-infrared (Fig. 41.1). It has been estimated [41.2] that more than 90% of optical components are based on glasses. Given the massive worldwide installation of optical communications fibers since the 1980s [41.5], and depending on the definition assigned to the term *dielectric component*, this statistic might actually be closer to 100%.

Befitting their central role in optical technology, there are numerous excellent reviews covering the optical properties, processing, and photonic functionalities of a range of inorganic glasses [41.3, 6–10] which target bulk glass and fiber optic applications. Further, one of the distinguishing attributes of glasses is their compositional flexibility. Hundreds of commercial glasses are available, and many more have been studied in research laboratories and reported in the literature. Thus, it is not possible to capture the full range of glassy materials (even restricted to inorganic glasses) within a short review paper. A recent monograph by *Yamane* and *Asahara* [41.6] provides a comprehensive and highly recommended treatment of glass technology and its application in photonics. However, reviews on the comparative properties of inorganic glasses for integrated photonics are few and far between. This is an emerging area and for this purpose in this short review our modest aims are as follows:

1. With a focus specifically on *integrated* photonics, the main advantages of glasses are discussed and the most-studied glass systems are reviewed in brief. Integrated photonics (integrated optics) refers to guided wave photonic devices fabricated on planar platforms. Glass substrates (with waveguides defined by ion diffusion) have played an important role in integrated optics. However, the advantages of integrated glass devices compared to crystalline materials such as III–V semiconductors or lithium niobate have been less apparent in applications so far. For this reason, emphasis is placed on glassy thin films that can enable photonic functionality on other substrates. In practice, substrates of interest might be printed circuit boards or semiconductor wafers.
2. For the most part, the discussion is restricted to inorganic glasses. Some reference is made to organic glasses, since their properties and advantages may be comparable with inorganic glasses and polymeric dielectric materials. Good reviews covering polymer materials for integrated photonics are available [41.11].

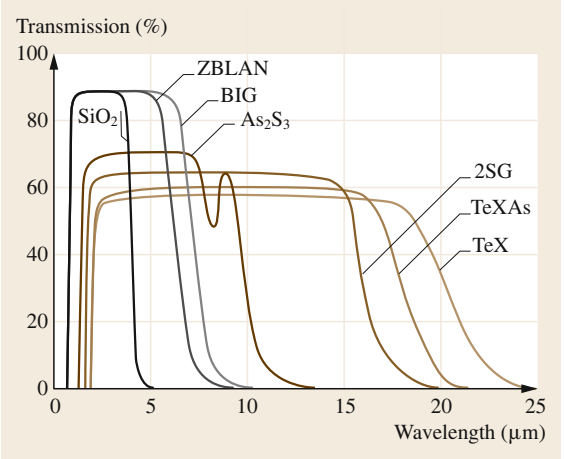


Fig. 41.1 Transmission plots are shown for SiO_2 glass and some representative fluoride and chalcogenide glasses. ZBLAN and BIG are heavy-metal fluoride glasses, 2SG is a selenide glass, and TeXAs and TeX are tellurium-based chalcogenide glasses. (After [41.15])

3. To date, the most important inorganic glass system for integrated photonics is the silica-on-silicon planar lightwave circuit (PLC) platform. We only mention this briefly as the literature [41.12, 13] provides an overview of that technology. Rather, the goal is to provide some insight into the advantages that are globally associated with glassy materials. Further, an emphasis is placed on the unique advantages (lower processing temperatures, higher refractive index contrast, optically active functionality, etc.) enabled by some glass systems.
4. In the context of emerging trends, the review also reports on the application of laser-based deposition and femtosecond laser techniques for waveguide inscription in hybrid and monolithic integration of active-passive functions [41.14]. We also report materials based methodological approaches for integrating photonic crystals (PCs) with high-speed data networks where there is much need for integration on the backplane of printed circuit boards (PCBs). Since the top layer materials present in PCBs for electronic circuits are polymer based, such a choice of material limits the capability of offering active-passive device integration which is essential for high-speed (10 Gb/s) data communication (datacom). Materials issues for active-passive integration are also reviewed.

With these goals in mind, the chapter is organized as follows. In Sect. 41.1, we highlight the main advantages of glasses for photonic integration. These attributes include the unique processing options afforded by the

existence of the glass transition and forming properties in the metastable regime of thermal processing of glasses. Sect. 41.2 provides an overview of the most-studied materials for glass-based integrated optics. This includes the standard glass formers (oxide, halide, and chalcogenide glasses) as well as several materials that are generally classified as ceramics (Si_3N_4, Al_2O_3, etc.) but which can also be formed as amorphous thin films for integrated photonics. In Sect. 41.3, some unique advantages of glasses for realization of integrated light sources are reviewed. Sect. 41.4 provides a summary of emerging trends and some thoughts on future prospects for glass as a material for photonic integration.

41.1 Main Attributes of Glasses as Photonic Materials

One of the most generic features of a glassy material, in contrast with crystalline solids and liquids, is the existence of a characteristic glass transition temperature (T_g). In short, if a viscous liquid is cooled fast enough, the liquid passes through the freezing temperature of the material without undergoing a phase transition to equilibrium or metastable equilibrium solid crystalline states (Fig. 41.2). Rather, it becomes a supercooled liquid, whose rate of change of specific volume (or enthalpy) with decreasing temperature remains constant and equivalent to that of the molten liquid. At T_g, the slope of these curves (for example, specific volume versus temperature) changes abruptly (but in a continuous way) to a value that slowly relaxes with time towards the nearest equilibrium state. Below T_g, the quenched material is an amourph solid or glass.

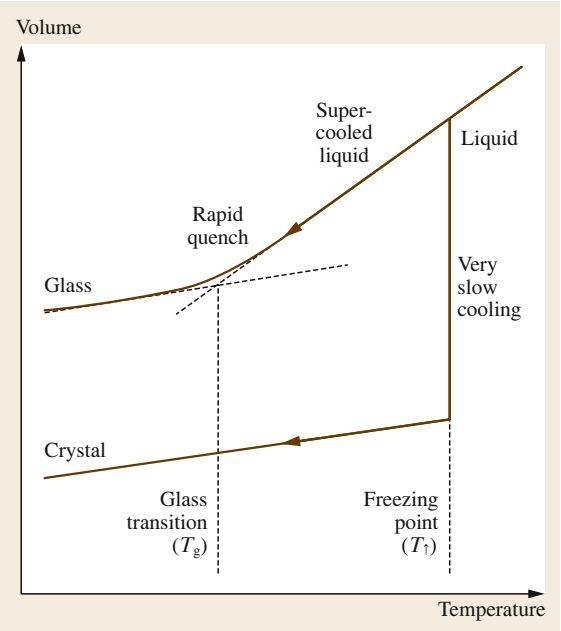


Fig. 41.2 Schematic plot illustrating the change in specific volume of a glass with temperature. (After [41.16])

While glass forming has been a staple of human technological innovation for thousands of years, the physical processes underlying the glass transition are still a matter of some debate [41.16–18]. Several empirical facts are well known, however. The glass transition temperature of a glass is dependent on thermal history, e.g., quench or glass-formation rate (which is the rate of cooling from the melt, or for film formation) and even the age of the glass [41.19]. The quench rate dependence of T_g is given by an Arrhenius-type equation [41.20]

$$Q_i = Q_o \exp\left[-B\left(\frac{1}{T_g} - \frac{1}{T_m}\right)\right], \qquad (41.1)$$

where Q_o and B are constants, and have different values for different glasses. T_m is the melting point.

However, on the basis of empirical observations T_g is in general estimated using the formula: $T_g \approx 2T_m/3$ [41.17]. This rough estimation, however, works only with silicates, phosphates, and borate glasses. For a vast majority of inorganic glasses which are not strongly network forming, the 2/3rd rule does not apply. For glasses that cannot be formed easily by melt quenching, such as amorphous Si [41.18], the direct experimental determination of T_g is not always possible, and indirect methods are sometimes used to estimate an effective T_g.

Central to the technological application of glasses is the ability to control their viscosity by way of temperature [41.5]. For stable glass formers, it is possible to transition between states of very low viscosity (molten liquid), low viscosity (supercooled liquid), and high viscosity (frozen liquid, i. e., glass). This remarkable variation in mechanical properties with temperature, spanning several orders of magnitude in viscosity, underpins both the ancient and modern manufacture of glass products (by molding, extrusion, float processes, etc.). It is interesting to note that T_g can be defined as the approximate temperature at which the shear viscosity attains 10^{12} Pa s. Further, viscosity varies rapidly in the

vicinity of T_g. For so-called strong glass formers (such as SiO_2), the viscosity η varies according to an Arrhenius equation of the form $\eta = A\exp(B/k_BT)$, where A and B are temperature-independent constants and k_B is the Boltzmann constant [41.17]. To form glasses into desired shapes, the temperature is typically controlled such that the viscosity is in the 10^2–10^7 Pa s range, depending on the process and the stability of the particular glass composition [41.6, 21].

It was once held that only certain materials could form glasses, but it is now believed that the glassy state is attainable for all types of materials [41.16]. Accordingly, the literature contains references to various families of glasses, including inorganic (ceramic), organic (polymeric), and metallic glasses.

41.1.1 The Glass Transition and Control of Viscosity

The surface tension of liquids is a well-known enabler of optical quality surfaces. The interfacial or surface tension property of a glass-forming liquid together with viscosity and quench rate determine the optical surface quality [41.22].

The manufacture of modern single-mode fibers is a good example of the enabling nature of the glass transition. In a typical process [41.4, 6, 23], highly purified glass layers are deposited either on the inside walls of a fused silica tube by chemical vapor deposition (CVD) or on the outside surface of a ceramic rod by flame hydrolysis deposition (FHD) or flame oxidation deposition (FOD). The hollow glass cylinder (after removal of the ceramic rod in the latter case) is then heated (sintered) to $\approx 1600\,^{\circ}C$, such that it collapses to form a solid glass preform (typically 1 m in length and 2 cm in diameter). 1600 °C is the approximate softening point of silica glass, typically defined as corresponding to a viscosity of $\approx 10^7$ Pa s [41.6]. The fiber drawing process involves local heating of the preform to a temperature of approximately 2000 °C (corresponding to a viscosity of $\approx 10^5$ Pa s). Thus, viscosity control is critical at nearly all stages of the process. Further, the drawing process allows tremendous control over the optical and geometrical properties of the fiber, in part due to the self-regulating properties of surface tension effects in the supercooled liquid. As an added benefit, inhomogeneities in the original preform tend to be averaged out in the process [41.24]. The amazing properties of silica fibers (tensile strength comparable to that of steel, low loss, and extremely tight tolerances for material and geometrical parameters) are well known, and attest to the technological versatility of (silica) glasses. It is also interesting to note that the standard technique for splicing sections of silica fibers together, which is achieved with less than 0.1 dB loss [41.5], is based on a reflow process wherein surface tension contributes to their alignment.

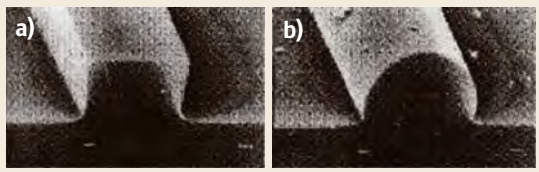


Fig. 41.3 (a) A glass waveguide core is shown before and (b) after a selective reflow process. (After [41.25])

Within integrated optics, the manufacture and optimization of PLC devices relies directly on the glass transition in many cases [41.12, 13]. PLC devices are manufactured in doped SiO_2 glasses deposited (by CVD, FHD, or sol–gel techniques) on a silicon wafer. The undercladding is often pure SiO_2, which can be formed by thermal oxidation of the silicon wafer. Typically, dopants in the core and uppercladding layers act to modify both the refractive index and the thermomechanical properties of the glass. In particular, controlled reflow (by heating the glasses near T_g) of these layers often plays an important part in the process. Reflow is used to consolidate the interface between layers, to smoothen the rough sidewalls of the waveguide core (formed by dry etching processes), and to assist in gap filling and planarization of the cladding layer over the patterned core layer [41.12, 24–26]. Typically, the layers are doped such that the glass transition decreases in the upward direction. In that way, a layer can be reflowed without affecting the underlying layers. As an example, a *pinned based reflow* is shown in Fig. 41.3 [41.25]. The circle-section shaped core arises from selective reflow of the core material while its interface with the undercladding glass remains fixed. The upper boundary of the core is shaped by surface tension.

Surface tension driven manufacturing of more complex designs of optical components such as silica-glass based optical microcavities [41.27–29], nm-scale silica wires [41.30], photonic crystal fibers [41.31], microball lenses [41.32], and microlenses [41.33, 34] have been widely studied. In these cases, the reduction of scattering losses by surface tension-mediated smoothing of interfaces is particularly important.

As discussed further in Sect. 41.2, high-index contrast waveguides, photonic crystals, and microcavities are major themes within integrated optics. High index contrast (between the core and cladding of a waveguide, for example) is desirable because it enables strong confinement of light to small waveguides or small res-

onant cavities. This is essential for the realization of high-density integrated optics [41.12]. The downside of high-index contrast is that scattering at interfaces increases with index contrast. This is clarified by the so-called Tien model for waveguide loss due to interfacial scattering in a symmetric slab waveguide [41.35]

$$\alpha_S = \frac{2\sigma^2 k_0^2 h}{\beta} \Delta n^2 E_S^2 \,, \tag{41.2}$$

where σ is the standard deviation (characteristic amplitude) of the roughness, k_0 is the free-space wavenumber, h is the inverse of the penetration depth of the mode into the cladding, β is the propagation constant of the mode, and E_S is the normalized electric field amplitude at the core-cladding interface. Further, $\Delta n^2 = n_{core}^2 - n_{clad}^2$, where n_{core} and n_{clad} are the core and cladding refractive indices, respectively. This equation explicitly shows the strong dependence of scattering loss on index contrast. A more accurate expression, derived by considering the statistics of roughness (not just the amplitude), is the Payne–Lacey model [41.36]

$$\alpha_S = 4.34 \frac{\sigma^2}{k_0 \sqrt{2} d^4 n_{core}} g f \,, \tag{41.3}$$

where d is the core thickness, g is a function that depends mainly on the core-cladding index offset, and f is a function that depends on the statistics of the roughness. If $r(z)$ is the random function describing the line-edge roughness (i. e., the deviation of the core-cladding interface from a straight line), the autocorrelation of $r(z)$ is often well described by an exponential function [41.37]

$$R(u) = \langle r(z) r(z+u) \rangle \approx \sigma^2 \exp\left(\frac{-|u|}{L_c}\right) , \tag{41.4}$$

where L_c is the so-called correlation length. It can be shown that for waveguide geometries of typical interest, low scattering loss requires both σ and L_c to be low. As a second illustration of the impact of interface roughness, it can be shown that the surface scattering limited quality (Q) factor for whispering gallery modes in a spherical microsphere is approximately [41.29]

$$Q_{SS} = \frac{\lambda^2 D}{2\pi^2 \sigma^2 L_c} \,, \tag{41.5}$$

where D is the microsphere diameter. Again, minimization of scattering losses is correlated with minimization of σ and L_c.

It is interesting to compare recent results for high-index-contrast structures in crystalline materials and glasses. Much of the work on very highly confining waveguides and photonic crystals has focused on semiconductor material systems, especially the silicon-on-insulator (SOI) system [41.35, 36] and III–V semiconductor systems [41.38]. Generally, structures are defined by electron-beam lithography or photolithography, followed by a dry etching process. Each of these steps has limited precision and contributes to the overall roughness [41.37]. With use of sophisticated processes, roughness parameters are typically on the order of $\sigma \approx 5\,\text{nm}$ and $L_c \approx 50\,\text{nm}$, resulting in losses on the order of $3-10\,\text{dB/cm}$ for submicron strip waveguides [41.36]. Significant improvement in these numbers is mostly reliant on advancements in lithography or etching processes. These same issues limit the performance of other high-index contrast structures formed in semiconductors. For example, the Q factor of semiconductor microdisk and microring resonators typically does not exceed 10^5 [41.39], and the loss of photonic crystal defect waveguides in semiconductors remains impractically large [41.40].

By comparison, glasses (at least certain glasses) offer the potential for much smaller characteristic roughness. If a glass is formed from a supercooled liquid state, surface tension effects can produce glass surfaces with nearly atomic-level smoothness. The roughness of a melt-formed or reflowed glass surface is determined by surface capillary waves, which are small amplitude fluctuations at a liquid surface that become frozen in place at T_g [41.31]. The resulting root mean square (RMS) roughness is approximately given by [41.41]

$$\sigma \approx \sqrt{\frac{k_B T_g}{\gamma(T_g)}} \,, \tag{41.6}$$

where $\gamma(T_g)$ is the surface tension of the supercooled liquid at T_g. For silica glass, this equation predicts RMS roughness on the order of 0.1 nm or less, and the correlation length of this inherent roughness is estimated to be as small as 3 nm [41.29]. Note that glasses with lower T_g and/or higher surface tension can in theory exhibit even lower values of surface roughness. Interestingly, sub-nm surface roughness has been experimentally verified for both melt-formed and fractured surfaces of silica glass [41.41] and indirectly from loss measurements in photonic crystal fibers [41.31]. Sub-nm roughness has also been determined for wet-etched surfaces of SiO_2 glass [41.42] and is not untypical for glass films deposited by evaporation.

The smoothness of glass surfaces has important implications for realization of low-loss microphotonic devices, as evidenced by several recently reported results. While little work has been conducted on small core, high-index-contrast planar waveguides in glass,

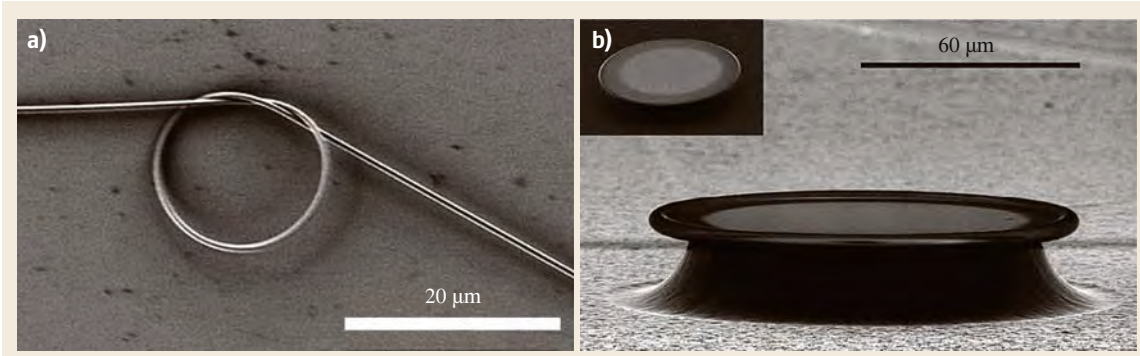


Fig. 41.4 **(a)** A submicron silica glass wire with sub-nm surface roughness is shown. (After [41.30]). **(b)** A silica glass toroid sitting atop a silicon post is shown. The atomic-level smoothness of the toroid surface was achieved using a selective reflow process. (After [41.43])

a method for fabricating submicron silica wires was recently reported [41.30] (Fig. 41.4). The authors estimated the RMS surface roughness of their wires to be less than 0.5 nm, and demonstrated losses on the order of 1 dB/cm at 633 nm for 400 nm diameter wires. Sub-nm surface roughness has also been measured in the case of sol–gel glass microlenses formed by a reflow process [41.33]. Silica-glass-based microsphere cavities, typically formed by melting the end of fused silica fiber, exhibit Q factors of nearly 10^{10} [41.28, 29], the highest for any solid state microcavity. Recently, this concept has been extended [41.27, 43] to the manufacture of silica toroid microresonators on a silicon substrate (Fig. 41.4). The latter devices exhibit Q factors exceeding 10^8. Finally, glass-based photonic crystal fibers have been demonstrated to exhibit scattering loss near fundamental limits set by surface capillary waves [41.31]. This last result contrasts with the situation mentioned above for semiconductor photonic crystals, which to date are significantly compromised by optical scattering.

In summary, processing enabled by the glass transition clearly offers unique benefits for manufacture of microphotonic structures. It is necessary to add a caveat to this discussion, however. Many amorphous materials are conditional glass formers, and do not offer the full range of processing advantages outlined above. As an extreme example, amorphous silicon films typically crystallize at temperatures well below their effective glass transition temperature [41.18], and are not amenable to reflow processes. To varying degrees, this is true of many *glassy* films used in integrated optics, such as silicon oxynitride and alumina. Further, the tendency towards crystallization varies greatly even amongst the traditional glass formers, as discussed below. Thus, chalcogenide and fluoride glasses will not generally support the range of viscosity control that enabled many of the SiO_2-based devices mentioned. Having said that, novel techniques can sometimes circumvent crystallization problems. For example, pressure can be used as a tool for forming glass devices at temperatures below the onset of severe crystallization (but high enough to enable the glass to flow). Chalcogenide glass-based lenses molded by the application of pressure and temperature are now commercially available [41.44, 45]. Another interesting example is the recent work on extruded channel waveguides reported by *Mairaj* et al. [41.46]. Finally, much research has been directed at the fabrication of microphotonic elements in organic glasses by hot embossing [41.11]. In this technique, a hard master (such as a silicon wafer) is pre-patterned with the negative image of the desired photonic circuitry. The glass is heated above T_g and the master is pressed against the glass such that the image of its pattern is transferred to the softened glass. While studied mainly in polymer, this technique might also be applicable to low-cost fabrication of photonic structures in *soft* inorganic glasses.

41.1.2 Metastability

A glass is a metastable material, having been frozen as an amorphous solid possessing excess internal energy relative to the corresponding crystal. Metastability is a double-edged sword. On the one hand, it presents technological limitations with respect to the processing and use of a glass device. As an example, it is well known that commercial introduction of organic glass (polymer)-based optical devices has been hindered by poor stability of many polymers at elevated temperatures or under exposure to high intensity light. On the other hand, metastability offers unique options for fabrication and optimization of photonic microstructures. Specifically, glass properties (optical, mechani-

cal, chemical) can often be tailored or adjusted by the careful addition of energy, in the form of heat, light, electron beams, or ion beams. An example is provided by the commercially important fiber Bragg gratings (FBG), wherein UV light is used to induce a stable refractive index change in the Ge-doped core of a standard SiO_2-based fiber. In the following subsections, the implications of metastability for glass-based integrated photonics are discussed.

Devitrification

Well below its glass transition temperature, a glass sits in a local energy minimum and is impeded by its own viscosity from reaching the lower energy crystalline state. A frequently cited example is that of silicate glass windows in ancient buildings, which have not exhibited any significant crystallization. However, addition of sufficient energy (such as heating a glass to some temperature above T_g) can result in crystallization of a glass. Controlled crystallization is a technologically useful means of modifying the mechanical, thermal, or optical properties of a glass, used in the production of glass-ceramics. However, devitrification produces a polycrystalline material, increases optical scattering loss, and is usually a problem to be avoided in manufacture of photonic devices.

As mentioned in Sect. 41.1.1, crystallization tendencies vary widely between various glass systems. The difference between the onset temperature for crystallization and T_g, $\Delta T = T_x - T_g$, is one of the parameters that define the stability of a particular glass. The high stability against devitrification of silicate glasses is one of numerous reasons that they have traditionally dominated glass technology. This stability underlies the manufacture of the structures discussed in Sect. 41.1.1. Other glass forming systems have lower stability against devitrification, often making their technological application a greater challenge [41.46, 47]. For example, a major research thrust in the field of fluoride and chalcogenide glass fibers is the identification of new compositions that exhibit improved stability against crystallization during the fiber drawing process [41.48].

For integrated optics, this form of stability is not always so critical. Amorphous thin films created by chemical vapor deposition, sputtering, evaporation, etc. can often be processed at temperatures below the onset of any significant crystallization. This enables the use of conditional glass formers, such as Si_3N_4, Al_2O_3 and semiconductors, in their amorphous state. However, it is interesting to review the early development of integrated optics, much of it focused on the identification of suitable thin film materials. Studies [41.49–51] that compared waveguide fabrication in polycrystalline (ZnO, ZnS, Al_2O_3, etc.) and amorphous (Ta_2O_5, polymer, sputtered glass) thin films showed that losses were orders of magnitude higher for the polycrystalline films. The lowest losses were initially obtained for polymer films [41.50], attributed to their amorphous structure and the smoothness of surfaces arising from spin casting of films.

If present, uncontrolled crystallization within films can be a dominant source of propagation loss [41.52]. For conditional glass formers, such as Al_2O_3 and TiO_2, crystallite formation must be carefully avoided during the deposition process [41.53]. While less of an issue for natural glass formers, the possibility for crystallization must always be considered when a glass film is subjected to high-temperature processing steps.

Structural Relaxation

Having been frozen in some metastable configuration at the time of formation, an amorphous solid will always be subject to some degree of aging (changes in the material properties on some time scale). At temperatures below T_g, the glass is usually inhibited from crystallization by a significant internal energy barrier. However, subtler changes in the network structure can occur, such as the transition from the initial state to a more stable (but still amorphous) second state. This so-called structural relaxation is manifested in many ways – experiments often probe the change in time of specific volume, structural signatures (such as by Raman spectroscopy), or optical properties of the glass. The rate of structural relaxation is highly dependent on temperature, being relatively fast at temperatures near T_g and significantly slower for temperatures far below T_g [41.18]. In fact, annealing of glasses at a temperature near T_g is a standard technique used to promote rapid structural relaxation and lessen the dependence of material properties on the processing history [41.6]. In many cases, this annealing step is not practical, and structural relaxation over the in-use lifetime of a glass-based device must be considered.

As the following discussion will illustrate, the rate of structural relaxation at temperature T_1 exhibits a logarithmic dependence on the temperature difference $T_g - T_1$. Thus, T_g is an important parameter in characterizing the long-term stability of a glass component. For example, SiO_2 has a very high T_g ($\approx 1100\,^\circ C$) and does not exhibit significant relaxation at room temperature, either on the time scale of typical experiments or over the typical lifetimes envisioned for glass devices. In fact, aging effects in silica fibers are predominately associated with the growth of defects formed in the glass at the time of manufacture. Environmental moisture is a particular contributor to this aging process. On the

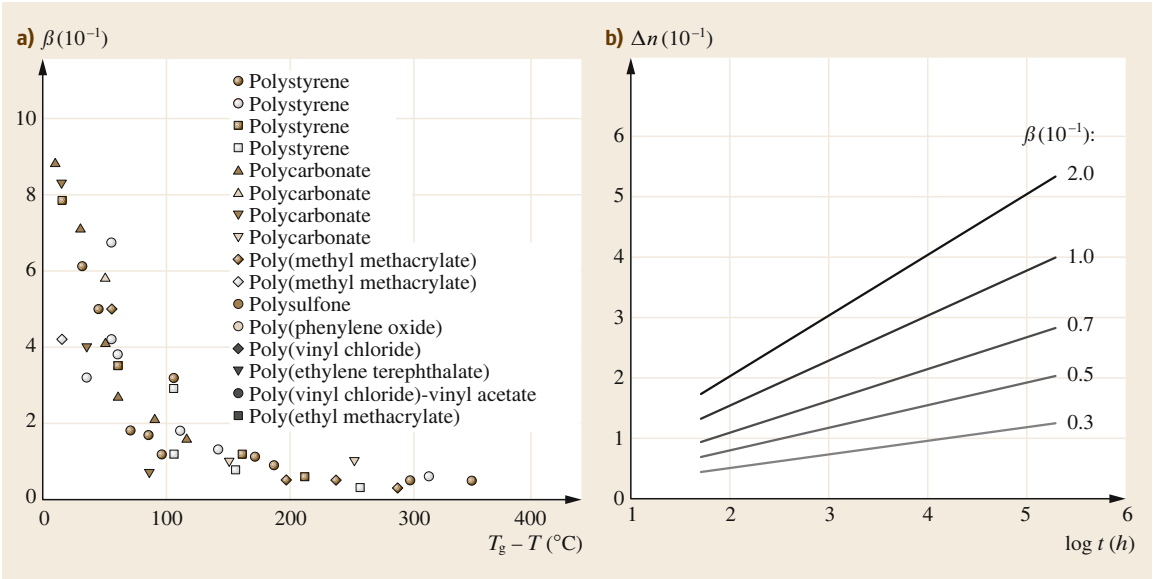


Fig. 41.5 **(a)** A plot of the structural relaxation rate parameter versus the supercooling temperature for a wide range of organic glasses is shown. **(b)** The predicted change in refractive index versus time for an organic glass, with the rate constant β as a parameter. (After [41.54])

other extreme, structural relaxation in low T_g organic glasses is known to be an issue, and intrinsic aging effects in polymers have been widely studied [41.55]. In particular, low T_g glasses exhibit densification with time (relaxation of their specific volume) and a corresponding change in their refractive index. This is a particular concern for interferometric optical devices (gratings, resonators, Mach–Zehnder interferometers), which necessitate tight control of material indices.

Volume relaxation can be characterized by a rate parameter β, defined as

$$\beta = \frac{1}{V}\left[\frac{\partial V}{\partial(\log t)}\right]_{P,T} , \tag{41.7}$$

where V is volume, t is time, P is pressure, and T is temperature. Using the Lorentz–Lorenz expression, the following relationship has been derived from (41.7) [41.54]

$$\left[\frac{\partial n}{\partial(\log t)}\right]_{P,T} = \frac{-\beta\left(\frac{\partial n}{\partial T}\right)_{P,t}}{\alpha} , \tag{41.8}$$

where n is refractive index, $\partial n/\partial T$ is the thermo-optic coefficient, and α is the volume coefficient of thermal expansion. *Zhang* et al. [41.54] further showed that, for polymers, β is approximately a universal function of $T_g - T_1$ as shown in Fig. 41.5. This implies that the time rate of change of refractive index due to structural relaxation is strongly correlated with the parameter $\Delta T_{SC} = T_g - T_1$, termed the supercooling temperature. They also assessed the implications for polymer-based telecommunications devices, concluding that T_g needs to be higher than 300 °C for many applications. This result provides a nice illustration of the impact of T_g on device stability.

While it is not necessarily reasonable to extend the foregoing results directly to inorganic glasses, the essential features of structural relaxation will be similar. The chalcogenide glasses in particular are sometimes called inorganic polymers, partly because of their softness and mechanical flexibility relative to silica-based glasses. Further, the range of T_g for chalcogenide glasses ($\approx$ 50–550 °C) is similar to that of organic polymers, and studies on chalcogenide glasses [41.19] support the conclusion that structural relaxation would present similar restrictions on inorganic and organic glasses. Thus, lifetime restrictions due to relaxation should be considered in the application of any glass having relatively low T_g, including chalcogenide glasses, fluoride glasses, and many nonsilicate oxide glasses [41.6].

Both the processing temperatures (during manufacture) and the in-use temperatures are important considerations in assessing the stability of a glass. Telecommunication devices are typically designed to withstand in-use temperatures between −40 and 85 °C [41.56].

As an example of a more demanding application, it has been proposed that glasses are good candidates for fabrication of integrated photonic elements on future integrated circuit chips [41.57]. In-use temperatures on modern microprocessors can reach 100–200 °C.

Finally, it should be noted that any energy applied to the glass (light, electron beams, etc.) might induce structural relaxation. This has been most widely studied for the chalcogenide glasses, which exhibit a wide array of photoinduced structural changes [41.58]. A fascinating example is the so-called photoinduced fluidity phenomenon [41.59], in which intense sub-bandgap light can reduce the viscosity of a chalcogenide glass by several orders of magnitude (causing the glass to flow or melt). This effect is completely athermal, and in fact is often enhanced at low temperatures. While an extreme example, it illustrates the possibility of aging effects due to nonthermal processes in glass devices. For photonic devices, the main concern is usually the effect of light exposure over the in-use lifetime.

Metastability as Enabler

One highly desirable implication of metastability is the possibility for direct patterning of photonic structures in glasses using energetic beams (electron beams, ion beams, light). These effects can also be exploited for postfabrication trimming of devices [41.60], which is important since many optical components have tolerances beyond the capabilities of practical fabrication processes.

Beam-induced effects in glasses are generally linked to their random network structure. That is, unlike crystals, glasses contain a range of internal *defects* – wrong bonds, dangling (missing) bonds, impurity atoms that act as network modifiers, as well as their inherent variation in bond angles. Addition of energy can cause a glass to undergo a transition from one metastable state to another. This transition is often accompanied by a change in the density of defects or by a structural change – rearrangement or reorientation of bonds. Associated with these structural changes are changes in the physical, chemical, and optical properties of the glass. Most existing knowledge about these induced structural changes is of an empirical nature, whereas the underlying physical mechanisms are still the subject of research.

The ability to induce transitions between metastable states is probably a universal property of amorphous matter. Perhaps for practical reasons, such as the availability of light sources of suitable wavelength, metastability was initially studied in amorphous semiconductors and low-bandgap chalcogenide glasses. It should be added, though, that chalcogenide glasses

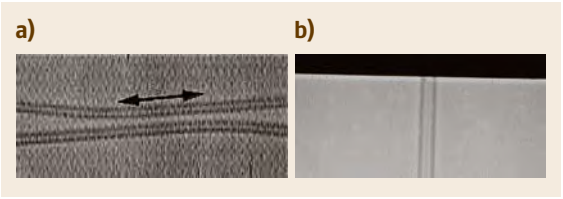


Fig. 41.6 **(a)** A directional coupler is shown, as written in a multicomponent silicate glass by an ultrafast Ti:sapphire laser. (After [41.64]). The *black arrow* indicates the coupling region. **(b)** Microscope image of a rib waveguide formed by direct exposure of a chalcogenide glass (As_2Se_3) in a UV mask aligner, followed by wet etching. (After [41.65])

seem to be unique with respect to the variety of induced structural changes they exhibit or at least the magnitude of those changes [41.58]. This might be related to the relatively weak, highly polarizable covalent bonds that are characteristic of chalcogenide glasses. Nevertheless, the later discovery [41.61] of photoinduced refractive index changes in Ge-doped silica glass prompted a huge research and development effort because of the technological importance of silicate glasses. The main outcome of this work was the fiber Bragg grating (FBG), which is a critical component in modern telecommunications and sensing networks. FBGs are manufactured by inducing a quasi-permanent index change within the Ge-doped core of a standard silica fiber, by exposure to a UV laser beam. Writing wavelengths are in the 150–350 nm range, corresponding to photon energies below the nominal bandgap of SiO_2. The mechanisms underlying the index increase are still debated [41.12, 61]. However, it is believed that contributing factors are absorption by defects in the glass (possibly generating further defects) and densification due to some sort of structural relaxation [41.62]. Typical index changes for standard fiber cores are in the 10^{-4}–10^{-3} range, but several techniques (addition of other dopants, hydrogen-loading, etc.) have been developed to enable changes as high as 10^{-2}. Given the discussion in the preceding section, it is reasonable to ask whether photoinduced changes are themselves stable. The answer, in some cases, seems to be yes. Researchers have concluded that fiber Bragg gratings can have an operational lifetime of greater than 20 years, even at elevated temperatures of 80 °C [41.63].

Photosensitivity has been widely explored as a means to directly pattern integrated optics devices, especially in silica-based glasses [41.66]. In general, direct patterning is induced using photon energies close to the nominal bandgap energy of the glass or using

intense light of lower photon energy [41.62]. In the latter case, photoinduced effects arise from nonlinear effects such as multiphoton absorption. In particular, the widely available Ti:sapphire laser has been applied to the writing of microstructures in a variety of glasses [41.64]. Because of the nonlinear nature of the writing process, writing with sub-bandgap light allows the direct patterning of 3-dimensional structures (Fig. 41.6).

Photosensitivity has been demonstrated within all of the standard glass families, and in many other amorphous thin films (Sect. 41.2). Because of their large photosensitivity, direct patterning of integrated optics structures in chalcogenide glasses is possible using low-power light sources [41.65, 67].

41.1.3 Glass as Host Material

For completeness, it should be mentioned that glasses are unique in their ability to incorporate a wide range of dopants, sometimes in very high concentration. This is related to the random network structure of a glass. Foreign species are much more likely to find a suitable bonding site, or simply space to reside, inside a glass versus a crystal. These foreign species can be rare-earth ions, transition-metal ions, or semiconductor or metal nanoparticles, for example. As a result, a glass provides significant scope for active functionality. We can transform a passive glass into a laser glass, a nonlinear optical glass, or a magneto-optic glass, for example, by appropriate doping [41.3, 6].

41.2 Glasses for Integrated Optics

In general, thin films can be amorphous, polycrystalline, or crystalline. *Tien* and *Ballman* [41.51] provided an early review of waveguide results achieved for various materials lying within each of these categories. Unless the crystal grain size is sufficiently small relative to the wavelength of interest [41.53], polycrystalline films are too lossy for integrated optics. It should be noted that high attenuation might be tolerable if circuit length is sufficiently short. The present discussion is concerned with amorphous films, which Tien further subcategorized as low index ($n < 1.7$), medium index ($1.7 < n < 2$), and high index ($n > 2$). For convenience, we will follow a similar approach in the following sections. This categorization is somewhat arbitrary, especially when considering material systems (such as the silicon oxynitride system) that enable a range of refractive index. It is interesting to note that 1.7 is the approximate upper limit for the refractive index of organic glasses [41.56]

As noted in Sect. 41.1, increased circuit density is one of the primary goals of integrated optics research [41.12]. Whether employing traditional total internal reflection effects or photonic bandgap materials, increased density relies on high index contrast between at least two compatible materials. Note that the terminology *high index* and *high index contrast* are rather imprecise. For example, in silica PLC technologies core-cladding index differences of $\Delta n \approx 0.02$ have been labeled as *superhigh* index contrast [41.13]. This is a very small value, however, relative to the index contrasts that characterize SOI photonic wire waveguides [41.36].

41.2.1 Low-Index Glassy Films

SiO_2-Based Glasses

As mentioned in the introduction, the SiO_2 on silicon PLC system is the most widely used and developed glass-based integrated optics platform. PLC development (mainly in the 1990s) was driven by fiber-optic long-haul communication systems, especially the emergence of wavelength-division multiplexing (WDM). Various devices, but especially arrayed-waveguide grating (AWG) wavelength demultiplexers, were developed to a very high degree of sophistication by the telecommunications industry. Good early [41.24] and recent [41.12, 13] reviews are available in the literature. A brief overview of the technology is given below, as it illustrates some of the advantages and challenges associated with glass-based photonic integration.

As mentioned in Sect. 41.1.1, commercial PLC waveguides are fabricated primarily by FHD [41.13] or by CVD [41.12]. In addition, considerable research has been conducted on sol–gel synthesis [41.68], with the aim of reducing fabrication costs and providing greater flexibility over the choice of glass compositions. Typically, the undercladding is pure silica glass. To raise its refractive index, the core layer is doped with Ge or P. Ge-doped SiO_2 is known for its photosensitivity, as discussed in Sect. 41.1.2. Addition of P lowers the viscosity and characteristic reflow temperature, enabling the processing options discussed in Sect. 41.1.1. Typical relative index offset Δ between the core and cladding is in the 0.3–2% range, corresponding to minimum waveguide bend radius R in the 2–25 mm range [41.13]. After reactive ion etch-

ing (RIE) to form an approximately square waveguide core, an upper cladding is deposited and subjected to heat treatment. The upper cladding is often a borophosphosilicate glass (BPSG), partly to enable reflow. Further, the B and P dopants have opposite effects on refractive index allowing a nearly symmetric waveguide structure to be obtained.

Given the target applications, it is not surprising that PLC technology placed a great emphasis on efficient coupling between the integrated waveguides and external fiber waveguides. Thus, waveguide cross-sectional dimensions and refractive index contrast between core and cladding layers were tailored to provide a good impedance match (low reflection and good modal overlap) to standard fiber. By employing essentially the same glass (doped SiO_2) as that used to construct fiber, it is even possible to achieve intimate reflowed (fusion type) coupling between the integrated and fiber guides [41.24]. The emphasis on impedance matching is due to the great importance of minimizing insertion loss in fiber systems, but presents some practical limitations:

1. The relatively low index contrast ($\approx 10^{-3}$) between core and cladding necessitates thick glass films. For example, the undercladding or buffer layer, often a thermally grown SiO_2 layer, must typically exceed 12 μm in order to negate radiation losses into the high-index silicon substrate. Further, high-temperature anneals (typically 900–1150 °C) are required to drive out hydrogen impurities and to reflow the core and upper cladding layers. The combination of thick films, high-temperature anneals, and thermal expansion mismatch can result in wafer bending and damage to the glass films. This is partly alleviated in practice by depositing identical layers on both sides of the Si wafer [41.12], which adds cost and complexity.
2. Since modal area and minimum bend radius scale inversely with core-cladding index contrast, traditional PLC waveguides do not support high-density optical integration. As mentioned above, index contrasts in the 0.3–2% range correspond approximately to bending radii in the 25–2 mm range [41.13].

In recent work at Corning [41.69], very low-loss (< 0.1 dB/cm) waveguides and ring resonators were realized in plasma enhanced chemical vapor deposition(PECVD)-grown silica-germania waveguides having an index contrast as high as 4%. Such high index contrasts can accommodate bending radii of less than 1 mm, which is comparable to the range explored recently by IBM researchers using silicon oxynitride materials [41.70].

Amorphous Aluminium Oxide

Sapphire (crystalline Al_2O_3) is amongst the most important solid-state laser hosts. Alumina (polycrystalline or amorphous Al_2O_3) is an important industrial material in its own right, possessing outstanding mechanical and thermal properties [41.21]. Amorphous Al_2O_3 films are of interest for integrated optics for several reasons [41.53]. First, the refractive index is relatively high (although falling within the low index range specified above), typically $n \approx 1.65$. Second, Al_2O_3 is an excellent host for rare-earth and transition-metal dopants. As discussed in Sect. 41.3, rare-earth dopants of interest are typically trivalent, matching the valency of Al ions in Al_2O_3 [41.71]. As a result, rare-earth ions can be incorporated easily into the alumina matrix. In short, Al_2O_3 can homogeneously dissolve large concentrations of rare-earth ions and is therefore of interest for realization of integrated amplifiers and light sources. Finally, Al_2O_3 films typically have excellent transparency from the UV to mid-IR range.

Fluoride Glasses

Heavy-metal fluoride glasses (typically fluorozirconate glasses) such as ZBLAN have been widely studied since their discovery in 1974 [41.48]. One of their outstanding attributes is a wide transparency range, extending from the UV well into the mid-IR (Fig. 41.1). They have received considerable attention as fiber optic materials, because theory predicts a minimum absorption well below that of silica glass. Further, rare-earth ions exhibit the greatest number of useful radiative transitions when embedded in fluoride glasses [41.72]. This is due to their wide transparency window and low characteristic phonon energies (Sect. 41.3).

Fluoride glasses typically have refractive indices in the 1.47–1.57 range, which is advantageous in terms of being well matched to silica glasses. Thin film deposition of these complex multicomponent glasses is difficult, and thermal expansion mismatches with standard substrates create further challenges. For these reasons, fluoride glasses have not been widely explored for applications in integrated optics. The so-called PZG fluoride glasses (PbF_2-ZnF_2-GaF_3) have been successfully deposited using straightforward evaporation techniques, enabling relatively low-loss waveguides and erbium-doped amplifiers on a silicon platform [41.73].

41.2.2 Medium-Index Glassy Films

Silicon Oxynitride

Silicon oxynitride (SiON) films are generally deposited by a CVD technique. The promise of this material system for integrated optics was identified in early

work [41.51]. One of the main attributes of SiON is that it is a standard material system employed in silicon microelectronics, and the thin film technology has been developed accordingly [41.53, 70]. Because of this, SiON is currently being studied as potential material for on-chip interconnects [41.57]. Further, the system enables a continuous range of refractive index from approximately 1.45 (SiO_2) to 2 (Si_3N_4) at 1550 nm wavelength. This index range has been extended to approximately 2.2 by deposition of nonstoichiometric, silicon-rich nitride films [41.74]. SiON and SiN have been amongst the most explored materials for realization of microring resonator structures [41.60, 74].

As an optical material, SiON has some drawbacks. The main one is the presence of hydrogen impurities in films deposited by traditional techniques. Overtones due to hydrogen bonds (mainly N–H and O–H) can produce impractically large values of loss in the 1300 and 1550 nm telecommunications bands [41.69]. Long term, high-temperature annealing (typically at $>$ 1000 °C) is required to reduce this loss. Interestingly, structural relaxation on the time scale of hours and days has been observed in such annealed films [41.53]. Perhaps related to this metastability, UV light has been used as a means to trim the refractive index of SiN-based microring resonators [41.60]. Modified deposition processes that can produce SiON films having low stress and low hydrogen content (without requiring a high temperature annealing step) have been reported recently [41.70, 74]. However, inherent film stress can limit the maximum thickness to a few hundred nanometers in some cases [41.57].

41.2.3 High-Index Glassy Films

Amorphous Metaloxides

Many amorphous metaloxides (Y_2O_3, Nb_2O_5, Ta_2O_5, etc.) have traditionally been used as high index layers in optical thin film stacks [41.56]. It is logical to look at these materials in the hunt for high-index thin films for integrated optics. Often these oxides are miscible with SiO_2, making it possible to tune the refractive index and other material properties over a range of values (as in the case of silicon oxynitride) [41.75].

TiO_2 films are used as dielectric layers, optical coatings, and to protect underlying materials from mechanical damage or corrosion [41.53]. The high dielectric constant of TiO_2 is widely exploited in microelectronics applications. It also has useful optical properties, including a refractive index as high as 2.6 (depending on film deposition details) and good transparency from the UV to mid-IR. The TiO_2-SiO_2 material system has good properties for sol-gel synthesis of thin films [41.68]. Integrated waveguides based on TiO_2 have been studied [41.53, 75], and AWG devices in TiO_2-rich oxide glass with $n \approx 1.9$ were recently reported [41.76].

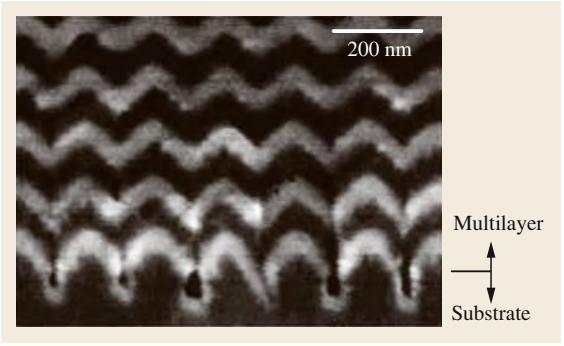


Fig. 41.7 A three-dimensional photonic crystal fabricated using the autocloning growth technique is shown. The *bright* and *dark* layers are TiO_2 and SiO_2 glass, respectively. (After [41.80])

Ta_2O_5 (with $n \approx 2.1$) is another well-studied material, used for example as a high dielectric constant material in microelectronics. It has low absorption in the visible to near infrared wavelength range, and has been recently studied as a material for active integrated optics [41.77, 78]. Further, Ta_2O_5/SiO_2 has been applied to the study of compact microring resonators [41.79].

TiO_2 and Ta_2O_5 have been used in the fabrication of three-dimensional photonic crystals with fundamental bandgaps in the visible and near infrared [41.80, 81]. An example is shown in Fig. 41.7.

Heavy-Metal Oxide Glasses

These glasses are based primarily on the oxides of bismuth (Bi_2O_3), lead (PbO), tellurite (TeO_2) and germanium (GeO_2) [41.9]. Amongst these heavy-metal oxide glass families, except for GeO_2 the majority of glass-forming compositions are conditional glass formers. They have relatively low phonon energies and transmit further into the infrared than most oxide glasses [41.82]. They also have interesting magneto-optic [41.6] and nonlinear optical [41.3] properties.

The major difference between the classical network forming like bismuth, silica and tellurite glass families is the presence of multiple structural units (TeO_3, TeO_4, and $TeO_{3+\delta}$ in tellurite); each of which has a lone pair electron site which may bridge with polyvalent cations and enable network bridging with phosphates, germanates, tungstates, and borates[41.83].

While TeO_2 is a conditional glass former, the addition of other oxides can result in stable glasses with many interesting properties [41.82]. They typically have tunable refractive index in a wide range from 1.68 to greater than 2 [41.84], low phonon energy,

large optical nonlinearities, and a high acousto-optic figure of merit. Further, they have been widely studied as hosts for rare-earth ions. Of particular interest has been the wide bandwidth of the 1550 nm emission exhibited by erbium in tellurite glass [41.85]. Because of these numerous attractive properties, tellurite glasses and amorphous TeO_2 films ($n \approx 2.2$) have received some attention for applications in integrated optics [41.86].

Chalcogenide Glasses

Many chalcogenide alloys, based on S, Se, or Te, are excellent glass formers. In fact, Se is the only element able to easily form a glass on its own [41.19]. These glasses are characterized by very low phonon energies, good transparency in the mid-infrared, and a wide array of beam-induced structural changes [41.58]. Further, chalcogenide glasses are essentially amorphous semiconductors, with an electronic absorption edge typically in the visible to near infrared. Related to this, they have very high refractive indices, ranging from 2.2 for sulfide glasses to greater than 3 for some telluride glasses. They are currently of industrial interest for night-vision optics [41.44, 45], fibers for infrared transmission [41.48], and optical and electronic memory elements [41.58].

Chalcogenide glasses have many appealing features for integrated optics. High-quality thin films can often be deposited by straightforward techniques such as evaporation or sputtering. Their high indices make them suitable for realization of compact waveguides [41.65] and photonic crystals [41.87]. Their beam-induced properties enable many unique processing options. Some chalcogenide alloys have good potential as laser glasses [41.48], with the ability to uniformly dissolve large concentrations of rare-earth ions. Perhaps of greatest interest, the chalcogenide glasses are amongst the most promising materials for nonlinear integrated optics [41.88, 89]. On this last point, several chalcogenide glasses have ultrafast optical Kerr response 2 to 3 orders of magnitude higher than that of SiO_2, while satisfying basic figure of merit criteria for all-optical switching. Results for metal-doped chalcogenide glasses [41.89] suggest even higher nonlinearities might be attainable. Combined with the possibility of realizing highly confining waveguides, this makes chalcogenide glasses one of few material systems that can realistically satisfy long-standing goals for nonlinear processing in integrated optical circuits. For example, it has been estimated [41.88] that 1 pJ pulses of 1 ps duration (≈ 1 W peak power) could induce phase shifts (due to self-phase modulation) greater than π in feasible chalcogenide glass waveguides with length on the order of a few centimeters. This is predicated on realization of low-loss waveguides with modal area of $\approx 1\,\mu m^2$. Recent experimental results suggest that this goal is within reach [41.65].

41.3 Laser Glasses for Integrated Light Sources

Active functionality includes means to generate and detect light (especially stimulated emission and absorption) and means to control (switch, modulate, etc.) light signals. Numerous material properties are employed in active photonics, including thermo-optic, acousto-optic, magneto-optic, electro-optic, and nonlinear Kerr effects. Many glasses have attractive thermo-optic or acousto-optic properties, and a few examples were cited in Sect. 41.2. Also in Sect. 41.2, some glasses with promising nonlinear optical properties were discussed. Elsewhere in this volume, *K. Tanaka* has provided an excellent review of nonlinearities in photonic glasses. A recent, thorough review of magneto-optic glasses is also available [41.6].

Arguably, the most critical element for photonic integration is an integrated light source. Rare-earth-doped glasses are well-established laser media, used especially for the realization of bulk and fiber lasers. Considerable effort has been directed towards development of integrated amplifiers and lasers based on such glasses. In the following, we attempt to highlight some ways in which light sources based on glasses are uniquely enabling, relative to those based on crystalline materials. The performance advantages discussed below combined with the properties discussed in earlier sections (fabrication options) make glasses particularly attractive.

41.3.1 Advantages of Glass-Based Light Sources

Stimulated-emission devices in glass are almost always based on trivalent rare-earth dopants [41.6]. Thus, the term laser glass can usually be equated with the term rare-earth-doped glass. Rare-earth-doped laser glasses have been widely studied and reviewed [41.71, 72]. Further, recent and comprehensive reviews on Nd- and Er-doped integrated glass amplifiers and lasers are available [41.6, 90]. Relative to single crystal hosts, glass hosts result in rare-earth ions exhibiting broadened luminescence lines and lower peak stimulated-emission cross-sections [41.72]. This property is a result of the random network structure of glasses; embedded rare-earth ions exist in a range of local environments. The

broadened, weaker emission is of great importance for the realization of broadband, low-noise fiber amplifiers.

It should be noted that the semiconductor injection laser is an extremely advanced technology, and is the dominant type of integrated light source at present and for the foreseeable future. Semiconductor lasers have important advantages over any glass-based device demonstrated to date. First, semiconductor gain media typically have gain coefficients of the order $\approx 100\,\mathrm{cm}^{-1}$ [41.23]. By comparison, glasses require high rare-earth dopant concentration to achieve gain coefficients exceeding $1\,\mathrm{cm}^{-1}$. Thus, semiconductor optical amplifiers (SOAs) and lasers have cavity lengths measured in tens to hundreds of μm while it is typical for glass waveguide amplifiers and lasers to be measured in cm. Second, semiconductor light sources are pumped electrically while glass devices are typically pumped optically. Electrical pumping is highly desirable for optoelectronic integration of photonic devices on electronic chips. However, as discussed below, integrated glass waveguide lasers have important advantages of their own [41.91].

The lifetime of the metastable lasing level in rare-earth-doped glasses is usually on the order of ms, much longer than the ns lifetimes typical of semiconductor gain media. This long lifetime implies that the gain does not change rapidly with variations in input power (pump or signal). This is an essential feature of the commercially important erbium-doped fiber amplifier (EDFA); the long lifetime ($\approx 10\,\mathrm{ms}$) of the $^4\mathrm{I}_{13/2}$ level of erbium in silicate glass contributes to low-noise operation, high pump efficiency, and low crosstalk between wavelength channels in a WDM system [41.92]. Further, since the relaxation oscillations in glass waveguide lasers occur at relatively low frequency, glass lasers can be mode-locked at correspondingly much lower repetition rates compared to semiconductor lasers [41.91]. This can enable much higher peak intensities from the glass laser.

Related to the discussion in Sect. 41.1, lower cavity loss (higher cavity Q) is generally possible for glass devices. Further, the long metastable lifetime of the rare-earth transition allows glass lasers to have linewidths approaching the Schawlow–Townes limit [41.93]

$$\Delta\nu = \frac{2\pi h\nu_0(\Delta\nu_\mathrm{c})^2}{P}\left(\frac{N_2}{N_2 - N_1}\right), \tag{41.9}$$

where ν_0 is the laser center frequency, P is the laser output power, N_2 and N_1 are the population densities of the upper and lower lasing levels, and $\Delta\nu_\mathrm{c} = (1/(\pi t_\mathrm{c}))$ with t_c the photon lifetime in the cavity. The linewidth ($\Delta\nu$) of a glass laser system and noise figure is underpinned by the uncertainty equation which is given by [41.94]

$$\Delta\varphi\Delta n \geq \frac{h}{4\pi}\,, \tag{41.10}$$

where $\Delta\varphi$ is the phase difference, and Δn is the population difference. The high Q of glass laser cavities coupled with relatively high output powers enables linewidths less than 10 kHz, orders of magnitude below that of semiconductor DFB lasers [41.91]. The high cavity Q and long metastable lifetime is also advantageous for achieving ultrastable passive mode-locking, with low timing jitter and pulse-to-pulse power variation.

Finally, glasses offer the possibility of integration on various substrates. While semiconductor lasers are inherently integrated structures, they are not easily transportable between platforms. For example, III–V semiconductor lasers have not shown great promise (in spite of heroic efforts in some cases) to satisfy the desire for a compact, truly integrated light source on the silicon electronics platform. Glasses (perhaps doped with semiconductor nanocrystals) are viewed by some as the more promising route to achieving such a goal [41.71].

41.3.2 Alternative Glass Hosts

The theoretical maximum gain (cm^{-1}) of a rare-earth-doped glass waveguide amplifier can be expressed [41.90] as $\gamma_\mathrm{p} = \Gamma\sigma_\mathrm{p}N_\mathrm{RE}$, where σ_p is the peak (versus wavelength) stimulated emission cross-section (cm^2) for the transition of interest and N_RE is the volume density (cm^{-3}) of the rare-earth ions. Γ is a dimensionless factor (lying between 0 and 1) that accounts for the spatial overlap of the waveguide mode (at the wavelength to be amplified) and the active ions producing the gain. It can be optimized through waveguide design, irrespective of the glass host, and will not be considered further here. The expression for γ_p neglects all waveguide losses (due to scattering, etc.) and assumes that all of the rare-earth ions have been promoted to the desired lasing level; i. e., a complete population inversion. It is therefore an ideal and unattainable limit, but is useful for framing the following discussion.

Since compactness is a central goal of integrated waveguide lasers, alternative glass hosts can be compared on the basis of the maximum gain (γ_max) that they enable in practice (for a given transition of a given rare-earth ion). Further, since low-noise operation relies on a near complete population inversion [41.90], it is desirable that $\gamma_\mathrm{max} \approx \gamma_\mathrm{p}$. In simple terms, the glass should dissolve a large concentration of the rare-earth ion (high N_RE), should result in a large stimulated-emission cross-section for the desired transition, and should enable the

realization of a nearly complete population inversion. The importance of other practical considerations, such as stability, processing options, and physical properties of the glass, will depend on the intended application. Some hosts provide unique advantages, such as flexible pumping options, as discussed in Sect. 41.3.3.

As mentioned, laser transitions in glasses are generally provided by radiative decay between two energy levels of a trivalent rare-earth ion. Perhaps the most important example is the $^4I_{13/2}$ to $^4I_{15/2}$ transition of Er^{3+}, which produces luminescence in the 1500–1600 nm wavelength range. Once an ion has been promoted (by pumping) to the upper lasing level, it will eventually transition to another state by interactions with the glass, impurities in the glass, the photon fields (at the signal or pump wavelength), or with other rare-earth ions in its vicinity [41.72]. The metastable lifetime of the upper lasing level can be expressed [41.90]

$$\frac{1}{\tau} = A + W_{\mathrm{MP}} + W_{\mathrm{ET}} + W_{\mathrm{IMP}} \,, \tag{41.11}$$

where A is the effective rate of spontaneous radiative decay to all lower lying levels, W_{MP} is the rate of nonradiative decay due to multiphonon energy exchanges with the glass, W_{ET} is the rate of nonradiative energy transfer due to interactions between closely spaced rare-earth ions, and W_{IMP} is the rate of energy transfer to quenching impurity centers in the glass. The first three terms on the right will be discussed below. For the last term, a classic example is the quenching of the 1550 nm luminescence band of erbium due to resonant energy transfer to OH^- impurities in silica glass [41.72].

The choice of a particular glass host will impact several important properties of a given transition: pumping efficiency, peak gain, linewidth, metastable lifetime, etc. Alternative glasses can be compared on the basis of a few key parameters, as discussed in the following subsections. Representative data for erbium in various glasses is given in Table 41.1.

Stimulated-Emission Cross-Section

Stimulated transitions of rare-earth ions in glass tend to be predominately driven by electric dipole interactions [41.72]. For a given transition of interest, the spontaneous-emission probability can be expressed in cgs units as [41.6, 95]

$$A = \frac{64\pi^4 e^2 \chi}{3h(2J+1)\lambda_p^3} S \,, \tag{41.12}$$

where J is the total angular momentum of the upper lasing level, λ_p is the peak emission wavelength, and χ is the local field correction factor. For electric-dipole interactions of an ion in a dielectric medium, $\chi \approx n((n^2+2)^2)/9$, with n the refractive index of the host glass. S is the quantum-mechanical line strength for the transition. Further, the peak stimulated-emission cross-section can be expressed in terms of the spontaneous-emission probability

$$\sigma_p = \left(\frac{\lambda_p^4}{8\pi c n^2 \Delta\lambda_{\mathrm{eff}}}\right) A \,, \tag{41.13}$$

where $\Delta\lambda_{\mathrm{eff}}$ is the effective linewidth of the transition, n is refractive index and c is the vacuum speed of light. From (41.12) and (41.13), the host-dependent factors that influence σ_p are the refractive index, the line strength, the effective linewidth, and to a lesser extent the peak emission wavelength (which typically varies only slightly between different hosts). The local field correction factor is significant in hosts with large refractive index, and can result in an enhancement of the stimulated-emission cross-section and a reduction of the radiative lifetime. This is especially true for chalcogenide glass hosts, which typically have refractive index in the 2–3 range.

Metastable Lifetime

In the limit of low rare-earth-dopant concentration, W_{ET} in (41.11) is zero because the ions are sufficiently well separated to negate their interaction. If the difference in energy between the upper lasing level and the adjacent state is several times the effective phonon energy, then W_{MP} can be neglected to first order. Further neglecting impurity quenching, we can then assert that $\tau_0 \approx \frac{1}{A}$, where τ_0 is the metastable lifetime in the limit of low rare-earth concentration. From Einstein's relations, $A \approx \sigma_p$, so it follows that $\tau_0 \approx \frac{1}{\sigma_p}$. In short, hosts that result in an enhancement of the peak stimulated-emission cross-section (because of an enhancement of the electric-dipole interaction or because of a high local field correction factor) will also result in a reduction in metastable lifetime. In other words, both stimulated and spontaneous emission rates are enhanced.

The inverse scaling of stimulated-emission cross-section and metastable lifetime represents a tradeoff, as it is generally desirable for the lifetime to be as large as possible. Some of the advantages of long lifetime were discussed in Sect. 41.3.1. In addition, the pumping efficiency (gain per applied pump power) of a waveguide amplifier scales directly with τ [41.6, 90] and, therefore, the threshold for continuous wave (CW) lasing scales inversely with τ [41.96].

Concentration Quenching

Glasses differ greatly in the amount of a given rare-earth dopant that they are able to dissolve. To avoid

Table 41.1 Representative parameters for erbium ions embedded in various types of glass. (After [41.68, 71, 72])

Glass host	Refractive index n	Peak stimulated emission cross-section σ_p (10^{-21} cm^2)	Metastable lifetime τ (ms)	Quenching concentration ρ_q (10^{20} cm^{-3})	Effective luminescence bandwidth (nm)
Silica	1.46	7	12	–	11
Amorphous Al_2O_3	1.64	6	7.8	–	55
Aluminosilicate	1.5	5.7	10	3.9–6.0	43
Phosphate	1.56	8	10	3.9–8.6	27
Fluoride	1.53	5	9	3.8–5.3	63
Tellurite	2.1	13	3.3	–	80
Sulfide	2.4	20	2.5	3.2	–

problematic ion–ion interactions, the rare-earth ions should be incorporated homogeneously into the glass structure. In the extreme case of high concentration, the rare-earth ions will form microscopic clusters (phase separation). Such clustering is highly detrimental, as typically all of the ions within a cluster are effectively removed from the desired stimulated-emission process [41.90]. As is well known [41.71, 72], the onset of clustering occurs at quite small values ($\approx 0.1\%$) in pure silica glass. The addition of Al_2O_3 to SiO_2 allows silica-based fibers to dissolve a significantly higher concentration. For example, a value of 10–20 for the Al ion to Er ion concentration ratio has been shown to greatly reduce clustering of Er ions in silica glass [41.96]. Interestingly, the addition of Ga (with similar ratio) to chalcogenide glasses has been shown to provide a similar reduction in rare earth-ion clustering [41.48]. These additives essentially modify the glass network, and create sites for isolated rare-earth ions to be incorporated. For similar reasons, multicomponent aluminosilicate and phosphosilicate glasses have been favored in the development of integrated waveguide amplifiers, where the rare-earth concentration must be orders of magnitude higher than in fibers [41.90].

Even in the absence of significant clustering, ion–ion interactions can occur at high concentrations. This is simply due to the reduction in inter-ion spacing, and is exacerbated by any nonuniform (nonhomogeneous) distribution of the rare-earth ions in the glass host. These interactions are manifested by a reduction in the metastable lifetime, often well described by the semiempirical expression [41.68, 72]

$$\tau(\rho) = \frac{\tau_0}{1 + \left(\frac{\rho}{\rho_Q}\right)^p}\,, \tag{41.14}$$

where ρ is the rare earth-ion concentration, ρ_Q is the so-called quenching concentration, and p is a fitting parameter ($p \approx 2$ for interactions between pairs of ions). The parameter ρ_Q is useful for comparing glasses in terms of their ability to uniformly dissolve a given ion.

Table 41.2 Characteristic maximum phonon energies for a variety of glass hosts. (After [41.6, 72])

Glass host	Phonon energy (cm^{-1})
Borate	1400
Phosphate	1200
Silicate	1100
Germanate	900
Tellurite	700
Heavy-metal fluoride	500
Chalcogenide (sulfide)	450
Chalcogenide (selenide)	350

Phonon Energies

The characteristic phonon energies of a glass depend on the weights of its constituent atoms and the strength and nature (ionic or covalent) of its bonds. Typical values are shown in Table 41.2. The rate of multiphonon decay (W_{MP}) between two energy levels depends exponentially on the number of phonons required to bridge the energy gap. Thus, the phonon energy has a great impact on the ultimate efficiency of a desired radiative transition. Low phonon energy can be a good or bad thing, depending on the transition of interest and the particular pumping scheme.

Mid- to far-infrared transitions of rare-earth ions can exhibit reasonably high quantum efficiency in low phonon energy hosts, such as fluoride, tellurite, and especially chalcogenide glasses. If the same ions are embedded in silicate or phosphate glasses, these transitions are completely quenched by nonradiative processes at room temperature. For this reason, rare-earth-doped chalcogenide glasses are of interest for realization of long-wavelength amplifiers and lasers [41.48, 71]. Also unique to low phonon energy hosts is the possibility of efficient upconversion lasers [41.97]. In simple terms, the long lifetimes of numerous energy levels allows processes such as ion–ion interactions and excited state absorption (ESA) to efficiently populate the higher energy levels. By contrast, these levels are rapidly depopulated by phonons in oxide glasses. Conversely, population of the higher levels is highly detrimental if the desired transition is between two lower levels. For

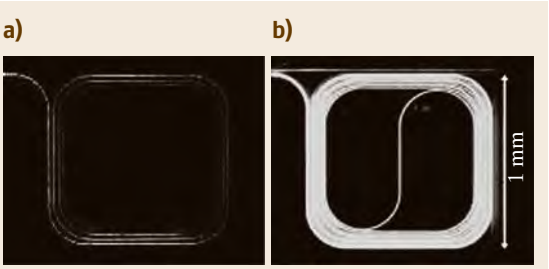


Fig. 41.8a,b Spiral geometry demonstrated for Al_2O_3-based waveguide amplifiers, where erbium was introduced by ion implantation **(a)** and co-sputtering **(b)**. The bright green emission is a result of ion–ion interactions causing up-conversion for the co-sputtered sample. (After [41.99])

example, the efficient pumping of EDFAs using 980 nm wavelength sources relies on the rapid decay (via multiphonon processes) of ions from the $^4I_{11/2}$ pumping level to the $^4I_{13/2}$ lasing level. In fluoride and chalcogenide glass hosts, ions raised above the $^4I_{11/2}$ level tend to become trapped in higher levels (so-called *population bottlenecking* [41.72]). Cerium co-doping has been shown to alleviate this problem [41.73]. Another approach is addition of light elements to the glass network, to increase the phonon energy [41.85].

41.3.3 Progress Towards Integrated Light Sources in Glass

Per the preceding discussion, important goals for glass-based lasers include size reduction and the need for simplified optical or (ideally) electrical pumping schemes. It should be noted that integration brings its own advantages. As is well known, waveguide lasers have greatly enhanced pump efficiency (gain per applied pump power) relative to their bulk solid-state counterparts. This is due to the confinement of the pump beam and the stimulated emission beam within a small volume. As explored recently in [41.98], this benefit scales with confinement. Amplifiers based on very high-index-contrast waveguides can have extremely high pump efficiency, as well as requiring only a small amount of wafer real estate. Thus, while the gain per unit length is limited by concentration quenching in rare-earth-doped glasses, devices can still be very compact through the use of folded spiral waveguide layouts as illustrated in Fig. 41.8 [41.99].

In terms of maximizing gain per unit length, phosphate glasses have produced the best results to date. This is due to their ability to uniformly dissolve large concentrations of rare-earth ions, and the high cross-sections exhibited by those ions. The high phonon energy can also be an advantage as discussed above. Phosphate glass disks (typically ≈ 1 mm thick) have been widely used for realization of compact microlasers [41.100] and ultrafast modelocked lasers with pulse length as short as 220 fs [41.101] and repetition rate exceeding 40 GHz [41.102]. Phosphate glass fiber lasers with a length of only 7 cm have recently generated 9.3 W of output power [41.103]. Sputtered phosphate glass waveguides have produced gains exceeding 4 dB/cm [41.104]. More recently, a phosphate glass co-doped with 8% erbium and 12% ytterbium was used to realize ion-exchange waveguides exhibiting 4.1 dB gain in only 3 mm length [41.105]. At such high concentrations, the rare-earth constituents can hardly be called dopants as they exceed the concentration of some of the host glass constituents.

On the theme of simplified pumping options, so-called broadband sensitizers [41.106] are of great interest. A broadband sensitizer is essentially a co-dopant that is able to efficiently absorb pump energy over a wide wavelength range and subsequently excite nearby rare-earth ions to the desired lasing level. This might enable planar waveguide amplifiers that are pumped transversely using low-cost broadband light sources [41.71]. Silicon nanoclusters (embedded in oxide glass) are receiving the most attention at present. In addition to acting as a broadband sensitizer, there is evidence that these nanoclusters can enhance the stimulated emission cross-section of co-doped rare-earth ions by 1–2 orders of magnitude. *Han* et al. [41.107] have reported a waveguide amplifier (on a silicon platform) exhibiting signal enhancement of 7 dB/cm, in spite of a very low erbium concentration. The same group has also reported gain using transverse pumping by light-emitting diode (LED) arrays [41.108]. Finally, the presence of a large concentration of silicon clusters makes electrical pumping a realistic possibility for some silicon-rich oxide glasses [41.71].

It seems very likely that chip-scale glass-based lasers are on the verge of playing a central role in photonics. These lasers will deliver moderate power in a very small package, and will offer extreme spectral purity or stable, high-repetition rate pulses. They might even satisfy the long-standing desire for a silicon-based laser [41.71, 101], facilitating the convergence of photonics and electronics.

41.3.4 Integration of Dissimilar Materials Using Laser Processing

Integration of an inorganic active glass thin film with a semiconductor and a polymer requires additional control of material properties such as thermal and stress mismatch of the integrated materials. Integration of these dissimilar materials opens up a new avenue for

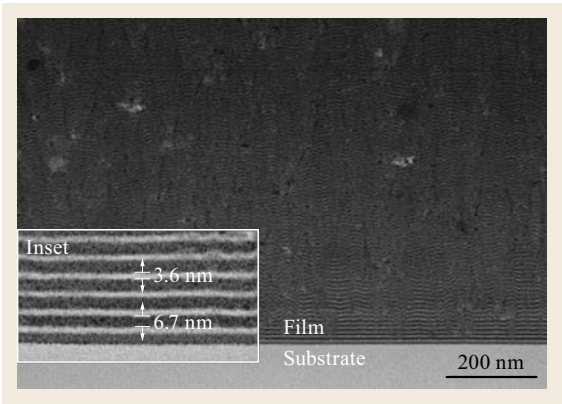


Fig. 41.9 Transmission electron microscopy (TEM) image of a cross-section of the superlattice thin film. The *inset* shows the glass (*dark*) and PDMS (*white*) interlayer thicknesses. (After [41.109], © IOP Publishing, reproduced with permission. All rights reserved)

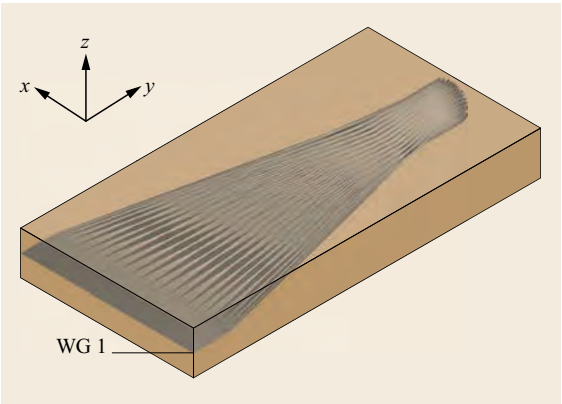


Fig. 41.10 A 3-D waveguide structure. The position of waveguide 1 (WG1) is indicated in the sketch. The waveguide number then increases sequentially along the linear array. (After [41.116], © The Optical Society)

novel photonic integrated device engineering with potential applications in optical and data communication. In photonic integrated device engineering one must consider thermal and stress mismatch as this is crucial for the structural stability of a device, which must be capable of processing signals in an active gain medium. Integration of a nanocrystalline semiconductor on an amorphous dielectric material has been demonstrated in the engineering of optoelectronic devices. Optically active or passive photonics devices, III–V semiconductor on glasses and on Si/SiO_2, and glass on an electro-optic layer such as a III–V semiconductor are some common examples of integration of dissimilar materials [41.110–114].

Recently, the integration of Er^{3+}-ion-doped phosphate-modified tellurite glass thin film with III–V semiconductor and polymer substrates has been reported [41.14, 109]. The active glass thin films were deposited using the pulsed-laser deposition technique and exhibited a refractive index close to the bulk glass target [41.115]. The optical channel waveguides fabricated using a femtosecond laser inscription method offered much smoother channel sidewalls and 2.5 dB internal gain at 1565 nm on using a 150 mW and 1480 nm pump source [41.14]. It was also reported that increasing the deposition temperature up to 400 °C during the pulsed-laser deposition process increases the photoluminescence lifetimes of the active glass thin films up to 4.3 ms, and it has no detrimental effect on the performance of the III–V semiconductor laser fabricated using this approach to integrating dissimilar materials.

Integration of organic and inorganic materials using the laser processing approach allows to control the growth parameters and optoelectronic properties of the integrated devices. For example, superlattice structures of Er^{3+}:Yb^{3+}:Ce^{2+} doped phosphate-modified tellurite glasses and polydimethyl siloxane (PDMS) polymer were successfully grown using pulsed-laser deposition at a deposition temperature of 100 °C with interlayer thicknesses as low as 7 nm for glass and 4 nm for polymer [41.115] as shown in Fig. 41.9.

The successful integration of dissimilar materials illustrates the possibility of engineering structural and thermal compatibility of organic–inorganic materials at nanometer size using the pulsed laser deposition technique. The introduced approach to integrating dissimilar materials opens up a new route to efficient and durable photonic device fabrication with potential applications in a range of optical amplifiers [41.109], and biological sensors [41.117] where integration is impossible using conventional material engineering techniques.

The limited solubility of rare-earth ions in silica glasses and silicon wafers hampers the development of low-loss photonics integrated circuits. Therefore, innovative fabrication methods are required to tackle this limitation which is related to the chemical and physical properties of the host materials and rare-earth ions. *Chandrappan* et al. [41.118] introduced a new technique based on ultrafast laser plasma-assisted hybrid multicomponent material integration of rare-earth-ion-rich tellurite and silica to enhance rare earth-ion solubility in integrated photonic silica-based components. Using this method, high multi-ion rare-earth-doping concentrations of Er^{3+}- and Yb^{3+}-ions (1.63×10^{21} atom/cm^3) were successfully achieved, without any segregation with Er^{3+}:$^4I_{13/2} \rightarrow {}^4I_{15/2}$ and transition lifetime of 9.1 ms in the silica substrate by

controlling the layer thickness of the modified silicate, which is impossible using by conventional methods.

When using conventional mix and melting glass-making methods to produce tellurite-modified silica glasses, phase separation and crystallization affinity increases due to the chemical dissimilarity of the raw materials. Recently, a novel technique based on the ultrafast laser-generated plasma-assisted process was employed for incorporation of rare-earth-ion-doped glasses into silica substrates [41.119]. By implanting the rare-earth ions into a silica glass network using the laser plasma-assisted method a metastable and homogeneous glass structure can be achieved, which is not possible using conventional methods. Increasing the rare earth-ion concentration in the bulk target glass enhances the refractive index and photoluminescence intensity of the modified silica substrate whilst the photoluminescence lifetime of the rare-earth ions decreases due to the quenching effect.

One application of the integration of dissimilar materials in the engineering of photonic components is the photonic lantern, which has potential uses in astrophysics and remote sensing. The photonics lantern is a device that efficiently couples light between a multimode optical fiber and a set of single-mode optical fibers as shown in Fig. 41.10. Ultrafast laser inscription can be used to fabricate two- and three-dimensional integrated optical waveguides that efficiently couple light from a multimode waveguide to a two- or three-dimensional array of single-mode waveguides with a loss of approximately 2 dB [41.116, 120, 121].

41.4 Summary

Glasses have numerous advantages (ease of component fabrication, low scattering loss, flexible atomic structure) that have made them a workhorse material for traditional bulk optics applications. These properties can provide equal or greater advantage in the realization of microphotonic devices and circuits. The glass transition can enable manufacturing techniques unavailable to other materials, and the realization of devices having a surface roughness approaching fundamental limits. Owing to this, glasses have enabled the highest Q microcavities and some of the lowest-loss photonic crystal and microphotonic waveguides reported to date. The metastability of glass enables a rich array of processing and postprocessing options. Direct patterning of waveguides, gratings, and other microphotonic elements by energetic beams is widely studied. These methods of material modification are also promising for postfabrication trimming of devices. Conversely, metastability has implications for processing and aging of glass devices.

Rare-earth-doped glasses offer numerous advantages relative to other solid-state laser media, especially for realization of ultra-low-noise single frequency and ultrafast lasers. Recent advances have greatly increased the maximum gain per unit length, and point to the potential for compact, on-chip glass-based light sources. Progress with respect to 2nd- and 3rd-order nonlinear effects in glasses is ongoing. It is expected that cm-scale switching or pulse-shaping devices based on glasses will become feasible, at least for niche applications.

Overall, it is clear that glasses can contribute greatly to the development of compact, low-loss, multifunction optics integrated with electronics.

Acknowledgments. The authors would like to acknowledge Prof. Animesh Jha for his support and guidance, and helpful technical discussions on the aspects of integrated optics and relevant fabrication. M.I. acknowledges the support from the RCUK Basic Technology Research grants (Ref: EP/D048672/1, EP/D04801X/1) and EPSRC Pathways to Impact (EP/I50110X/1).

References

41.1 H. Rawson: Glass and its history of service: Part A, IEE Proceedings **135**(6), 325 (1988)
41.2 W.J. Tropf, M.E. Thomas, T.J. Harris: *OSA Handbook of Optics*, Vol. II, 2nd edn. (McGraw-Hill, New York 1995)
41.3 K. Hirao, T. Mitsuyu, J. Si, J. Qiu: *Active Glass for Photonic Devices, Photoinduced Structures and Their Application* (Springer, Berlin, Heidelberg 2001)
41.4 Z. Yin, Y. Jaluria: Neck-down and thermally induced defects in high-speed optical fiber drawing, J. Heat Transfer **122**(2), 351 (2000)
41.5 W.A. Gambling: IEEE J. Sel. Top. Quantum Electron. **6**, 1084 (2000)
41.6 M. Yamane, Y. Asahara: *Glasses for Photonics* (Cambridge Univ. Press, Cambridge 2000)

41.7 P.W. France (Ed.): *Fluoride Glass Optical Fibres* (Blackie, Glassow, London 1990)
41.8 I.D. Aggarwal, G. Lu: *Fluoride Glass Fiber Optics* (Academic, Boston 2013)
41.9 A. Jha, B. Richards, G. Jose, T. Teddy-Fernandez, P. Joshi, X. Jiang, J. Lousteau: Rare-earth ion doped TeO_2 and GeO_2 glasses as laser materials, Prog. Mater. Sci. **57**(8), 1426 (2012)
41.10 K. Fujiura, T. Kanamori, S. Sudo: *Fiber Materials and Fabrications, in Optical Fiber Amplifiers–Materials, Devices and Applications* (Artech House, Norwood 1997)
41.11 H. Ma, A.K.-Y. Jen, L.R. Dalton: Adv. Mat. **14**, 1339 (2002)
41.12 M.R. Poulsen, P.I. Borel, J. Fage-Pederson, J. Hubner, M. Kristensen, J.H. Povlsen, K. Rottwitt, M. Svalgaard, W. Svendsen: Opt. Eng. **42**, 2821 (2003)
41.13 K. Okamoto: *Integrated Optical Circuits and Components, Design and Applications* (Dekker, New York 1999), Chap. 4
41.14 M. Irannejad, M. Pasha, G. Jose, P. Steenson, T.T. Fernandez, A. Jha, Q. Jiang, Z.Y. Zhang, R.A. Hogg, C. Evans, Z. Ikonic, P. Harrison: Active glass waveguide amplifier on GaAs by UV-pulsed laser deposition and femtosecond laser inscription, Laser Phys. Lett. **9**(5), 329 (2012)
41.15 J. Lucas: Curr. Opin. Solid State Mat. Sci. **4**, 181 (1999)
41.16 S. Torquato: Nature **405**, 521 (2000)
41.17 P.G. Debenedetti, F.H. Stillinger: Nature **410**, 259 (2001)
41.18 C.A. Angell, K.L. Ngai, G.B. McKenna, P.F. McMillan, S.W. Martin: J. Appl. Phys. **88**, 3113 (2000)
41.19 J.M. Saiter, M. Arnoult, J. Grenet: Phys. B: Condens. Matter **355**, 370 (2005)
41.20 A. Jha: *Inorganic Glasses for Photonics: Fundamentals, Engineering and Applications* (Wiley, New York 2016)
41.21 J.C. Anderson, K.D. Leaver, R.D. Rawlings, J.M. Alexander: *Materials Science*, 4th edn. (Chapman Hall, London 1990)
41.22 B. Hendriks, S. Kuiper: IEEE Spectrum **41**, 32 (2004)
41.23 G.P. Agrawal: *Fiber-Optic Communication Systems*, 2nd edn. (Wiley, New York 1997)
41.24 Y.P. Li, C.H. Henry: IEE Proc. Optoelectron. **143**, 263 (1996)
41.25 R.R.A. Syms, W. Huang, V.M. Schneider: Electron. Lett. **32**, 1233 (1996)
41.26 R.R.A. Syms, A.S. Holmes: IEEE Photonics Technol. Lett. **5**, 1077 (1993)
41.27 T.J. Kippenberg, S.M. Spillane, B. Min, K.J. Vahala: IEEE J. Sel. Top. Quantum Electron. **10**, 1219 (2004)
41.28 D.W. Vernooy, V.S. Ilchenko, H. Mabuchi, E.W. Streed, H.J. Kimble: Opt. Lett. **23**, 247 (1998)
41.29 M.L. Gorodetsky, A.A. Savchenkov, V.S. Ilchenko: Opt. Lett. **21**, 453 (1996)
41.30 L. Tong, R.R. Gattass, J.B. Ashcom, S. He, J. Lou, M. Shen, I. Maxwell, E. Mazur: Nature **426**, 816 (2003)
41.31 P.J. Roberts, F. Couny, H. Sabert, B.J. Mangan, D.P. Williams, L. Farr, M.W. Mason, A. Tomlinson, T.A. Birks, J.C. Knight, P.S.J. Russell: Opt. Express **13**, 236 (2005)
41.32 C.-T. Pan, C.-H. Chien, C.-C. Hsieh: Appl. Opt. **43**, 5939 (2004)
41.33 M. He, X.-C. Yuan, N.Q. Ngo, J. Bu, V. Kudryashov: Opt. Lett. **28**, 731 (2003)
41.34 M. He, X.-C. Yuan, J. Bu: Opt. Lett. **29**, 2004 (2004)
41.35 Y.A. Vlasov, S.J. McNab: Opt. Express **12**, 1622 (2004)
41.36 F. Grillot, L. Vivien, S. Laval, D. Pascal, E. Cassan: IEEE Photonics Technol. Lett. **16**, 1661 (2004)
41.37 T. Barwicz, H.I. Smith: J. Vaccum Sci. Technol. B **21**, 2892 (2003)
41.38 V. Van, P.P. Absil, J.V. Hryniewicz, P.-T. Ho: J. Light. Technol. **19**, 1734 (2001)
41.39 K.J. Vahala: Optical microcavities, Nature **424**, 839 (2003)
41.40 S.J. McNab, N. Moll, Y.A. Vlasov: Opt. Express **11**, 2927 (2003)
41.41 P.K. Gupta, D. Inniss, C.R. Kurkijian, Q. Zhong: J. Non-Crystal. Solids **262**, 200 (2000)
41.42 D.P. Bulla, W.-T. Li, C. Charles, R. Boswell, A. Ankiewicz, J. Love: Appl. Opt. **43**, 2978 (2004)
41.43 D.K. Armani, T.J. Kippenberg, S.M. Spillane, K.J. Vahala: Nature **421**, 925 (2003)
41.44 X.H. Zhang, Y. Guimond, Y. Bellec: J. Non-crystal. Solids **326**, 519 (2003)
41.45 X.H. Zhang, Y. Guimond, Y. Bellec: J. Non-crystal. Solids **327**, 519 (2003)
41.46 A.K. Mairaj, X. Feng, D.P. Shepherd, D.W. Hewak: Appl. Phys. Lett. **85**, 2727 (2004)
41.47 A.K. Mairaj, R.J. Curry, D.W. Hewak: Appl. Phys. Lett. **86**, 094102 (2005)
41.48 J.S. Sanghera, L.B. Shaw, I.D. Aggarwal: Rare earth doped infrared-transmitting glass fibers. In: *Rare-Earth-Doped Fiber Lasers and Amplifiers, 2nd edn*, ed. by H.J.F. Digonnet (Dekker, New York 2001) p. 449
41.49 P.K. Tien: Appl. Opt. **10**, 2395 (1971)
41.50 R. Ulrich: J. Vaccum Sci. Technol. **11**, 156 (1974)
41.51 P.K. Tien, A.A. Ballman: J. Vaccum Sci. Technol. **12**, 892 (1974)
41.52 J.M. Mir, J.A. Agostinelli: J. Vaccum Sci. Technol. A **12**, 1439 (1994)
41.53 J. Mueller, M. Mahnke, G. Schoer, S. Wiechmann: AIP Conf. Proc. **709**, 268 (2004)
41.54 Z. Zhang, G. Xiao, C.P. Grover: Appl. Opt. **43**, 2325 (2004)
41.55 D. Cangialosi, M. Wubbenhorst, H. Schut, A. van Veen, S.J. Picken: Phys. Rev. B **69**, 134206 (2004)
41.56 M.B.J. Diemeer: AIP Conf. Proc. **709**, 252 (2004)
41.57 N. Daldosso, M. Melchiorri, F. Riboli, F. Sbrana, L. Pavesi, G. Pucker, C. Kompocholis, M. Crivellari, P. Bellutti, A. Lui: Mat. Sci. Semicond. Proc. **7**, 453 (2004)
41.58 A.V. Kolobov (Ed.): *Photo-Induced Metastability in Amorphous Semiconductors* (Wiley-VCH, Weinheim 2003)
41.59 K. Tanaka: C.R. Chimie **5**, 805 (2002)
41.60 H. Haeiwa, T. Naganawa, Y. Kokubun: IEEE Photonics Technol. Lett. **16**, 135 (2004)

41.61 K.O. Hill: IEEE J. Sel. Top. Quantum Electron. **6**, 1186 (2000)
41.62 P.R. Herman, R.S. Marjoribanks, A. Oettl, K. Chen, I. Konovalov, S. Ness: Appl. Surf. Sci. **154–155**, 577 (2000)
41.63 M. Aslund, J. Canning: Opt. Lett. **25**, 692 (2000)
41.64 K. Minoshima, A.M. Kowalevicz, E.P. Ippen, J.G. Fujimoto: Opt. Express **10**, 645 (2002)
41.65 N. Ponnampalam, R.G. DeCorby, H.T. Nguyen, P.K. Dwivedi, C.J. Haugen, J.N. McMullin, S.O. Kasap: Opt. Express **12**, 6270 (2004)
41.66 C. Florea, K.A. Winick: J. Light. Technol. **21**, 246 (2003)
41.67 A. Zakery, Y. Ruan, A.V. Rode, M. Samoc, B. Luther-Davies: J. Opt. Soc. Am. B. **20**, 1844 (2003)
41.68 X. Orignac, D. Barbier, X. Du Min, R.M. Almeida, O. McCarthy, E. Yeatman: Opt. Mat. **12**, 1 (1999)
41.69 R.A. Bellman, G. Bourdon, G. Alibert, A. Beguin, E. Guiot, L.B. Simpson, P. Lehuede, L. Guiziou, E. LeGuen: J. Electrochem. Soc. **151**, 541 (2004)
41.70 G.-L. Bona, R. Germann, B.J. Offrein: IBM J. Res. Dev. **47**, 239 (2003)
41.71 A.J. Kenyon: Prog. Quantum Electron. **26**, 225 (2002)
41.72 M.J. Miniscalco: Optical and electronic properties of rare earth jons in glasses. In: *Rare-Earth-Doped Fiber Lasers and Amplifiers, 2nd edn*, ed. by M.J.F. Digonnet (Dekker, New York 2001) p. 17
41.73 Y. Gao, B. Boulard, M. Couchaud, I. Vasilief, S. Guy, C. Duverger, B. Jacquier: Opt. Mat. **27**, 195 (2005)
41.74 T. Barwicz, M.A. Popovic, P.T. Rakich, M.R. Watts, H.A. Haus, E.P. Ippen, H.I. Smith: Opt. Express **12**, 1437 (2004)
41.75 R. Rabady, I. Avrutsky: Appl. Opt. **44**, 378 (2005)
41.76 H. Hirota, M. Itoh, M. Oguma, Y. Hibino: IEEE Photonics Technol. Lett. **17**, 375 (2005)
41.77 C.-Y. Tai, J.S. Wilkinson, N.M.B. Perney, M. Caterina Netti, F. Cattaneo, C.E. Finlayson, J.J. Baumberg: Opt. Express **12**, 5110 (2004)
41.78 B. Unal, C.-Y. Tai, D.P. Shepherd, J.S. Wilkinson, N.M.B. Perney, M. Caterina Netti, G.J. Parker: Nd:Ta_2O_5 rib waveguide lasers, Appl. Phys. Lett. **86**, 021110 (2005)
41.79 Y. Kokubun, Y. Hatakeyama, M. Ogata, S. Suzuki, N. Zaizen: IEEE J. Sel. Top. Quantum Electron. **11**, 4 (2005)
41.80 T. Kawashima, K. Miura, T. Sato, S. Kawakami: Appl. Phys. Lett. **77**, 2613 (2000)
41.81 T. Sato, K. Miura, N. Ishino, Y. Ohtera, T. Tamamura, S. Kawakami: Opt. Quantum Electron. **34**, 63 (2002)
41.82 D.R. MacFarlane: Ceramics Int. **22**, 535 (1996)
41.83 A. Jha, B.D.O. Richards, G. Jose, T.T. Fernandez, C.J. Hill, J. Lousteau, P. Joshi: Review on structural, thermal, optical and spectroscopic properties of tellurium oxide based glasses for fibre optic and waveguide applications, Int. Mater. Rev. **57**(6), 357 (2012)
41.84 M. Irannejad: Fabrication and Characterization of Glass-Semiconductor Integrated Waveguide Amplifier by Pulsed Laser Deposition, Ph.D. Thesis (University of Leeds, Leeds 2012)
41.85 S. Hocde, S. Jiang, X. Peng, N. Peyghambarian, T. Luo, M. Morrell: Opt. Mat. **25**, 149 (2004)
41.86 R. Nayak, V. Gupta, A.L. Dawar, K. Sreenivas: Thin Solid Films **445**, 118 (2003)
41.87 A. Feigel, M. Veinger, B. Sfez, A. Arsh, M. Klebanov, V. Lyubin: Appl. Phys. Lett. **83**, 4480 (2003)
41.88 G. Lenz, S. Spalter: *Nonlinear Photonic Crystals* (Springer, Berlin, Heidelberg 2003)
41.89 K. Ogusu, J. Yamasaki, S. Maeda, M. Kitao, M. Minakata: Opt. Lett. **29**, 265 (2004)
41.90 D. Barbier, R.L. Hyde: Eibium doped glass waveguide devices. In: *Integrated Optical Circuits and Components, Design and Applications*, ed. by E.J. Murphy (Dekker, New York 1999) p. 83
41.91 B.E. Callicoatt, J.B. Schlager, R.K. Hickernell, R.P. Mirin, N.A. Sanford: IEEE Circuits Devices Mag. **19**, 18 (2003)
41.92 E. DeSurvire: Erbium-doped fiber amblifies: Basic physics and characteristics. In: *Rare-Earth-Doped Fiber Lasers and Amplifiers, 2 edn*, ed. by M.J.F. Digonnet (Dekker, New York 2001) p. 531
41.93 A. Yariv: *Optical Electronics in Modern Communications*, 5th edn. (Oxford Univ. Press, New York 1997)
41.94 E. Desurvire: *Erbium Doped Fiber Amplifiers, Principles and Applications* (Wiley, Hoboken 2002)
41.95 Q. Wang, N.K. Dutta: J. Appl. Phys. **95**, 4025 (2004)
41.96 M.J.F. Digonnet: Continious-wave silica fiber lasers. In: *Rare-Earth-Doped Fiber Lasers and Amplifiers, 2nd edn*, ed. by M.J.F. Digonnet (Dekker, New York 2001) p. 113
41.97 D.S. Funk, J.G. Eden: Visible fluride fiber lasers. In: *Rare-Earth-Doped Fiber Lasers and Amplifiers, 2nd edn*, ed. by M.J.F. Digonnet (Dekker, New York 2001) p. 171
41.98 S. Saini, J. Michel, L.C. Kimerling: J. Light. Technol. **21**, 2368 (2003)
41.99 P.G. Kik, A. Polman: J. Appl. Phys. **93**, 5008 (2003)
41.100 P. Laporta, S. Taccheo, S. Longhi, O. Svelto, C. Svelto: Opt. Mat. **11**, 269 (1999)
41.101 F.J. Grawert, J.T. Gopinath, F.O. Ilday, H.M. Shen, E.P. Ippen, F.X. Kartner, S. Akiyama, J. Liu, K. Wada, L.C. Kimerling: Opt. Lett. **30**, 329 (2005)
41.102 U. Keller: Nature **424**, 831 (2003)
41.103 T. Qiu, L. Li, A. Schulzgen, V.L. Temyanko, T. Luo, S. Jiang, A. Mafey, J.V. Moloney, N. Peyghambarian: IEEE Photonics Technol. Lett. **16**, 2592 (2004)
41.104 Y.C. Yan, A.J. Faber, H. de Waal, P.G. Kik, A. Polman: Appl. Phys. Lett. **71**, 2922 (1997)
41.105 F.D. Patel, S. DiCarolis, P. Lum, S. Venkatesh, J.N. Miller: IEEE Photonics Technol. Lett. **16**, 2607 (2004)
41.106 A. Polman, F.C.J.M. van Veggel: J. Opt. Soc. Am. B **21**, 871 (2004)
41.107 H.-S. Han, S.-Y. Seo, J.H. Shin, N. Park: Appl. Phys. Lett. **81**, 3720 (2002)
41.108 J. Lee, J.H. Shin, N. Park: J. Light. Technol. **23**, 19 (2005)
41.109 Z. Zhao, G. Jose, T.T. Fernandez, T.P. Comyn, M. Irannejad, P. Steenson, J.P. Harrington, M. Ward, N. Bamiedakis, R.V. Penty: Active glass–polymer superlattice structure for photonic inte-

gration, Nanotechnology **23**(22), 225302 (2012)
41.110 M. Madsen, K. Takei, R. Kapadia, H. Fang, H. Ko, T. Takahashi, A.C. Ford, M.H. Lee, A. Javey: Nanoscale semiconductor X on substrate Y – Processes, devices, and applications, Adv. Mater. **23**(28), 3115 (2011)
41.111 G.A. Antypas, J. Edgecumbe: Glass-sealed GaAs-AlGaAs transmission photocathode, Appl. Phys. Lett. **26**(7), 371 (1975)
41.112 M. Konagai, M. Sugimoto, K. Takahashi: High efficiency GaAs thin film solar cells by peeled film technology, J. Crystal Growth **45**, 277 (1978)
41.113 E. Yablonovitch, T. Gmitter, J.P. Harbison, R. Bhat: Extreme selectivity in the lift-off of epitaxial GaAs films, Appl. Phys. Lett. **51**(26), 2222 (1987)
41.114 H. Ko, K. Takei, R. Kapadia, S. Chuang, H. Fang, P.W. Leu, K. Ganapathi, E. Plis, H.S. Kim, S.-Y. Chen, M. Madsen, A.C. Ford, Y.-L. Chueh, S. Krishna, S. Salahuddin, A. Javey: Ultrathin compound semiconductor on insulator layers for high-performance nanoscale transistors, Nature **468**(7321), 286 (2010)
41.115 M. Irannejad, G. Jose, P. Steenson, A. Jha: Enhancement in optical and microstructure properties of Er^{3+}-doped phospho-tellurite glass thin film, Opt. Mater. **34**(8), 1272 (2012)
41.116 R.R. Thomson, R.J. Harris, T.A. Birks, G. Brown, J. Allington-Smith, J. Bland-Hawthorn: Ultrafast laser inscription of a 121-waveguide fan-out for astrophotonics, Opt. Lett. **37**(12), 2331 (2012)
41.117 T. Kakkar, N. Bamiedakis, T.T. Fernandez, Z. Zhao, M. Irannejad, P. Steenson, A. Jha, R. Penty, I. White, G. Jose: Glass–polymer superlattice for integrated optics, Opt. Eng. **53**(7), 071818 (2014)
41.118 J. Chandrappan, M. Murray, P. Petrik, E. Agocs, Z. Zolnai, A. Tempez, S. Legendre, D.P. Steenson, A. Jha, G. Jose: Doping silica beyond limits with laser plasma for active photonic materials, Opt. Mater. Express **5**(12), 2849 (2015)
41.119 J. Chandrappan, M. Murray, T. Kakkar, P. Petrik, E. Agocs, Z. Zolnai, D.P. Steenson, A. Jha, G. Jose: Target dependent femtosecond laser plasma implantation dynamics in enabling silica for high density erbium doping, Sci. Rep. **5**, 14037 (2015)
41.120 A. Ródenas, G. Martin, B. Arezki, N. Psaila, G. Jose, A. Jha, L. Labadie, P. Kern, A. Kar, R. Thomson: Three-dimensional mid-infrared photonic circuits in chalcogenide glass, Opt. Lett. **37**(3), 392 (2012)
41.121 R.R. Thomson, T.A. Birks, S.G. Leon-Saval, A.K. Kar, J. Bland-Hawthorn: Ultrafast laser inscription of an integrated photonic lantern, Opt. Express **19**(6), 5698 (2011)

42. Optical Nonlinearity in Photonic Glasses

Keiji Tanaka

A brief review of optical nonlinearity in photonic glasses is given. For third-order nonlinearity, the relationship between two-photon absorption and nonlinear refractive index is considered using a formalism developed for crystalline semiconductors. Stimulated light scattering and supercontinuum generation in optical fibers and waveguides are introduced. Prominent resonant-type nonlinearity in particle-embedded glasses is described. For second-order nonlinearity, a variety of poling methods are summarized. The nonlinearity also appears in crystallized and ceramic glasses. Finally, it is pointed out that various photoinduced changes can appear when excited by linear and nonlinear optical processes.

42.1 **Third-Order Nonlinearity in Homogeneous Glass** 1083
42.1.1 Experimental 1083
42.1.2 Theoretical Treatment 1084
42.1.3 Stimulated Light Scattering, Supercontinuum Generation, and Other Applications 1087
42.2 **Second-Order Nonlinearity in Poled Glass** 1088
42.3 **Particle-Embedded Systems** 1089
42.4 **Photoinduced Phenomena** 1090
42.5 **Summary** 1091
References 1092

Developments in optical fibers, integrated waveguides, and pulsed lasers have prompted increasing interest in optical nonlinearity in photonic glasses [42.1–3]. The third-order polarization yields several nonlinear phenomena, such as intensity-dependent absorption and refractive index, which can be utilized in power stabilizers, supercontinuum generation, all-optical switches, and so forth. The main target here is to attain high nonlinearity with fast response. However, in some cases, as mentioned later, the nonlinearity may be a nuisance. On the other hand, the second-order polarization that occurs in poled and crystallized glasses can be utilized in second-harmonic generation, for example. This chapter provides a brief review of optical nonlinearity in *inorganic* glasses.

This chapter begins with an overview of research trends (at the beginning of 2015). Figure 42.1 shows a yearly number of published papers, analyzed using Web of Science, having the topical words *nonlinear, optical, and glass*.

In the total number amounting to ≈ 4000, we see an abrupt rise in 1991 followed by an overall increase with a peak in 2007–2008. What led to these rises? The rise in 1991 seems to be triggered by fundamental interests. Actually, as shown in the statistics, the rise becomes clearer with additional words such as CdS. At the time, we actively examined fundamental properties of commercially available, semiconductor-doped, color-glass filters using pulsed YAG lasers. Such studies have continued to date, though the dopant has been gradually replaced by metallic particles such as Ag, Cu, Au, etc., which are capable of providing greater nonlinearity in combination with plasmonic effects. On the other hand, the origin of the 2007–2008 peak is vague. We see that the additional words such as *photonic crystal*, *Raman*, *ceramic*, *crystalized*, *poling*, or *all-optical* also result in the rise. Probably, many scientific experiments using Ti-sapphire lasers flourished around that time.

The statistics shed light on another noteworthy aspect. As known, there are three kinds of optical glasses namely, oxide, chalcogenide, and halide, of which oxide is the most commonly used one. However, we note that *oxide* and *chalcogenide* papers are comparable, giving the total numbers of ≈ 400, with much fewer being halide related. This relatively intense interest in chalcogenide glasses rests on the higher nonlinearity (Table 42.1), which promises wide applications, as reviewed in recent articles [42.4–6].

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_42

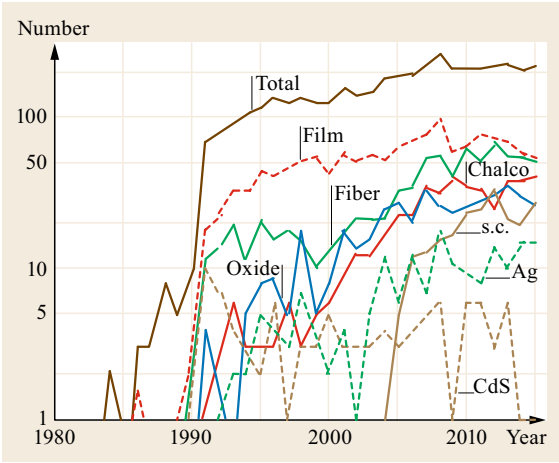


Fig. 42.1 Yearly numbers of published papers having topical words *nonlinear* + *optical* + *glass* (total), and those with a word added; *film*, *fiber*, *oxide*, *chalcogenide* (Chalco), *supercontinuum* (s.c.), *Ag* and *CdS* (analyzed using Web of Science)

Finally, what are the growing topics at the present time? As shown in Fig. 42.1, an additional word *supercontinuum* has been manifesting a rapid growth since ≈ 2005, which might contribute to commercializing supercontinuum fibers. *Photonic crystal fiber* shows a similar trend. It is also mentioned that *integrated*, which appeared in 1992 and fluctuated around 6 [42.7], rapidly rose to 14 in 2014. The topic includes, for example, proposals of polymer-glass hybrid systems [42.8] and also integrated devices [42.5, 9], which may be promising henceforth.

At this point we should note that, in many respects, the glass nonlinearity is competitive with that in organic polymers and crystals [42.1, 2]. Compared to polymers, glass is more stable, while polymers can provide greater nonlinearity with molecular designs. Therefore, these two types of materials are competitors in practical applications. Compared to crystals, the glass nonlinearity, specifically that of the second order, may be smaller than that in the crystal. However, a great advantage of glass is its ability to control its form at three scales. First, the atomic composition of the glass can be tailored continuously. For instance, we can obtain nonlinear optical glass with any refractive index in the range of 1.4–3.2 at a wavelength of $\approx 1\,\mu$m. Second, the atomic structure of the glass can be modified reasonably easily using, for example, light beams. Such modifications may be regarded as photoinduced phenomena, which can be employed in order to add second-order nonlinearity to selected regions and so forth. Finally, macroscopic forms can be changed into arbitrary bulk shapes, fibers [42.3], thin layers, channels, and micro-particles [42.10, 11].

Here, fibers and waveguides are practically important for nonlinear applications for two reasons. One is that the light power density can be increased by reducing the lateral size, for example, the fiber diameter or waveguide dimensions, to wavelength scales [42.12]. In this respect, nanowires [42.13], which are the waveguide with submicron cross sections, may be more promising. The other is fibers can provide long lengths for light–matter interactions that are not limited by diffraction. These scale factors produce apparent enhancements in nonlinear effects in glasses, although the intrinsic nonlinearity may be smaller than that in crystals. In addition, it should be emphasized that structured (holey, photonic crystal) fibers [42.14] are able to control dispersion characteristics, which can tune desired wavelengths. Needless to say, the accuracy in shaping fibers and waveguides, which are fabricated by several methods [42.3, 15, 16], is a key issue for suppressing the light-scattering loss.

Before proceeding further, it would be useful to introduce a nonlinear formula [42.1, 2]. For simplicity, we take the polarization P and the electric field E to be scalar quantities. Then, very simply, P can be written down in the CGS unit as

$$P = \chi^{(1)}E + \chi^{(2)}EE + \chi^{(3)}EEE + \dots , \qquad (42.1)$$

where the first term $\chi^{(1)}E$ depicts the conventional linear response, and $\chi^{(2)}$ and $\chi^{(3)}$ represent the second- and the third-order nonlinear susceptibilities, respectively. $\chi^{(1)}$ is related to the linear refractive index n_0 via $n_0 = (1+4\pi\chi^{(1)})^{1/2}$. On the other hand, the second term provides such time dependence as $\sin(\omega \pm \omega')t$, so that it could produce second-overtone (2ω) and dc signals. In a similar way, the third could modify the fundamental (ω) signal and generate a third-overtone (3ω). Nevertheless, since a refractive index changes with light frequency, it is not straightforward in the overtone generations to satisfy the so-called phase-matching condition, for example, $\varphi_{3\omega} = 3\varphi_{\omega}$, where φ is the phase of electric fields in glasses. Microscopically, the nonlinear terms arise through several mechanisms, such as electronic, atomic (including molecular motions), electrostatic, and thermal processes. Among these, electronic processes can provide the fastest responses, with fs–ps time scales, which will be needed for optical information technologies. Accordingly, we will focus on such processes from this point onward.

The chapter is divided as follows. In Sect. 42.1, we treat conventional glasses. These are optically isotropic, which gives $\chi^{(2)} = 0$, because the isotropic structure

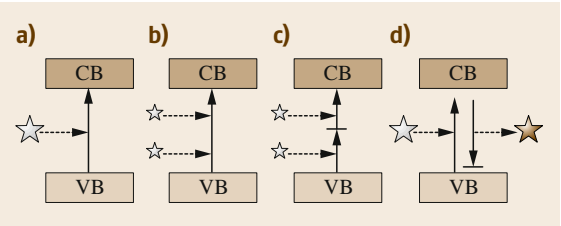


Fig. 42.2a–d Schematic illustrations of **(a)** one-photon absorption, **(b)** two-photon absorption, **(c)** two-step absorption via a mid-gap state, and **(d)** Raman-scattering process in a semiconductor with a valence band (VB) and a conduction band (CB)

appears to be centrosymmetric. In this case, the lowest nonlinear term becomes $\chi^{(3)}EEE$. When $\chi^{(3)}EEE$ can be written as $\chi^{(3)}IE$, where I is the light intensity, we can define the refractive index n as $n = n_0 + n_2 I$, where n_2 is sometimes called the second-order index of refraction [42.2]. Incidentally, in nonlinear optics, the notation poses perplexing problems due to several definitions of parameters. And, the conversion $n_2\,(\mathrm{cm}^2/\mathrm{W}) \approx 0.04\chi^{(3)}(\mathrm{esu})/n_0^2$ is frequently employed [42.2]. On the other hand, for light absorption, which is proportional to $\langle E\mathrm{d}P/\mathrm{d}t\rangle$, where $\langle\cdots\rangle$ represents a time average, we need to take the imaginary parts of $\chi^{(1)}$ and $\chi^{(3)}$, which cause one- and two-photon absorption, into account (Fig. 42.2). In such cases, an effective absorption coefficient can be written as $\alpha\,(\mathrm{cm}^{-1}) + \beta\,(\mathrm{cm/W})I\,(\mathrm{W/cm}^2)$. n_2 and β will be connected through nonlinear Kramers–Kronig relations.

Section 42.2 focuses on noncentrosymmetric systems. From the mid-1980s onward, several kinds of *poling* treatments have been found to add $\chi^{(2)}$ to glasses [42.17]. For instance, *Österberg* and *Margulis* [42.18] demonstrated intense second-harmonic generation in laser-irradiated glass fibers. Such work began with silica, and has been directed toward more complicated glasses. The magnitude of $\chi^{(2)}$ obtained by poling is likely to be smaller than those observed in crystals, due to disordered atomic structures in glasses, but the structural controllability such as spatially selected crystallization may offer some advantages. Section 42.3 introduces glasses incorporating fine semiconductor and metal particles. Such nanostructured glasses are known to exhibit large optical nonlinearities [42.1, 2, 19]. However, as will be described, their features remain to be studied. For ion-doped glass systems, including fiber lasers and amplifiers, readers may refer to Chaps. 40 and 43 and a recent review [42.20]. In Sect. 42.4, we consider photoinduced phenomena. Intense light needed to produce optical nonlinearity is likely to modify the structure of glass. Light that is even more intense may cause damage. While the damaging can be useful for optical engraving, the structural modification may be more important. Actually, commercial fiber Bragg gratings [42.3] have been produced using excimer lasers. Nonlinear optical excitation possibly plays there an important role [42.21]. Also, three-dimensional waveguides are produced through controlled light focusing [42.16, 22]. Section 42.5 provides a summary of the chapter. Parts of this chapter are based on the author's review [42.23].

42.1 Third-Order Nonlinearity in Homogeneous Glass

42.1.1 Experimental

Substantial data are available for n_2 at transparent wavelengths [42.1, 2]. However, to the author's knowledge, almost all experiments have utilized lasers, and spectral dependence has rarely been reported. We should also note that, in comparison to n_0 measurements, n_2 evaluations are much more difficult. The biggest difficulty, which is common to all nonlinear experiments, is the light intensity. Normally, light is pulsed and focused, producing complex temporal and spatial profiles, which in turn means that intensity evaluations are not straightforward. For instance, if the profile is Gaussian, can we simply use the peak value to evaluate the nonlinearity? Because of such evaluation difficulties and also some other characteristics inherent to glasses, including thermodynamical quasi-stability, reported n_2 values vary, even for a reference SiO_2 glass, over the range of 1×10^{-7}–$3\times10^{-7}\,\mathrm{cm}^2/\mathrm{GW}$. In other glasses, values that differ by a factor of ≈ 10 have been reported [42.1, 24–26].

However, we can see a general trend in n_2 in Table 42.1. Halide glasses (BeF_2) have the smallest values, light-metal oxides (BK-7) and SiO_2 have similar values, heavy-metal oxides such as PbO-SiO_2 (SF-59) have larger values [42.27], while the largest belong to the chalcogenides (As_2S_3) [42.1]. The ultimate n_2 value reported so far may be $8\times10^{-3}\,\mathrm{cm}^2/\mathrm{GW}$, seen in $Ag_{20}As_{32}Se_{48}$ at wavelength of 1.05 μm [42.28], which is $\approx 10^5$ times as large as that of SiO_2.

Figure 42.3 shows β spectra for SiO_2 [42.29, 30], As_2S_3, and two PbO-SiO_2 glasses [42.27]. Bi_2O_3-B_2O_3 glasses [42.31, 32] exhibit similar features to those of lead glass. We can see that all the β spectra seem to

Table 42.1 Linear (E_g, n_0, α_0) and nonlinear (n_2, β, β_{max}) optical properties and figures of merit ($n_2/2\beta\lambda_0$, n_2/α_0) in some glasses. E_g is the optical bandgap energy [42.22], n_0 is the refractive index [42.22], α_0 is the attenuation coefficient [42.27], n_2 is the intensity-dependent refractive index [42.1, 2, 22], β is the two-photon absorption coefficient [42.27], and β_{max} is the maximum value. Except β_{max}, the values are evaluated at a wavelength of 1−1.5 μm. BK-7 is a borosilicate glass and SF-59 indicates data for lead-silicate glasses with ≈ 57 mol % PbO

	E_g (eV)	n_0	α_0 (cm^{-1})	n_2 (10^{-7} cm^2/GW)	β (cm/GW)	β_{max} (cm/GW)	$n_2/(2\beta\lambda_0)$	n_2/α_0 (cm^3/GW)
SiO_2	10	1.5	10^{-6}	2	$< 10^{-2}$	1	< 0.1	0.2
BK-7	4	1.5		3				
SF-59	3.8	2.0		30	$< 10^{-1}$	10		
As_2S_3	2.4	2.5	10^{-3}	200	10^{-2}	50	10	0.02
BeF_2	10	1.3		0.8				

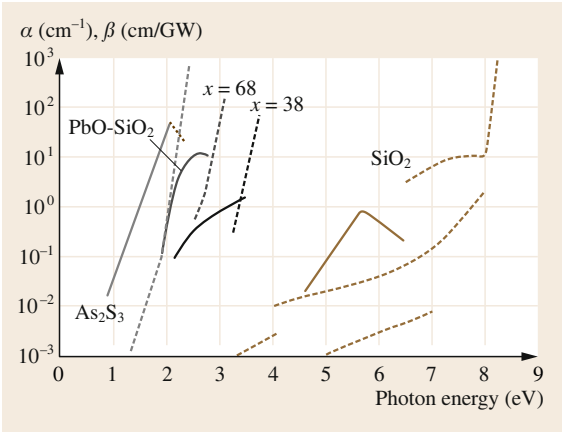


Fig. 42.3 One- (*dashed line*) and two-photon (*solid line*) absorption spectra, α and β, in SiO_2 [42.29, 30], 38PbO-62SiO_2, 68PbO-32SiO_2 [42.27], and As_2S_3 [42.27]. For SiO_2, in transparent regions ($\hbar\omega \leq 8.0$ eV) α should be regarded as an attenuation coefficient; these are not reproducible, probably due to impurities, defects, and light scattering. The *dotted line* for As_2S_3 at $\hbar\omega \geq 2.0$ eV indicates two-step absorption

have maxima at mid-gap regions, $\hbar\omega \approx E_g/2$. However, this is not an absolute rule, since at $\hbar\omega \geq 2.0$ eV As_2S_3 provides two-step absorption (Fig. 42.2c), which tends to mask two-photon absorption signals [42.33]. We also see that the maximal β ($\approx 10^0$ cm/GW) in SiO_2 is markedly smaller than those belonging to the other glasses. Note that such a marked difference does not exist in maximal α, which is $\approx 10^6$ cm^{-1} at the super-bandgap regions of SiO_2 and As_2S_3. It may be worth mentioning here that two-photon excitation produces photocurrents in some glasses [42.34, 35].

42.1.2 Theoretical Treatment

For $\chi^{(3)}$, theoretical treatments of glasses, liquids, and crystals are largely the same. Glasses and liquids can be treated similarly in the sense that both are optically isotropic. Hence, many ideas stemming from types of chemical bonds have been proposed [42.2]. On the other hand, the magnitude of $\chi^{(3)}$ in glass appears to be similar to that in the corresponding crystal, since electronic structures of both glass and crystal governed by short-range atomic structures [42.36], within ≈ 0.5 nm, which are much shorter than the wavelength (≈ 500 nm) of visible light. Specifically, since n_2 is governed by integrated electronic absorption [42.17], the magnitudes of glass and crystal are roughly the same. For instance, in crystalline and glassy SiO_2 at near-infrared wavelengths, the difference in the n_2 values of both seems to be comparable with their experimental accuracy; that is, the values for crystalline and glassy SiO_2 are $\approx 1.14 \times 10^{-13}$ and 0.85×10^{-13} esu, respectively [42.1]. Then, using a formula derived for semiconductor crystals, we can apply a kind of band picture to glasses [42.27], as described below.

Estimation of Nonlinear Refractivity

Several semiempirical relationships have been proposed for $\chi^{(3)}$ or n_2 [42.1]. Such relations are useful, since nonlinear optical constants are much more difficult to evaluate than linear constants. Naturally, since they are simplifications they are also less accurate. The simplest may be the one provided by Wang, $\chi^{(3)} \approx \{\chi^{(1)}\}^4$, which can be regarded as a generalized Miller's rule [42.1]. *Boling* et al. [42.37] have also derived some relations, among which the simplest may be

$$n_2\left(10^{-13}\,\text{esu}\right) \approx 391\frac{(n_d - 1)}{\nu_d^{5/4}}\,, \qquad (42.2)$$

where n_d is the refractive index at the d-line ($\lambda = 588$ nm) and ν_d is the Abbe number, which is defined as $(n_d - 1)/(n_F - n_C)$, where n_F and n_C are the indices at $\lambda = 486$ and 656 nm, respectively. This relation contains only two macroscopic parameters, which can easily be obtained, and so it has often been used to

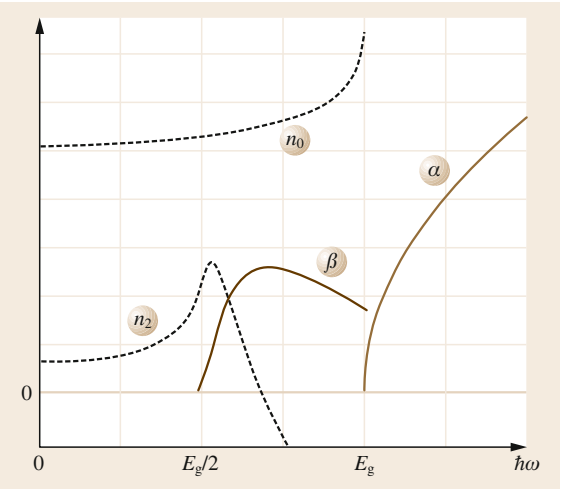


Fig. 42.4 Spectral dependence of linear absorption α, linear refractive index n_0, two-photon absorption β, and intensity-dependent refractive index n_2 in an ideal amorphous semiconductor with energy gap E_g. *Vertical scales* are arbitrary

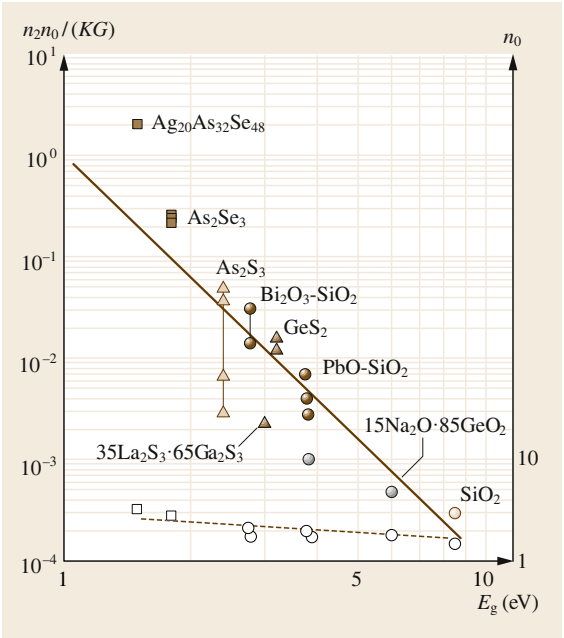


Fig. 42.5 The Sheik-Bahae relation $n_2 n_0 = KG/E_g^4$ (*solid line*), related data (*solid symbols*), the Moss rule $n_0^4 E_g = 77$ (*dashed line*), and related data (*open symbols*) for an oxide (*circles*), sulfide (*triangles*), and selenide (*squares*). The four data for As_2S_3 are obtained from different publications, while those for PbO-SiO_2 with slightly different E_g correspond to different compositions. The illustration is modified from its original form [42.27] due to the additions of $Ag_{20}As_{32}Se_{48}$ [42.28], Bi_2O_3-silica glasses [42.43], $35La_2S_3$-$65Ga_2S_3$ [42.44], and $15Na_2O$-$85GeO_2$ [42.45]

estimate n_2. It actually provides a good approximation for small n_d glasses with $n_d \leq 1.7$ [42.37]. For high n_d glasses, *Lines* [42.24] proposes a relation that contains the atomic distance, and others derive a more complicated formula [42.38, 39]. Ab initio calculations of the nonlinear refractivity in TeO_2-based glasses have also been reported [42.40]. However, these relations cannot predict wavelength dependence (Fig. 42.4), and hence, the estimated n_2 may be regarded as a long-wavelength limit. In addition, the material is regarded as transparent, or an insulator, so that relations between n_2 and β cannot be known.

Nonlinear optical properties of semiconductor crystals have been studied fairly deeply [42.2], and applying the same concepts to glasses appears to be tempting. *Tanaka* [42.27] has adopted a band picture, which was developed by *Sheik-Bahae* et al. [42.41], to glasses. Their formulation gives a universal relationship of

$$n_2 = \frac{KG\left(\hbar\omega/E_g\right)}{n_0 E_g^4}\,, \tag{42.3}$$

for crystals with energy gaps of 1–10 eV, where K is a fixed constant and $G(\hbar\omega/E_g)$ is a spectral function. For the bandgap energy E_g in glasses, we can take the so-called Tauc gap [42.36] if it is known, or otherwise the photon energy $\hbar\omega$ at $\alpha \approx 10^3\,\mathrm{cm}^{-1}$. As shown in Fig. 42.5, the universal line gives reasonable agreement with published data, while the agreement is not so satisfactory as it is for crystals [42.41]. The poorer agreement for glasses may be partly due to experimental difficulties. In addition, band tail states probably cause larger deviations in smaller bandgap glasses such as $Ag_{20}As_{32}Se_{48}$. Incidentally, as shown by the dashed line in the figure, the Moss rule $n_0^4 E_g = 77$ [42.42], which was originally pointed out for crystalline semiconductors, also gives satisfactory fits for glasses.

Nonlinear Kramers–Kronig Relationship

We can take another approach in order to obtain unified insight into optical nonlinearity, β and n_2. In this case, we start with absorption and then obtain the refractive index using Kramers–Kronig relationships (KKR). As illustrated in Fig. 42.6, this approach is able to provide a complete understanding of atomic structure and optical properties, since absorption spectra can be related to electronic structure more directly.

It is well known that, for glasses, and neglecting momentum conservation rules, $\alpha(\hbar\omega)$ and $\beta(\hbar\omega)$ can

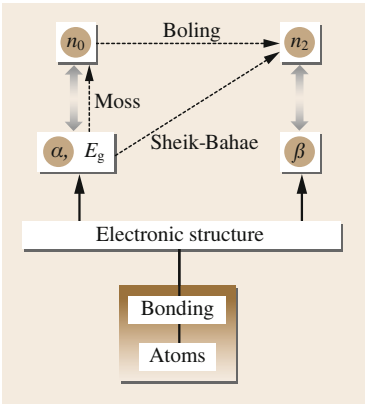


Fig. 42.6 Relationship between atomic structure, electronic structure, linear absorption α including bandgap energy E_g, linear refractive index n_0, two-photon absorption β, and nonlinear refractive index n_2. Absorption and refraction are related by the linear and nonlinear KKR (*double arrows*). The Moss rule may be regarded as a simplified KKR. The Boling and Sheik-Bahae relations connect n_0 and E_g to n_2, respectively

be written as [42.23]

$$\alpha(\hbar\omega) \propto |\langle\phi_f|H|\phi_i\rangle|^2 \int D_f(E+\hbar\omega)\, D_i(E)\, dE\,, \tag{42.4}$$

$$\beta(\hbar\omega) \propto \left|\frac{\Sigma_n \langle\phi_f|H|\phi_n\rangle\langle\phi_n|H|\phi_i\rangle}{E_{ni}-\hbar\omega}\right|^2 \times \int D_f(E+2\hbar\omega)\, D_i(E)\, dE\,, \tag{42.5}$$

where H represents the light–electron interaction, ϕ is a related electron wavefunction, D is the density of states, and the subscripts i, n, and f denote the initial, intermediate, and final states, respectively. α and β are linked to the electronic structure, i. e., the wavefunctions and densities of states, which is determined by the atomic species and bonding structures involved [42.36]. From $\alpha(\hbar\omega)$, $n_0(\omega)$ is expressed using the conventional KKR as [42.2]

$$n_0(\omega) = 1 + \left(\frac{c}{\pi}\right) \wp \int \left(\frac{\alpha(\Omega)}{\Omega^2-\omega^2}\right) d\Omega\,. \tag{42.6}$$

For a nonlinear KKR response Δn, *Hutchings* et al. [42.46] have derived the relation

$$\Delta n(\omega;\zeta) = \left(\frac{c}{\pi}\right) \wp \int \left(\frac{\Delta\alpha(\Omega;\zeta)}{\Omega^2-\omega^2}\right) d\Omega\,, \tag{42.7}$$

where $\Delta\alpha$ is the nonlinear absorption induced by an excitation at ζ and probed at Ω. Note that this relation is derived from the causality principle for nondegenerate cases, for example, two-beam experiments with different photon energies, ζ and Ω. However, this relation may also be applied to degenerate cases ($\zeta = \Omega$) as a rough approximation as [42.47]

$$n_2(\omega) = \left(\frac{c}{\pi}\right) \wp \int \left(\frac{\beta(\Omega)}{\Omega^2-\omega^2}\right) d\Omega\,. \tag{42.8}$$

Using these formulae, we can roughly predict how β and n_2 depend upon E_g. Equation (42.5) suggests that $\beta \propto 1/E_g^2$, provided that the E_g dependence is governed by $|\Sigma_n 1/(E_{ni}-\hbar\omega)|^2$. This relation may be consistent with the material dependence shown in Fig. 42.3, $\beta_{max} \propto 1/E_g^{2\sim3}$, which is comparable to $\beta \propto 1/E_g^3$, the equation theoretically derived and experimentally confirmed for crystalline semiconductors [42.2, 41]. Then, putting $\beta \propto 1/E_g^2$ into (42.8), and assuming $\int[1/(\Omega^2-\omega^2)]d\Omega \propto 1/E_g^2$, we obtain a strong dependence of n_2 on E_g, $n_2 \propto 1/E_g^4$, which is consistent with (42.3). Note that the form is contrastive to $n_0 \propto 1/E_g^{1/4}$ [42.42].

We can also calculate, under some plausible assumptions, the spectral dependence of β and n_2 for an ideal amorphous semiconductor, which contains no gap states [42.48]. Figure 42.4 shows that, at $\hbar\omega = E_g/2$, $\beta = 0$ and n_2 is maximal. This is similar to some degree to the feature at $\hbar\omega = E_g$ for α and n_0. The figure also shows that β becomes maximal at $E_g/2-E_g$. Note, however, that it is not clear whether we can link this photon energy dependence to the experimental results shown in Fig. 42.3, since the observed spectra appear to have fairly sharp peaks.

Figure of Merit

Several figures of merit have been proposed for evaluating the performances of optical devices which utilize n_2. For instance, *Mizrahi* et al. [42.49] emphasized the negative effects of two-photon absorption, and provided a criterion $n_2/(2\beta\lambda_0) > 1$. Then, using the above E_g dependence, we see that $n_2/\beta \propto E_g^{-2}$, so that small E_g materials are preferable. Actually, as listed in Table 42.1, in this figure of merit, As_2S_3 appears to be better than SiO_2. However, this criterion implicitly neglects α. *Lines* [42.24] uses n_2/α instead. Here, α arises from the so-called residual absorption, which is difficult to estimate quantitatively. In addition, it is the attenuation coefficient α_0 instead of the absorption α that could be decisive. Table 42.1 shows that, in this measure, SiO_2 behaves better than As_2S_3. These figures of merit, however, have presumed only nonresonant electronic contributions with sub-ps responses. Then, *Jha* et al. [42.19] utilize $n_2/\tau\alpha$, where τ is a response time. Nevertheless, theoretical predictions of τ remain to be studied.

Naturally, the most appropriate definition of the figure of merit depends on the application of interest. For instance, for an optical fiber device, the maximum length might be $\approx 1\,\mathrm{m}$ [42.50], since the device must be compact and have a fast response. In this case, the light absorption, which arises from $\alpha + \beta I$, should be suppressed to below $10^{-2}\,\mathrm{cm}^{-1}$, or the light propagation loss must be smaller than $\approx 1\,\mathrm{dB/m}$. However, for optical integrated circuits, an effective propagation distance may be $\approx 1\,\mathrm{cm}$, in which the attenuation could be as large as $1\,\mathrm{cm}^{-1}$.

We can suggest another idea by taking the spectral dependence, as shown in Fig. 42.4, into account. That is, the best material that has n_2 at some $\hbar\omega$ is the one which satisfies a bandgap condition of $\hbar\omega = E_g/2$, where α and β are zero and n_2 becomes maximal, provided that there is no gap-state absorption. In practical glass samples, absorption due to impurities, dangling bonds such as E' centers in oxide glasses [42.21, 36], and wrong bonds in chalcogenide glasses [42.36, 48] cannot be neglected. In fact, we can see in Table 42.1 and Fig. 42.3 that residual attenuation exists in nominally transparent regions, parts of which are undoubtedly caused by absorption [42.51]. Note that these mid-gap states also give rise to two-step absorption (Fig. 42.2). We also cannot neglect photoinduced phenomena induced by these photoelectronic excitations (Sect. 42.4). As a matter of fact, selecting an appropriate glass for a specific application is not a straightforward process.

42.1.3 Stimulated Light Scattering, Supercontinuum Generation, and Other Applications

Stimulated Raman scattering has been attracting considerable interest, because it can be utilized for broadband light amplification [42.2, 3, 52]. As illustrated in Fig. 42.2d, Raman scattering can be regarded as a kind of nondegenerate two-photon process. However, in contrast to two-photon absorption, one photon is absorbed by a virtual state at the same time as another photon is emitted through the photon–phonon interaction. And, in the same way with conventional stimulated emission processes, when the incident light is intense, Stokes-shifted light can be amplified. The nonlinear polarization related to this process is given by

$$P^{(3)}(\omega_s) \propto \mathrm{i}\chi_R^{(3)} I(\omega_L) E(\omega_s)\,,$$

or the emitted light intensity is

$$\Delta I(\omega_s) \propto \chi_R^{(3)} I(\omega_L) I(\omega_s)\,,$$

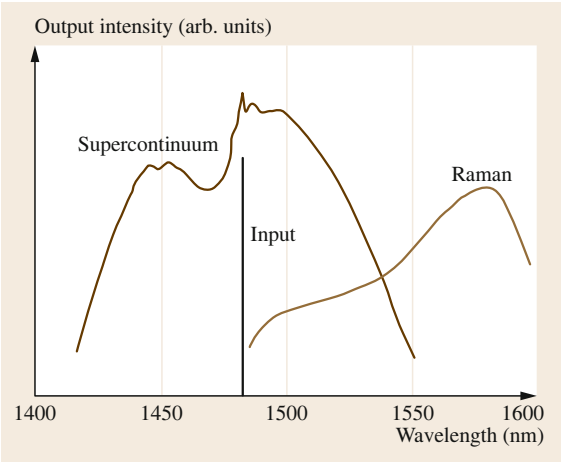


Fig. 42.7 Output spectra from a Raman fiber amplifier [42.54] and from a supercontinuum fiber generator [42.57] excited by 1483 nm light (schematic illustration)

where ω_L and ω_s ($< \omega_L$) are the exciting laser frequency and the Stokes-shifted frequency, respectively.

Most experiments have been performed using optical fibers [42.3, 50, 53] in order to obtain long interaction lengths. For instance, as illustrated in Fig. 42.7, Raman amplifiers comprising 5–10 km lengths of silica fibers can provide broadband gains of $\approx 10\,\mathrm{dB}$ [42.54]. Also interesting are Raman lasers using silica microspheres of diameter $\approx 70\,\mu\mathrm{m}$ [42.11] and amplifiers using submeter As_2Se_3 fibers [42.55]. It should also be mentioned here that stimulated Brillouin scattering in glasses has been explored recently [42.3, 56].

Supercontinuum generation in silica and other glass fibers has been studied intensively [42.1–3]. When pulsed or CW light propagates through a nonlinear medium, it may undergo spectral broadening. For instance, as illustrated in Fig. 42.7, a 350 m single-mode silica fiber excited by a 2.22 W CW laser with a wavelength of $1.483\,\mu\mathrm{m}$ has emitted 2.1 W light over a broad spectrum of 1.43–$1.53\,\mu\mathrm{m}$ [42.57]. The phenomenon appears under strong and prolonged nonlinear interactions, and accordingly, several mechanisms such as intensity-dependent refractive index changes, third-harmonic generation and stimulated Raman scattering could be responsible. Note that, unlike the stimulated light scattering described above, light of shorter wavelengths is also generated in this process. Such spectrum conversion fibers could be utilized as broadband light sources and optical amplifiers. The generation efficiency and bandwidth are becoming much improved using novel fiber structures consisting of, for example, heavy-metal oxide [42.58], tellurite [42.59, 60], chalcogenide [42.61, 62], and fluoride glasses [42.63].

However, in some compositions, atmospheric degradations have to be suppressed [42.64].

Finally, it is mentioned that the nonlinear refractivity in glasses has wider applications to all-optical switching [42.65], third-harmonic generation [42.66], THz-induced Kerr effect in As_2S_3 and As_2Se_3 [42.67], tunable sideband laser [42.68], and so forth. Nevertheless, the nonlinearity is a nuisance in some cases such as phase-modulated telecommunications [42.69]. It may also cause self-focusing damages [42.70].

42.2 Second-Order Nonlinearity in Poled Glass

It has been discovered that several kinds of poling methods can add second-order nonlinearity to glasses [42.17, 71]. At least five kinds have been demonstrated, which are listed in Table 42.2. Note that similar procedures are also employed for polymers. Most experiments utilize second harmonic signals to evaluate $\chi^{(2)}$ and, less commonly, electro-optical effects. Practical applications remain to be studied, while, for second-harmonic generations, the optical phase matching between exciting and nonlinearly generated light is a prerequisite.

The optical poling was demonstrated by *Österberg* and *Margulis* [42.18] using optical fibers. They found that exposing Ge-doped optical fibers of length ≈ 1 m to intense (≈ 70 kW) Nd:YAG laser light for ≈ 1 h could increase the second-harmonic signal to 0.55 kW. *Stolen* and *Tom* [42.72] utilized two light beams ($\times 1$ and $\times 2$, of Nd:YAG laser light) for induction, which reduced the exposure time to ≈ 5 min. However, the method was practical only for optical fibers, since the induced nonlinearity was relatively small, $\chi^{(2)} \approx 10^{-4}$ pm/V. Other glasses may be preferred, and for instance, in Ag-containing glass, laser-fabricated three-dimensional structures with ≈ 2 pm/V have been produced [42.73]. Second, the so-called thermal poling, which was actually electrothermal in nature, was demonstrated [42.74]. Application of high dc fields of 5×10^4 V/cm at $\approx 300\,^\circ$C to *bulk* SiO_2 samples induced a nonlinearity of ≈ 1 pm/V, a similar magnitude to that in quartz, which was evaluated from the second harmonic signals of 1.06 μm laser light. The induced nonlinearity is reasonably large and stable, and this method has been widely applied not only to oxide glasses [42.75–77] but also to chalcogenides [42.78]. Third, *Okada* et al. [42.79] applied corona-discharge poling to 7059 *films* at $\approx 200\,^\circ$C, which were deposited onto Pyrex glass substrates. Such corona-poling procedures, which can efficiently apply high DC fields, have often been also employed for organic polymers. Fourth, electron-beam poling could produce ≈ 1 pm/V in PbO-silica glass [42.80]. An advantage of this method is its high spatial resolution, which may hold promise in the fabrication of optical integrated circuits, despite the fact that a vacuum is needed. This method was applied to chalcogenide glasses as well, Ge-As-S, giving rise to ≈ 1 pm/V [42.81]. Fifth, proton implantation could add $\chi^{(2)}$ of ≈ 1 pm/V in fused silica with exposure times of 10^2 s [42.82]. Finally, *Fujiwara* et al. [42.83] demonstrated UV poling in Ge-doped SiO_2 subjected to electric fields of $\approx 10^5$ V/cm. The $\chi^{(2)}$ induced is reported to be ≈ 3 pm/V, which is comparable to that in $LiNbO_3$.

Two poling mechanisms of note have been proposed [42.17, 77]. One is that space charge produces a built-in electric field of E_{DC} ($\approx 10^6$ V/cm), which induces an effective $\chi^{(2)}$ of $3E_{DC}\chi^{(3)}$ [42.72]. This idea suggests a proportionality, $\chi^{(2)} \approx \chi^{(3)}$, and actually, $\chi^{(2)} \approx 10$ pm/V has been obtained in a high $\chi^{(3)}$ chalcogenide glass, $Ge_{25}Sb_{10}S_{65}$ [42.78]. On the other hand, E_{DC} is governed by the migration of ions such as Na^+ (impurities) under applied or generated electric fields [42.78, 84]. In agreement with this idea, $\chi^{(2)}$ tends to decay with a time constant of 10^1–10^6 days at room temperature, which is connected to the alkali ion mobility. The other idea is that oriented defects such as E' centers [42.21] are responsible for generating $\chi^{(2)}$. We can envisage that UV excitation produces defective dipoles, which are oriented with the static electric field.

It is plausible that the dominant mechanism depends on the poling method. Actually, we can divide the procedures listed in Table 42.2 into two types, depending upon whether or not the glass is heated during the poling process. The heating tends to enhance macroscopic ion migration, while it can relax microscopic defect

Table 42.2 Reported poling methods, applied objects, typical procedures, and induced $\chi^{(2)}$ values in silica. For references, see the main text. The $\chi^{(2)}$ values listed are compared with $\chi^{(2)}_{11} = 1$ pm/V in crystalline SiO_2 and $\chi^{(2)}_{22} = 5$ pm/V in $LiNbO_3$

Method	Object	Procedure	$\chi^{(2)}$ (pm/V)
Optical	Fiber	Nd:YAG laser, 1 h	10^{-4}
Thermal	Bulk	4 kV, 300 °C, 1 h	1
Corona	Film	5 kV, 300 °C, 15 min	1
e-Beam	Bulk	40 kV, 10 mA, 10 min	1
Proton	Bulk	500 kV, 1 mC, 100 s	1
Ultraviolet (UV)	Bulk	ArF laser, 10 kV	3

orientations at the same time [42.21, 36]. Therefore, it seems that the ion migration is responsible in thermal and corona poling, while defect orientation dominates in the other methods. In this context, poling at low temperatures may be a promising way of enhancing defect orientation.

Finally, we note that there exist some glassy materials which can exhibit $\chi^{(2)}$ without poling procedures. The so-called nano-crystallized glass and glass ceramics are well known, in which embedded crystals (or crystallites) are responsible for prominent $\chi^{(2)}$ [42.76, 85–87]. For instance, *Takahashi* et al. [42.86] demonstrated that a heat treatment of a BaO-TiO_2-SiO_2 glass sample at 760 °C for 1 h produced *oriented* $Ba_2TiSi_2O_8$ crystals, and the glass ceramic gives a prominent $\chi^{(2)}$ of $\approx 10\,\mathrm{pm/V}$. The thermal orientated crystallization has been applied not only to oxides [42.88] but also to chalcogenide and chalcohalide glass-ceramics [42.71] such as GeS_2-$CdGa_2S_4$ [42.89] and GeS_2-Ga_2S_3-AgCl [42.90]. Finally, we note that $\chi^{(2)}$ can appear in glasses through different mechanisms. It is known that $\chi^{(2)}$ emerges in TeO_2 glass-ceramics at interfaces [42.91], which probably cause strains. It was recently reported that sputter-deposited Ge-SiO_2 films exhibit $\chi^{(2)}$ amounting to 8 pm/V, which may be ascribed to oriented Ge-related defects [42.92].

42.3 Particle-Embedded Systems

Glasses that contain nanoparticles of semiconductors [42.93] and metals [42.94] have attracted considerable interest due to their unique third-order nonlinearities. Table 42.3 lists typical results [42.95–100]. Such glasses containing dispersed nanoparticles can be prepared by a variety of physical and chemical methods, for example, vacuum deposition and sol–gel techniques [42.1, 101]. For semiconductor systems, a lot of work has been performed using commercial semiconductor-doped color glass filters [42.93].

These nanoparticle systems efficiently function at close to the resonant wavelengths of some electronic excitations. This operating condition gives rise to at least three characteristics. First, the imaginary, not the real, part of $\chi^{(3)}$ may be more prominent, and accordingly, Table 42.3 compares absolute values. Second, the system exhibits a strong spectral dependence [42.1, 102]. For instance, Au–silica and CdSSe–silica are efficient at ≈ 580 and $\approx 800\,\mathrm{nm}$ [42.1], respectively. Third, the response time τ and the linear absorption α tend to become longer and higher. Actually, a trade-off between $\chi^{(3)}$ and τ (and α) seems to exist, for example, CdSSe-dispersed glasses show $\chi^{(3)}$ value of $\approx 10^{-9}$ esu with τ value of $\approx 20\,\mathrm{ps}$, while Au-dispersed glasses give $\approx 10^{-11}$ esu and $\approx 1\,\mathrm{ps}$ [42.1]. Linear absorption can be as large as $10^4\,\mathrm{cm}^{-1}$ [42.94], so these systems can be utilized as small devices, not as fibers. Note that, in pure silica at nonresonant infrared wavelengths, $\chi^{(3)} \approx 10^{-13}$ esu, $\tau \approx 10\,\mathrm{fs}$, and $\alpha \leq 10^{-5}\,\mathrm{cm}^{-1}$ [42.94].

Due to these inherent features, the particle-embedded system should surmount two problems for wide applications. One is the reduction of linear attenuation, including absorption and scattering, and the other is the shift of resonant wavelengths to the optical communication region, 1.3–1.5 μm. Is such a shift possible? What are the mechanisms that give rise to these prominent nonlinearities in particle-embedded systems?

When the particle is a semiconductor such as CuCl and CdSSe, excitons or confined electron–hole pairs are responsible [42.93]. Specifically, the excitons in semiconductor particles behave as two-level systems, and at the resonance frequency, $|\chi^{(3)}|$ is written as

$$|\chi^{(3)}| \approx \mathrm{Im}\{\chi^{(3)}\} \propto |\mu|^4 N T_1 T_2^2 \,, \tag{42.9}$$

where μ is the dipole moment of the exciton, N is the particle number, T_1 ($\approx 100\,\mathrm{ps}$) is the lifetime of the exciton, and T_2 ($\approx$ fs) is the dephasing time. A quan-

Table 42.3 Several particle systems, along with their preparation methods, $|\chi^{(3)}|$ values, and response times τ at the measured wavelength λ, as well as references. PLD and VE depict pulsed laser deposition and vacuum evaporation

System	Preparation	$\lvert\chi^{(3)}\rvert$ (esu)	τ (ps)	λ (nm)	Reference
Au (15 nm)/silica	Shell structure	10^{-9}	2	550	[42.95]
Cu (2 nm)/Al_2O_3	PLD	10^{-7}	5–450	600	[42.96]
Ag (20 nm)/BaO	VE	10^{-10}	0.2	820	[42.97]
Fe (4 nm)/$BaTiO_3$	PLD	10^{-6}		532	[42.98]
SnO_2 (10 nm)/silica	Sol–gel	10^{-12}		1064	[42.99]
CdS (4 nm)/silica	Sol–gel	10^{-11}		500	[42.100]

titative estimation predicts that closely packed CuCl particles of radius 40 nm could provide a $|\chi^{(3)}|$ enhancement of a factor of $\approx 10^3$ when compared with the bulk value [42.103].

When metal particles such as spherical Au particles with diameters of 10–50 nm are used, we can envisage local field enhancement and dynamic responses from conduction electrons, including plasmon effects [42.94, 104]. The result is approximately written as

$$\chi^{(3)} \approx p_\mathrm{m}\chi^{(3)}_\mathrm{m}\left|\frac{3\varepsilon_\mathrm{h}}{\varepsilon_\mathrm{m}+2\varepsilon_\mathrm{h}}\right|^2\left(\frac{3\varepsilon_\mathrm{h}}{\varepsilon_\mathrm{m}+2\varepsilon_\mathrm{h}}\right)^2 , \quad (42.10)$$

where p_m is the volume fraction of the metal particles, $\chi^{(3)}_\mathrm{m}$ is the bulk nonlinearity of the metal, and ε_h and ε_m are the linear dielectric constants of the host (glass) and the metal, respectively. The nonlinearity of the host is neglected here for the sake of simplicity. Also note that ε_h can be real, while ε_m is complex. We then see that the metal's nonlinearity $\chi^{(3)}_\mathrm{m}$ is decreased by the volume factor p_m, while $\chi^{(3)}_\mathrm{m}$ may be enhanced by local fields if $\mathrm{Re}\{\varepsilon_\mathrm{m}+2\varepsilon_\mathrm{h}\}\approx 0$, which determines the resonance wavelength. In agreement with this model, an Au-dispersed glass, for instance, gives a greater $\chi^{(3)}$ than an Au film [42.1].

However, since these systems are complex, consisting of particles and a host glass, a variety of situations arise. For the particle, as well as particle species, we need to consider the size, shape, and concentration. Here, needless to say, a narrow size distribution is preferred for investigating fundamental mechanisms [42.105]. The shape may be spherical, ellipsoidal [42.106], a rod, and so forth. The concentration determines the mean separation distance between particles. As the concentration increases, electrical particle–particle interactions appear [42.95], and then the particles eventually percolate [42.96], which may produce dramatic changes in the nonlinear response. On the other hand, the host material seems to be of secondary importance. Actually, liquids and organic materials as well as glasses have been employed [42.1]. However, some reports suggest that the nonlinearity is greatly affected by the surface states of semiconductor particles [42.107] and by the dielectric constant of a host surrounding metallic particles [42.1].

Finally, a few challenging results may be worth mentioning here. One is that particles can be arrayed to produce photonic structures that exhibit novel nonlinear properties, such as light confinement [42.108]. It is also demonstrated that an arrayed molecular structure, Se-loaded zeolite, gives higher nonlinearity by three orders of magnitude than those of glassy Se [42.109]. The other is the generation of second harmonic signals from oriented ellipsoidal Ag nanoparticles in silica [42.110], which have been produced through a kind of mechanical poling; tensile deformation under simultaneous heating.

42.4 Photoinduced Phenomena

The intense pulsed light, which is employed to produce nonlinear effects, is also likely to produce a variety of transitory and (quasi-)stable optical changes [42.107, 111]. Photochromic effects induced by subgap light are the result of transitory electronic changes [42.112]. The increase in the refractive index of SiO_2 glass induced by fs–ns laser light with $\hbar\omega$ on the range of 1–5 eV is known to occur along with quasi-stable structural changes [42.21]. The optical poling described in Sect. 42.2 can be regarded as a kind of photoinduced anisotropy. When the light is more intense than 1 GW/cm^2–1 TW/cm^2, depending on the duration of the light pulse, permanent damages are likely to occur [42.2]. These photostructural changes can be induced by electronic excitations, which may be triggered by one- and multi-photon processes, depending upon the photon energy and the light intensity. Temperature rises may also add some effects.

However, we should be careful to assert if these photoinduced phenomena are really excited by nonlinear processes. For instance, when a photoinduced phenomenon is induced by exposure to photons with an energy $\hbar\omega \approx E_\mathrm{g}/m$, where m is an integer, it is often asserted that m-photon absorption is responsible. Nevertheless, the phenomenon may be triggered by one-photon absorption by mid-gap states, which give weak absorption tails (Fig. 42.5) arising from defective structures in the glass [42.36, 113]. Otherwise, it may be triggered by m-step absorption processes [42.114], consisting of m one-photon absorptions in succession (Fig. 42.2c). These processes have probably been overlooked when analyzing experimental results.

In addition, as schematically illustrated in Fig. 42.8, the structural changes induced by linear and nonlinear excitation are not necessarily the same. For instance, in As_2S_3 glass, bandgap excitations induced by one-photon and two-photon absorption produce different changes [42.113]; one-photon excitation leads to photodarkening, while two-photon excitation results in an

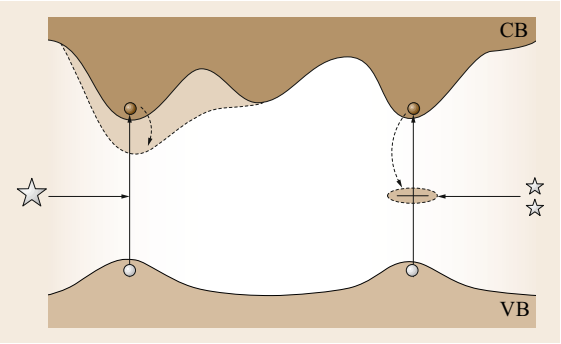


Fig. 42.8 Different relaxation paths from bandgap excitation due to one-photon (*left*) and two-photon (*right*) processes. For instance, the one-photon and two-photon excitations induce bonding strains and defect alterations, respectively

increase in the refractive index with no photodarkening. Such different changes depending upon excitation processes can appear in glasses because the *localized* atomic wavefunction plays an important role, in contrast to the extended Bloch wavefunctions in crystals. For instance, as we can see from (42.4), one-photon excitation occurs between wavefunctions with different parities, such as from s to p orbitals, since H is an odd function. In contrast, (42.5) shows that two-photon excitation tends to occur between states of the same parity, such as from p to p. In addition, $1/(E_{\rm ni}-\hbar\omega)$ in the equation leads to the possibility of *resonant and localized* two-photon excitation if $E_{\rm ni}-\hbar\omega\approx 0$ is satisfied for a mid-gap state. In such cases, the mid-gap state selectively absorbs the excitation energy, leading to an atomic transformation, which is likely to appear as a photoinduced macroscopic change.

42.5 Summary

At present, there is a big gap between the science and the technology of glasses. From a physical point of view, glass science lags far behind crystal science, mainly due to experimental difficulties associated with atomic structure determination and limited applicability of the formulation using spatially extended phonons and electron wavefunctions. On the other hand, it is fully unrealistic to imagine a world without optical glass fibers. And, cutting-edge optical processing needs, in addition to electro-optical devices, novel photo-optical devices, which may utilize the optical nonlinearity. These situations are promoting a deeper understanding of nonlinear photonic glasses and leading to a wider range of photonics applications, the result appearing as the increasing number of publications (Fig. 42.1).

However, the magnitude of optical nonlinearity may still be insufficient for fabricating compact devices. And, we have been exploring newer glass compositions and architectures. For the former, research appears to be gradually shifting from oxides to chalcogenides and more complicated systems, which possess higher nonlinearity with fast responses. However, few studies have been performed to predict the response time τ of nonlinearity in a glass at a particular excitation energy. For device structures, we are developing photonic-structured fibers, nanowires, snaky waveguides, and so forth, which may be assembled as functional bundles and also as plates for two-dimensional signal processing.

Third-order optical nonlinearity in homogeneous glasses has been studied in a fair amount of detail. Substantial experimental data have been obtained for n_2, which have been analyzed using empirical relations such as those from Boling. In contrast, less work has been done on the nonlinear absorption β. In this chapter, therefore, we have tried to present, in a coherent way, the absorption and the refractive index using semiconductor terminology. This approach can be used to connect nonlinear properties to the energy gap.

Two heterogeneous systems have aroused increasing interest. One is poled, crystallized, or ceramic glasses. Enhanced second-order nonlinearity has been reported for such systems, and this can sometimes be added selectively to desired locations. The other is the particle-embedded system, which can give greater third-order nonlinearity, and in some cases enhanced second-order nonlinearity too. However, the mechanism associated with this system remains to be studied. One experimental problem is to find a method that can reproduce particles of a fixed size and shape in arbitrary concentrations. These heterogeneous systems may be fabricated in combination with structured fibers or photonic crystals for improving nonlinear performances.

Finally, nonlinear optical excitations appear to play important roles in a variety of photoinduced phenomena in glasses. However, fundamental studies are still lacking in this area. The phenomenon of interest may be nonlinear in nature, or the linear excitations of gap

states may trigger successive changes. Attempting to understand the nonlinear photo-electro-structural process provides a challenging problem.

Acknowledgments. The author would like to thank his students, K. Sugawara and N. Minamikawa, for preparing illustrations and giving comments.

References

42.1 R.L. Sutherland: *Handbook of Nonlinear Optics*, 2nd edn. (Marcel Dekker, New York 2003)
42.2 R.W. Boyd: *Nonlinear Optics*, 3rd edn. (Academic, San Diego 2008)
42.3 G.P. Agrawal: *Applications of Nonlinear Fiber Optics*, 5th edn. (Academic, San Diego 2012)
42.4 A. Zakery, S.R. Elliott: *Optical Nonlinearities in Chalcogenide Glasses and Their Applications* (Springer, New York 2007)
42.5 B.J. Eggleton, B. Luther-Davies, K. Richardson: Nat. Photonics **5**, 141 (2011)
42.6 J.L. Adam, X. Zhang (Eds.): *Chalcogenide Glasses* (Woodhead Publishing, Oxford 2014)
42.7 M. Ferrera, L. Razzari, D. Duchesne, R. Morandotti, Z. Yang, M. Liscidini, J.E. Sipe, S. Chu, B.E. Little, D.J. Moss: Nat. Photonics **2**, 737 (2008)
42.8 C. Markos, I. Kubat, O. Bang: Sci. Rept. **4**, 6057 (2014)
42.9 J.D. LeGrange, M. Dinu, T. Sochor, P. Bollond, A. Kasper, S. Cabot, G.S. Johnson, I. Kang, A. Grant, J. Kay: Opt. Express **22**, 3600 (2014)
42.10 M.H. Field, J. Popp, R.K. Chang: Nonlinear optics in microspheres. In: *Progress in Optics*, Vol. 41, ed. by E. Wolf (North Holland, Amsterdam 2000) p. 1
42.11 S.M. Spillane, T.J. Kippenberg, K.J. Vahala: Nature **415**, 621 (2002)
42.12 E.C. Mägi, L.B. Fu, H.C. Nguyen, M.R. Lamont, D.I. Yeom, B.J. Eggleton: Opt. Express **15**, 10324 (2007)
42.13 M.A. Foster, A.C. Turner, M. Lipson, A.L. Gaeta: Opt. Express **16**, 1300 (2008)
42.14 T.M. Monro, H. Ebendorff-Heidepriem: Annual Rev. Mater. Res. **36**, 467 (2006)
42.15 S.J. Madden, D.-Y. Choi, D.A. Bulla, A.V. Rode, B. Luther-Davies, V.G. Ta'eed, M.D. Pelusi, B.J. Eggleton: Opt. Express **15**, 14414 (2007)
42.16 M. Hughes, W. Yang, D. Hewak: Appl. Phys. Lett. **90**, 131113 (2007)
42.17 Y. Quiquempois, P. Niay, M. Dounay, B. Poumellec: Curr. Opin. Solid State Mater. Sci. **7**, 89 (2003)
42.18 U. Österberg, W. Margulis: Opt. Lett. **11**, 516 (1986)
42.19 A. Jha, X. Liu, A.K. Kar, H.T. Bookey: Curr. Opin. Solid State Mater. Sci. **5**, 475 (2001)
42.20 J. Heo: Rare-earth-doped chalcogenide glass for lasers and amplifiers. In: *Chalcogenide Glasses*, ed. by J.L. Adam, X. Zhang (Woodhead Publishing, Oxford 2014)
42.21 G. Pacchioni, L. Skuja, D.L. Griscom: *Defects in SiO_2 and Related Dielectrics: Science and Technology* (Kluwer, Dordrecht 2000)
42.22 E.M. Vogel, M.J. Weber, D.M. Krol: Phys. Chem. Glasses **32**, 231 (1991)
42.23 K. Tanaka: J. Mater. Sci. Mater. Electron. **16**, 633 (2005)
42.24 M.E. Lines: J. Appl. Phys. **69**, 6876 (1991)
42.25 A. Zakery, S.R. Elliott: J. Non-Cryst. Solids. **330**, 1 (2003)
42.26 J. Requejo-Isidro, A.K. Mairaj, V. Pruneri, D.W. Hewak, M.C. Netti, J.J. Baumberg: J. Non-Cryst. Solids **317**, 241 (2003)
42.27 K. Tanaka: J. Non-Cryst. Solids. **338–340**, 534 (2004)
42.28 K. Ogusu, J. Yamasaki, S. Maeda, M. Kitao, M. Minakata: Opt. Lett. **29**, 265 (2004)
42.29 T. Mizunami, K. Takagi: Opt. Lett. **19**, 463 (1994)
42.30 A. Dragonmir, J.G. McInerney, N. Nikogosyan: Appl. Opt. **41**, 4365 (2002)
42.31 K. Imanishi, Y. Watanabe, T. Watanabe, T. Tsuchiya: J. Non-Cryst. Solids **259**, 139 (1999)
42.32 H. Ebendorff-Heidepriem, P. Petropoulos, S. Asimakis, V. Finazzi, R. Moore, K. Frampton, F. Koizumi, D. Richardson, T. Monro: Opt. Express **12**, 5082 (2004)
42.33 K. Tanaka: Appl. Phys. Lett. **80**, 177 (2002)
42.34 R.C. Enck: Phys. Rev. Lett. **31**, 220 (1973)
42.35 P.S. Weitzman, U. Osterberg: J. Appl. Phys. **79**, 8648 (1996)
42.36 S.R. Elliott: *Physics of Amorphous Materials*, 2nd edn. (Longman Scientific, Essex 1990)
42.37 N.L. Boling, A.J. Glass, A. Owyoung: IEEE J. Quantum Electron. **QE-14**, 601 (1978)
42.38 V. Dimitrov, S. Sakka: J. Appl. Phys. **79**, 1741 (1996)
42.39 J. Qi, D.F. Xue, G.L. Ning: Phys. Chem. Glasses **45**, 362 (2004)
42.40 S. Suehara, P. Thomas, A. Mirgorodsky, T. Merle-Mejean, J.C. Champarnaud-Mesjard, T. Aizawa, S. Hishita, S. Todoroki, T. Konishi, S. Inoue: J. Non-Cryst. Solids **345/346**, 730 (2004)
42.41 M. Sheik-Bahae, D.J. Hagan, E.W. Van Stryland: Phys. Rev. Lett. **65**, 96 (1990)
42.42 T.S. Moss: *Optical Properties of Semiconductors* (Butterworths, London 1959) p. 48
42.43 N. Sugimoto, H. Kanbara, S. Fujiwara, K. Tanaka, K. Hirao: Opt. Lett. **21**, 1637 (1996)
42.44 S. Smolorz, I. Kang, F. Wise, B.G. Aiken, N.F. Borrelli: J. Non-Cryst. Solids **256/257**, 310 (1999)
42.45 O. Sugimoto, H. Nasu, J. Matsuoka, K. Kamiya: J. Non-Cryst. Solids **161**, 118 (1993)
42.46 D.C. Hutchings, M. Sheik-Bahae, D.J. Hagan, E.W. Van Stryland: Opt. Quantum Electron. **24**, 1 (1992)
42.47 M. Sheik-Bahae, D.C. Hutchings, D.J. Hagan, E.W. Van Stryland: IEEE J. Quantum Electron. **QE-27**, 1296 (1991)
42.48 K. Tanaka: *Optoelectronic Materials and Devices*, Vol. 1, ed. by G. Lucovsky, M. Popescu (INOE, Bucharest 2004)

42.49 V. Mizrahi, K.W. DeLong, G.I. Stegeman, M.A. Saifi, M.J. Andrejco: Opt. Lett. **14**, 1140 (1989)
42.50 G.I. Stegeman, R.H. Stolen: J. Opt. Soc. Am. B **6**, 652 (1989)
42.51 K. Tanaka, T. Gotoh, N. Yoshida, S. Nonomura: J. Appl. Phys. **91**, 125 (2002)
42.52 M.N. Islam: IEEE J. Sel. Top. Quantum Electron. **8**, 548 (2002)
42.53 J. Ballato, P. Dragic: Materials **7**, 4411 (2014)
42.54 J. Bromage: J. Lightwave Technol. **22**, 79 (2004)
42.55 R. Ahmad, M. Rochette: IEEE J. Sel. Top. Quantum Electron. **20**, 0900706 (2014)
42.56 I.V. Kabakova, R. Pant, H.G. Winful, B.J. Eggleton: J. Nonlinear Opt. Phys. Mater. **23**, 1 (2014)
42.57 A. Zheltikov: Appl. Phys. B **77**, 143 (2003)
42.58 R. Buczynski, H.T. Bookey, D. Pysz, R. Stepien, I. Kujawa, J.E. McCarthy, A.J. Waddie, A.K. Kar, M.R. Taghizadeh: Laser Phys. Lett. **7**, 666 (2010)
42.59 A.K. Abeeluck, C. Headley: Appl. Phys. Lett. **85**, 4863 (2004)
42.60 J.H.V. Price, X. Feng, A.M. Heidt, G. Brambilla, P. Horak, F. Poletti, G. Ponzo, P. Petropoulos, M. Petrovich, J. Shi, M. Ibsen, W.H. Loh, H.N. Rutt, D.J. Richardson: Opt. Fiber Technol. **18**, 327 (2012)
42.61 W. Gao, M. El Amraoui, M. Liao, H. Kawashima, Z. Duan, D. Deng, T. Cheng, T. Suzuki, Y. Messaddeq, Y. Ohishi: Opt. Express **21**, 9573 (2013)
42.62 C.R. Petersen, U. Møller, I. Kubat, B. Zhou, S. Dupont, J. Ramsay, T. Benson, S. Sujecki, N. Abdel-Moneim, Z. Tang, D. Furniss, A. Seddon, O. Bang: Nat. Photonics **8**, 830 (2014)
42.63 J. Swiderski, M. Michalska: Opt. Laser Technol. **52**, 75 (2013)
42.64 O. Mouawad, F. Amrani, B. Kibler, J. Picot-Clémente, C. Strutynski, J. Fatome, F. Désévédavy, G. Gadret, J.-C. Jules, O. Heintz, E. Lesniewska, F. Smektala: Opt. Express **22**, 23912 (2014)
42.65 M. Galili, J. Xu, H.C. Mulvad, L.K. Oxenløwe, A.T. Clausen, P. Jeppesen, B. Luther-Davies, S. Madden, A. Rode, D-..Y. Choi, M. Pelusi, F. Luan, B. Eggleton: Opt. Express **17**, 2182 (2009)
42.66 A.H. Reshak, O.S. Klymovych, G.L. Myronchuk, O.V. Zamuruyeva, O.F. Zmiy, Z.A. Alahmed, J. Chyský, J. Bila, H. Kamarudin: J. Appl. Phys. **116**, 143102 (2014)
42.67 M. Zalkovskij, A.C. Strikwerda, K. Iwaszczuk, A. Popescu, D. Savastru, R. Malureanu, A.V. Lavrinenko, P.U. Jepsen: Appl. Phys. Lett. **103**, 221102 (2013)
42.68 L. Zhu, W. Liu, C. Fang: Appl. Phys. Lett. **103**, 061110 (2013)
42.69 R. Slavik, F. Parmigiani, J. Kakande, C. Lundström, M. Sjödin, P.A. Andrekson, R. Weerasuriya, S. Sygletos, A.D. Ellis, L. Grüner-Nielsen, D. Jakobsen, S. Herstrøm, R. Phelan, J. O'Gorman, A. Bogris, D. Syvridis, S. Dasgupta, P. Petropoulos, D.J. Richardson: Nat. Photonics **4**, 690 (2010)
42.70 J.B. Ashcom, R.R. Gattass, C.B. Schaffer, E. Mazur: J. Opt. Soc. Am. B – Opt. Phys. **23**, 2317 (2006)
42.71 V. Nazabal, I. Kityk: Second harmonic generation in chalcogenide glasses. In: *Chalcogenide Glasses*, ed. by J.L. Adam, X. Zhang (Woodhead, Oxford 2014)
42.72 R.H. Stolen, H.W.K. Tom: Opt. Lett. **12**, 585 (1987)
42.73 J. Choi, M. Bellec, A. Royon, K. Bourhis, G. Papon, T. Cardinal, L. Canioni, M. Richardson: Opt. Lett. **37**, 1029 (2012)
42.74 R.A. Myers, N. Mukherjee, S.R.J. Brueck: Opt. Lett. **22**, 1732 (1991)
42.75 B. Ferreira, E. Fargin, J.P. Manaud, G. Le Flem, V. Rodriguez, T. Buffeteau: J. Non-Cryst. Solids **343**, 121 (2004)
42.76 G.S. Murugan, T. Suzuki, Y. Ohishi, Y. Takahashi, Y. Benino, T. Fujiwara, T. Komatsu: Appl. Phys. Lett. **85**, 3405 (2004)
42.77 M. Dussauze, V. Rodriguez, A. Lipovskii, M. Petrov, C. Smith, K. Richardson, T. Cardinal, E. Fargin, E.I. Kamitsos: J. Phys. Chem. C **114**, 12754 (2010)
42.78 M. Guignard, V. Nazabal, F. Smektala, J.L. Adam, O. Bohnke, C. Duverger, A. Moreac, H. Zeghlache, A. Kudlinski, G. Martinelli: Adv. Funct. Mater. **17**, 3284 (2007)
42.79 A. Okada, K. Ishii, K. Mito, K. Sasaki: Appl. Phys. Lett. **60**, 2853 (1992)
42.80 P.G. Kazansky, A. Kamal, P.S. Russell: Opt. Lett. **18**, 683 (1992)
42.81 Q.M. Liu, F.X. Gan, X.J. Zhao, K. Tanaka, A. Narazaki, K. Hirao: Opt. Lett. **26**, 1347 (2001)
42.82 L.J. Henry, B.V. McGrath, T.G. Alley, J.J. Kester: J. Opt. Soc. Am. B **13**, 827 (1996)
42.83 T. Fujiwara, M. Takahashi, A.J. Ikushima: Appl. Phys. Lett. **71**, 1032 (1997)
42.84 O. Deparis, C. Corbari, G. Kazansky, K. Sakaguchi: Appl. Phys. Lett. **84**, 4857 (2004)
42.85 V. Pruneri, P.G. Kazansky, D. Hewak, J. Wang, H. Takebe, D.N. Payne: Appl. Phys. Lett. **70**, 155 (1997)
42.86 Y. Takahashi, Y. Benino, T. Fujiwara, T. Komatsu: Appl. Phys. Lett. **81**, 223 (2002)
42.87 I. Chung, M.G. Kanatzidis: Chem. Mater. **26**, 849 (2014)
42.88 N. Toyohara, Y. Benino, T. Fujiwara, T. Komatsu: Solid State Commun. **140**, 299 (2006)
42.89 H. Tao, C. Lin, S. Gu, C. Jing, X. Zhao: Appl. Phys. Lett. **91**, 011904 (2007)
42.90 M. Guignard, V. Nazabal, A. Moreac, S. Cherukulappurath, G. Boudebs, H. Zeghlache, G. Martinelli, Y. Quiquempois, F. Smektala, J.-L. Adam: J. Non-Cryst. Solids **354**, 1322 (2008)
42.91 R.T. Hart, K.M. Ok, P.S. Halasyamani, J.W. Zwanziger: Appl. Phys. Lett. **85**, 938 (2004)
42.92 I. Kawamura, K. Imakita, M. Fujii, S. Hayashi: Appl. Phys. Lett. **103**, 201117 (2013)
42.93 G. Banfi, V. Degiorgio, D. Ricard: Adv. Phys. **47**, 510 (1998)
42.94 F. Gonella, P. Mazzoldi: Metal nanocluster composite glasses. In: *Handbook of Nanostructured Materials and Nanotechnology*, Vol. 4, ed. by H.S. Nalwa (Academic, San Diego 2000) p. 81
42.95 Y. Hamanaka, K. Fukuta, A. Nakamura, L.M. Liz-Marzan, P. Mulvaney: Appl. Phys. Lett. **84**, 4938 (2004)
42.96 R. Del Coso, J. Requejo-Isidro, J. Solis, J. Gonzalo, C.N. Afonso: J. Appl. Phys. **95**, 2755 (2004)

42.97 Q.F. Zhang, W.M. Liu, Z.Q. Xue, J.L. Wu, S.F. Wang, D.L. Wang, Q.H. Gong: Appl. Phys. Lett. **82**, 958 (2003)
42.98 W. Wang, G. Yang, Z. Chen, Y. Zhou, H. Lu, G. Yang: J. Appl. Phys. **92**, 7242 (2002)
42.99 A. Clementi, N. Chiodini, A. Paleari: Appl. Phys. Lett. **84**, 960 (2004)
42.100 S.G. Lu, Y.J. Yu, C.L. Mak, K.H. Wong, L.Y. Zhang, X. Yao: Microelectron. Eng. **66**, 171 (2003)
42.101 M. Nogami, S.T. Selvan, H. Song: Photonic glasses: Nonlinear optical and spectral hole burning properties. In: *Handbook of Advanced Electronic and Photonic Materials and Devices*, Vol. 5, ed. by H.S. Nalwa (Academic, San Diego 2001) p. 141
42.102 J. He, W. Ji, G.H. Ma, S.H. Tang, H.I. Elim, W.X. Sun, Z.H. Zhang, W.S. Chin: J. Appl. Phys. **95**, 6381 (2004)
42.103 Y. Li, M. Takata, A. Nakamura: Phys. Rev. B **57**, 9193 (1998)
42.104 T. Cesca, P. Calvelli, G. Battaglin, P. Mazzoldi, G. Mattei: Opt. Express **20**, 4537 (2012)
42.105 R. Sato, M. Ohnuma, K. Oyoshi, Y. Takeda: Phys. Rev. B **90**, 125417 (2014)
42.106 S. Mohan, J. Lange, H. Graener, G. Seifert: Opt. Express **20**, 28655 (2012)
42.107 A. Puzder, A.J. Williamson, F. Gygi, G. Galli: Phys. Rev. Lett. **92**, 217401 (2004)
42.108 M. Ajgaonkar, Y. Zhang, H. Grebel, C.W. White: Appl. Phys. Lett. **75**, 1532 (1999)
42.109 K. Tanaka, A. Saitoh: Appl. Phys. Lett. **94**, 241905 (2009)
42.110 A. Podlipensky, J. Lange, G. Seifert, H. Graener, I. Cravetchi: Opt. Lett. **28**, 716 (2003)
42.111 R.C. Jin, Y.C. Cao, E.C. Hao, G.S. Metraux, G.C. Schatz, C.A. Mirkin: Nature **425**, 487 (2003)
42.112 Y. Watanabe, Y. Kikuchi, K. Imanishi, T. Tsuchiya: Mater. Sci. Eng. B **54**, 11 (1998)
42.113 K. Tanaka: Philos. Mag. Lett. **84**, 601 (2004)
42.114 K. Kajihara, Y. Ikuta, M. Hirano, H. Hosono: Appl. Phys. Lett. **81**, 3164 (2002)

Part E Novel Ma

Part E Novel Materials and Selected Applications

43 Solar Cells and Photovoltaics
Stuart Irvine, Swansea, UK

44 Disordered Semiconductors on Mechanically Flexible Substrates for Large-Area Electronics
Peyman Servati, Vancouver, Canada
Arokia Nathan, Cambridge, UK

45 Photoconductors for X-Ray Image Detectors
M. Zahangir Kabir, Montreal, Canada
Safa Kasap, Saskatoon, Canada

46 Phase-Change Memory Materials
Alexander V. Kolobov, Tsukuba, Japan
Junji Tominaga, Tsukuba, Japan
Paul Fons, Tsukuba, Japan

47 Carbon Nanotubes and Bucky Materials
Mark Baxendale, London, UK

48 Graphene
Henry H. Radamson, Beijing, China, Stockholm, Sweden

49 Magnetic Information-Storage Materials
Charbel Tannous, Brest, France
R. Lawrence Comstock†

50 High-Temperature Superconductors
Rainer Wesche, Villigen, Switzerland

51 Molecular Electronics
Michael C. Petty, Durham, UK
Takashi Nagase, Sakai, Japan
Hitoshi Suzuki, Higashihiroshima, Japan
Hiroyoshi Naito, Sakai, Japan

52 Organic Materials for Chemical Sensing
Asim K. Ray, Middlesex, UK

53 Packaging Materials
Darrel Frear, Chandler, USA

54 Organic Solar Cells
Masahiro Hiramoto, Okazaki, Japan
Yusuke Shinmura, Okazaki, Japan

55 Materials for Terahertz Engineering
Roger Lewis, Wollongong, Australia

56 Metamaterials
Vassili Fedotov, Southampton, UK

57 Thermoelectric Materials
Donald T. Morelli, East Lansing, USA

58 Transparent Conductive Oxides
Hideo Hosono, Yokohama, Japan
Kazushige Ueda, Kitakyushu, Japan

59 Inorganic Perovskite Oxides
Tatsumi Ishihara, Fukuoka, Japan

43. Solar Cells and Photovoltaics

Stuart Irvine

Photovoltaic solar cells are gaining wide acceptance for producing clean, renewable electricity. This has been based on more than 40 years of research that has benefited from the revolution in silicon electronics and compound semiconductors in optoelectronics. This chapter gives an introduction into the basic science of photovoltaic solar cells and the challenge of extracting the maximum amount of electrical energy from the available solar energy. In addition to the constraints of the basic physics of these devices, there are considerable challenges in materials synthesis. The latter has become more prominent with the need to reduce the cost of solar module manufacture as it enters mainstream energy production. The chapter is divided into sections dealing with the fundamentals of solar cells and then considering six very different materials systems, from crystalline silicon through to polycrystalline thin films and perovskites. These materials have been chosen because they are all either in production or have the prospect of being in production over the next few years. Many more materials are being considered in research and some of the more exciting, excitonic cells and nanomaterials are mentioned. However, there is insufficient space to give these very active areas of research the justice they deserve. I hope the reader will feel sufficiently inspired by this topic to read further and explore one of the most exciting areas of semiconductor science. The need for high-volume production at low cost has taken the researcher along paths not normally considered in semiconductor devices and it is this that provides an exciting challenge.

43.1 Figures of Merit for Solar Cells 1098
43.2 Crystalline Silicon 1100
43.3 Amorphous Silicon 1102
43.4 GaAs Solar Cells 1102
43.5 CdTe Thin-Film Solar Cells 1104
43.6 $CuInGaSe_2$ ($CIGS_2$) Thin-Film Solar Cells 1105
43.7 Excitonic PV 1107
43.8 Conclusions 1107
References 1108

Photovoltaic (PV) devices are a method of converting radiant solar energy into electrical energy. Most of our energy sources, including fossil fuels, hydroelectric and wind power actually come from solar radiation but are indirect conversions into electricity. Another class of solar energy conversion is the heating of water in solar thermal panels. Although the conversion efficiency can be high, they do not generate the thermal energy necessarily where and when it is needed, so storage is required. Direct generation of electric energy is attractive because it is a versatile energy form, rapidly converted into heat, mechanical or light energy. Photovoltaic energy is the main source of energy in the rapidly expanding satellite market with high-efficiency photovoltaic modules producing more than 1 kW of power. Terrestrial applications are also rapidly growing with an estimated total installed capacity worldwide of over 177 GW at the end of 2014 [43.1], and this has been increasing annually by 30%. This, however, is small compared with the total amount of electrical energy consumed each year, but is now supplying more than 1% of primary energy demand. Solar energy is very attractive as it is completely nonpolluting and can help to reduce the amount of fossil fuels that we burn to generate electricity. World CO_2 emissions grew by 4% a year up to 2012. Any contribution from non-fossil-fuel alternatives such as solar energy will help to reduce this annual burden of CO_2 emissions that is now a widely accepted cause of global warming. So how much solar energy is available for conversion to electricity? The total solar energy falling on the Earth's surface each year is huge and 10 000 times the current consumption of energy. We only need to capture a tiny fraction of this to make a major contribution to

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_43

our electricity supply but this will mean incorporating solar electric panels into most buildings; to achieve this it will need to be much cheaper to compete with current fossil-fuel electricity generation.

The range of applications outlined here place different demands on the design of the photovoltaic module and hence the materials solutions may well be different. Cost has already been mentioned as paramount for terrestrial power requirements, but for space applications resistance to ionising radiation is important. This section will consider the materials, the implications of competing technologies and how well they meet the criteria for different applications. The first section will deal with how a photovoltaic cell operates and introduces the figures of merit that are used to compare photovoltaic cells and relate performance to theoretical performance. Solar cell technology is promoted as a clean fuel technology that does not introduce CO_2 and pollutants into the atmosphere. However, for this to be a truly environmentally friendly technology, the manufacturing and eventual disposal will also have to be environmentally safe and this depends on the materials used and the fabrication technologies. These factors will be considered for each material system.

43.1 Figures of Merit for Solar Cells

In this section we will consider the operation of a photovoltaic cell and the significant parameters that characterize its performance. The objective of a photovoltaic cell is to capture as much of the solar energy as possible and convert this into electrical energy. The solar energy reaching the Earth's atmosphere fits a black-body distribution for a body at 5800 K. This spectrum becomes highly structured, particularly in the infrared part of the spectrum, by absorption bands due to atmospheric gases. By the time the solar radiation reaches the Earth's surface, it no longer fits a black-body distribution [43.2]. The shift in the spectral distribution will obviously affect the efficiency of absorption of the solar radiation, particularly considering that all semiconductor materials will display a cutoff wavelength dictated by the bandgap of the semiconductor. It is necessary to specify the atmospheric absorption when quoting efficiency, as the depth of atmosphere that the solar radiation has passed through will affect both the spectral distribution and the total amount of energy. The measurement used is air mass (AM) and is defined as 0 for solar radiation outside the atmosphere and 1 for radiation reaching the ground when the Sun is at its zenith. For shallower angles the solar radiation has to penetrate a larger depth of atmosphere and the AM is therefore going to be greater than 1. For space applications AM 0 is the appropriate condition and for terrestrial applications (depending on the position on the Earth) AM 1.5 is a typical value quoted, which corresponds to a solar angle of 48°. All this assumes that there is no cloud cover, which will further reduce the integrated intensity and modify the solar spectrum. Although cloud cover will reduce the amount of solar energy available for conversion, useful amounts of electricity can still be generated.

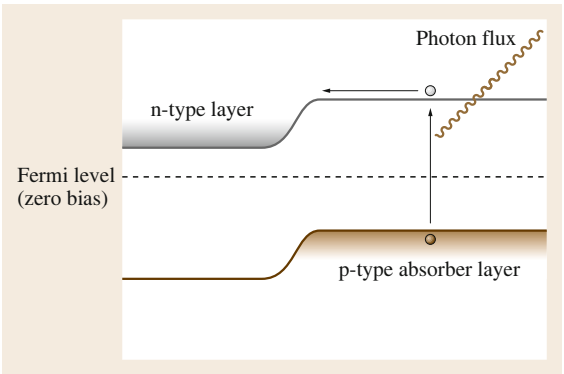


Fig. 43.1 Schematic of an energy-band diagram for a p-n junction solar cell showing photoabsorption to create an electron-hole pair and diffusion of the electron towards the junction

The available solar energy also decreases from $1367\,W/m^2$ outside the atmosphere (AM 0) to $1040\,W/m^2$ for AM 1, when the sun is directly overhead. In practice, for terrestrial applications the available solar energy is considerably less than this where, in general, AM > 1 and is further reduced by cloud cover.

A photovoltaic cell is basically a diode with a photogenerated current. A band diagram is shown schematically in Fig. 43.1. Absorption of radiation can occur on both sides of the junction, creating minority carriers that can diffuse towards the junction. A photocurrent is generated if the minority carriers can drift across the junction without recombination. In practice, the junction is shallow and absorption will occur predominantly on one side where there is a greater depth of absorbing material. There are also heterojunction p-n devices where less absorption will occur in the wider-bandgap layer, as shown in Fig. 43.2. So, one side of the junction is considered to be the absorber layer and it is the spectral absorption characteristics of this layer that will

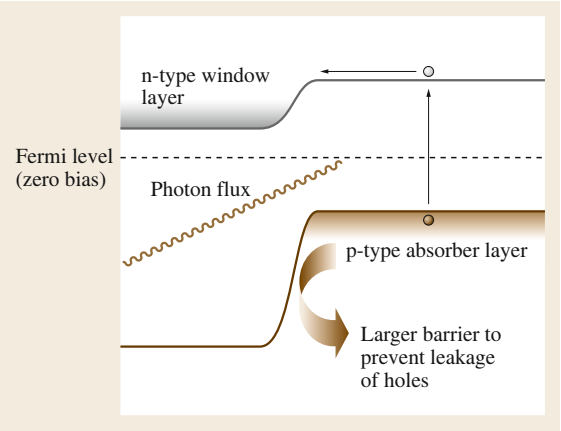


Fig. 43.2 Band diagram for a heterojunction PV cell showing illumination through the wide-bandgap window layer

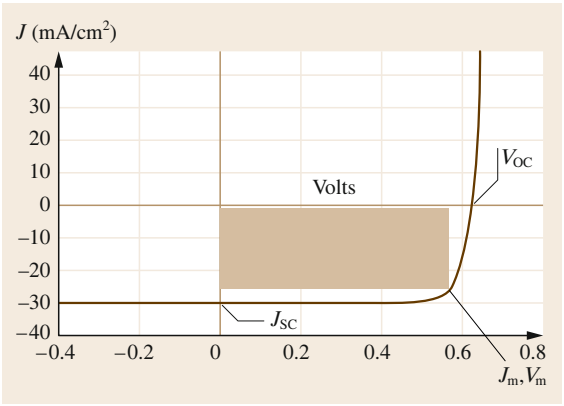


Fig. 43.3 Ideal J–V characteristic for a photovoltaic cell, according to (43.1) with J_s equal to $30\,\mathrm{mA/cm^2}$. The parameters J_{sc} and V_{oc} are indicated on the graph. The maximum extracted power is shown as the *shaded area*

determine the maximum absorption possible from the available solar radiation.

An ideal diode characteristic for an illuminated cell is shown in Fig. 43.3. In the dark the J–V plot would go through the origin and, as the light intensity increases, the short-circuit current becomes increasingly negative, indicating the presence of a photogenerated current. The equation for the J–V characteristic of this ideal device is

$$J = J_s\left(e^{\frac{qV}{k_B T}} - 1\right) - J_L\,, \tag{43.1}$$

where J_s is the saturation current in reverse bias under zero illumination, q is the charge on the carrier, V is the applied voltage, k_B is Boltzmann's constant, T is the temperature of the cell and J_L is the photogenerated current. In the ideal cell this is equal to the short-circuit current, indicated as J_{sc} on the J–V curve in Fig. 43.3. Power can be extracted from the device in the $+V$, $-J$ quadrant of the J–V plot and the load will determine the operating point in this quadrant. The power is determined from the product JV at this operating point, shown graphically in Fig. 43.3. The maximum power will correspond to the operating point that will give the largest JV area on this graph. For the ideal diode characteristics given in (43.1), this maximum power is given by the following (43.2)

$$\begin{aligned} P_m &= J_m V_m \\ &= J_m\left[V_{oc} - \frac{k_B T}{q}\ln\left(1 + \frac{qV_m}{k_B T}\right) - \frac{k_B T}{q}\right] \\ &= J_m\left(\frac{E_m}{q}\right), \end{aligned} \tag{43.2}$$

where E_m is the maximum energy that can be extracted per photon and depends on the band parameters for the semiconductor absorber layer, which determine V_{oc} and V_m. These parameters are marked on the J–V plot in Fig. 43.3. We now have two fundamental parameters which will limit the efficiency of the cell:

- The fraction of solar photons absorbed in the cell
- The electrical energy created per photon.

The first factor can be calculated by integrating over the solar spectrum for the appropriate AM number and including the cutoff wavelength of the semiconductor absorber layer.

$$\eta_{abs} = \frac{\int_{E_g}^{\infty} n_E(E)\mathrm{d}E}{\int_0^{\infty} n_E(E)\mathrm{d}E}\,. \tag{43.3}$$

This is shown graphically in Fig. 43.4 for AM 1.5. Any photons with energy less than the bandgap will not be absorbed and will not contribute to the photocurrent.

The second efficiency factor mentioned above implies that not all the photon energy will be converted into electrical energy, even if one photon absorbed constitutes one minority carrier crossing the junction. The electrical energy per carrier is given by the factor E_m in (43.2), so the maximum power of the device is the product of the absorption rate of photons and the mean electrical energy created per photon. This product is represented by the inner shaded area of the solar spectrum shown in Fig. 43.4. The difference between curve 1 and curve 2 is simply the energy lost per photon because not all the photon energy is converted into electrical energy. Different semiconductor materials will have different efficiencies primarily because of different values for the bandgap. The ideal value for E_m will

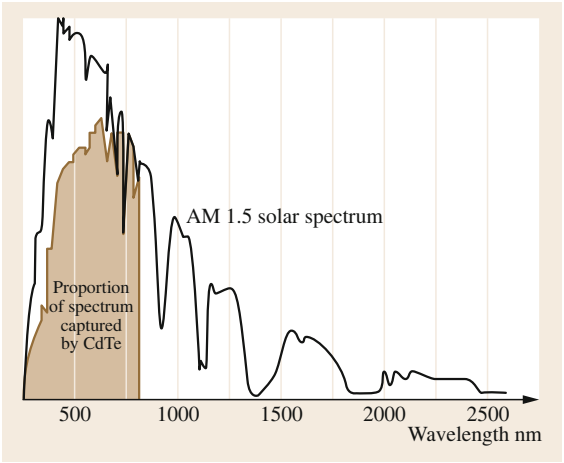


Fig. 43.4 Graphical representation of the maximum energy that can be extracted from a CdTe solar cell with a bandgap energy of 1.45 eV

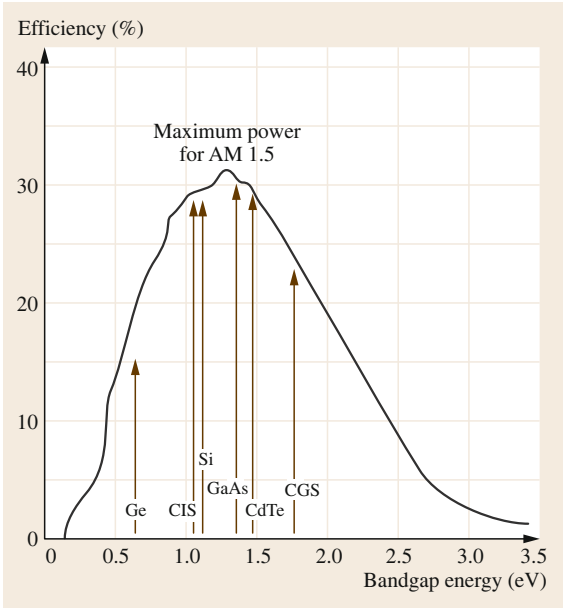


Fig. 43.5 Plot of ideal efficiency against bandgap energy for a single-junction cell for AM 1.5 illumination conditions. (After [43.2])

track the bandgap, so for narrower-bandgap materials there will be a larger proportion of photons absorbed but less electrical energy per photon. The function of efficiency for semiconductors with different bandgaps, taken from curve 2 in Fig. 43.4, is plotted in Fig. 43.5 and shows that the optimum efficiency occurs for semiconductors with a bandgap in the near-infrared region, around 1.5 eV. This represents the best compromise between absorption of solar radiation and transferring the optimum amount of energy per photon into electrical energy.

The maximum efficiency is predicted for Si, InP, GaAs and CdTe, which are in the region of 30% for AM 1.5 irradiation (1 sun illumination) and is called the Shockley–Queisser limit [43.3]. This means that the very best conversion efficiency for a single-junction cell is approximately one third of the available solar energy. In practice, photovoltaic cells have efficiencies considerably less than this due to optical reflections, poor junction characteristics and carrier recombination. These are materials issues that are often traded off against production costs, for example using polycrystalline rather than single-crystal material. Conversely, higher conversion efficiencies can be achieved using multiple junctions, which are more expensive but attractive for space applications and when used in combination with solar concentrators, so less surface area of the expensive multijunction cell is required. Other factors that can influence the choice of photovoltaic materials include the following:

- Absorption coefficient
- Contact resistance
- Abundance of raw materials
- Toxicity of materials
- Stability of materials and junctions
- Radiation resistance.

These factors will be considered in the following examples of different photovoltaic systems to assess the merits of different materials. It is probably true to say that there is not one ideal material but different applications can make one material more attractive than another.

43.2 Crystalline Silicon

Over 85% of the current world production of solar modules are made from either single-crystal or multigrain silicon. This is the most mature of the photovoltaic materials and has benefited considerably from the size of the silicon semiconductor industry. This has ensured a ready supply of material and processing tools suitable for large-scale production. However, crystalline silicon does suffer from a fundamental disadvantage in that the bandgap of silicon is indirect, which means that the absorption coefficient is much lower than

a direct-bandgap semiconductor such as CdTe. In practice this means that a much thicker piece of material is needed to absorb all the solar radiation greater than the bandgap energy. This requires wafers of silicon thicker than 100 μm and means that this material is not suitable for thin-film technology. The highest performance solar electric modules are made from wafers of single-crystal silicon cut from Czochralski-grown crystals, up to 30 cm in diameter. This was originally developed on the back of the silicon electronics industry, but the rapid increase in production volume of PVs is now driving production of silicon crystals. The conversion efficiency for single-crystal PV modules is around 20% with the potential for increases in the near future to 22%.

Multicrystalline silicon involves the relatively cheap path of casting silicon ingots that are not seeded but produce a random grain size of the order of 1 cm across. The ingots are cast in blocks, larger than 100 kg, and then sliced into wafers 300 μm thick and with an area of up to 20 cm^2. The grain boundaries cause these wafers to be mechanically weaker than the single-crystal wafers and they will typically be thicker. This causes some trade off in price. Each cell can be expected to contain grain boundaries, so loss of photogenerated charge at the grain boundary can cause some loss of efficiency. Typical module efficiencies for multicrystalline silicon are currently around 17%.

The junction is formed by phosphorus implantation to form a p-n homojunction and contacted by printing of a metal grid, usually of a Ni-Au alloy. Patterning of the surface prior to implantation and contacting has achieved the highest efficiency cells by improving light collection [43.4]. Another recent innovation that has contributed to higher efficiencies is the *V*-grove or buried junction. This is shown schematically in Fig. 43.6 and entails the grooving of the p-type silicon substrate, implantation of phosphorus to create a buried junction and filling with metal contact alloy. This helps to improve the collection efficiency of the cell, particularly in the blue part of the spectrum. These manufacturing steps have to handle large volumes and be cheap. Cheaper alternatives to ion implantation involve thermal diffusion from a screen-printed paste or spin-on glass [43.5].

The main disadvantage of silicon is that the absorption coefficient is low because it is an indirect-bandgap semiconductor (2×10^3 cm^{-1} for Si compared with 1×10^5 cm^{-1} for CdTe in the green part of the spectrum). This means that the amount of material needed to absorb the solar radiation is greater than for a direct-bandgap semiconductor. The absorption can be improved by having a reflecting back contact that, if it is roughened, will increase the mean path length of

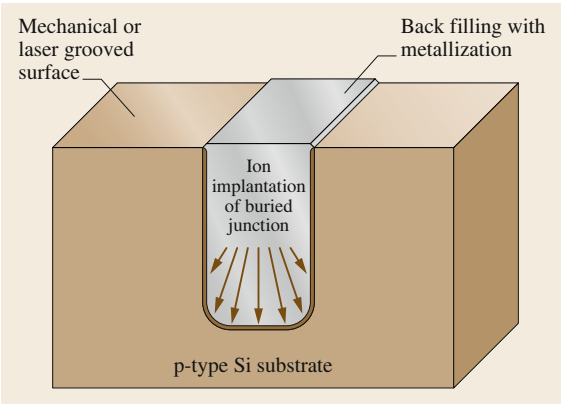


Fig. 43.6 Schematic of buried-contact technology for Si solar cells

the back reflected light in the silicon cell. This is particularly important for thin-film silicon devices where the amount of material is more of an issue, although the thickness of 30–100 μm can be compared with the much thinner layers used for direct-bandgap semiconductors (see CdTe and GaAs).

Thin-film polycrystalline silicon is grown by ribbon casting techniques. The pulling speeds are in the range of 10–1800 cm/min, and this represents a cheap method for the production of solar cell material but the conversion efficiencies are currently low. One way of increasing the absorption coefficient is to deposit a film of amorphous silicon (a-Si), which has a larger bandgap than crystalline silicon but is a direct-bandgap material. Amorphous silicon has led the way in cheap thin-film solar cells but suffers the disadvantages of low efficiency (< 10%) and poor long-term stability. These cells will be discussed in greater detail in the next section.

Crystalline silicon solar cells, over the last ten years, have approximately doubled in efficiency and this has been achieved by a combination of the following:

- Improved material quality, leading to improved minority-carrier diffusion length.
- Improved V_{oc} and fill factor through emitter and base doping and contact optimization.
- Improved J_{sc} through diffusion length improvement using phosphorus gettering, hydrogen passivation and buried contacts.
- Surface passivation and contact grid optimization.

All these aspects involve materials issues, either directly associated with the quality of the silicon or with contacting and passivation. It is also important to avoid degradation of the cell and the final encapsulation must avoid exposure of the cell to water. Multicrystalline silicon can be passivated with silicon nitride deposited by

plasma-assisted CVD (CVD) from SiH_4 and NH_3 to reduce surface recombination [43.6].

Silicon solar modules are becoming cheaper to produce, nontoxic and stability in a nonradiation environment is good. Multicrystalline silicon is not so suitable for space applications because of the combination of thick absorber layers, requiring greater weight per unit area and the sensitivity of the cell to degradation in a high-radiation environment. However, single-crystal silicon has dominated the space market, but it has now been replaced by higher efficiency III-V multijunction cells. Single-crystal modules for terrestrial applications produce the highest available output power with over 300 W (AM1.5) per panel.

43.3 Amorphous Silicon

Amorphous silicon offers the potential for a cheap production technology for terrestrial photovoltaic solar cells. The amorphous state displays different physical properties to the crystalline with a modified band structure. One consequence of this is that the absorption coefficient in the green part of the visible spectrum is a factor of ten higher at $2\times10^4\,\mathrm{cm}^{-1}$, which makes it more suitable for thin-film applications. The amorphous structure leaves dangling bonds, which pin the Fermi level and would normally prevent doping of the material. This is overcome by the inclusion of hydrogen, which passivates these dangling bonds, so the material is referred to as a-Si:H. The Si can also be alloyed with Ge, C, and N in a glow-discharge evaporator. These alloys are particularly useful for multijunction devices used for increasing the solar collection efficiency. In the laboratory single-junction cells have an efficiency around 10% [43.7] with multijunction cells going up to over 13%, where the a-Si junction is combined with nanocrystalline Si junctions [43.8]. In production an a-Si module has an efficiency of 7–10%, considerably less than the crystalline materials.

The most common method for producing a-Si:H for photovoltaics is by plasma-enhanced CVD from SiH_4 mixtures. The films are deposited onto a textured conducting oxide such as indium-tin oxide (ITO), which provides the electrical contact and increases the average light path in the absorber layer to increase absorption. The device structure is a p-i-n with absorption taking place in the middle (insulator) layer, which is only 0.5 μm thick. The p and n layers can be deposited by adding B_2H_6 or PH_3 respectively to the plasma.

One of the major disadvantages of a-Si:H is the instability and long-term deterioration under light illumination. This is caused by the rearrangement of dangling bonds, often associated with rearrangement of hydrogen atoms close to weak or dangling bonds. The energy comes from nonradiative bimolecular reactions and hence depends on the illumination intensity (for further details see the review by *Bloss* et al. [43.9]). In practice, this causes a downward drift in efficiency with time, which can be as much as 2% in 100 h. The cost of a-Si thin-film modules ten years ago was cheaper than crystalline silicon (3 US$/Wp) but has not kept track with falling prices and larger manufacturing volumes of crystalline silicon modules [43.10, 11]. One of the main technical challenges is to maintain stabilized efficiency above 10%. The relatively high cost of PV solar modules compared with more conventional sources of energy can only become attractive if it can operate efficiently over a long period of time. This means that the modules must remain efficient over a period of 20 years. Shorter operating lifetimes would have to be mitigated by much lower production cost.

In conclusion, a-Si:H is a low-cost technology for terrestrial applications and has found its way into low-power and other more bespoke applications such as small-scale standalone systems. The temperatures used in processing these modules is lower than the high temperatures needed to melt silicon for the crystalline silicon modules, which means that the energy payback times and carbon footprint are less. However, longer-term stability will affect the amortized cost of electricity compared with crystalline silicon PV modules.

43.4 GaAs Solar Cells

The III-V semiconductors have advantages over silicon due to their direct-bandgap photon absorption, with an absorption coefficient in the green of $8\times10^4\,\mathrm{cm}^{-1}$. This means that, for example, GaAs can theoretically yield over 30% efficiency (AM 1.5) for an absorber layer thickness of just 1–2 μm, compared with a hundred times this for crystalline silicon. The efficiency, stability and thin-film deposition technology make these cells attractive for space applications. From Fig. 43.5 it is clear that the bandgap of GaAs is well matched to the op-

timum for maximum efficiency. It is also noteworthy that the highest efficiency single-junction solar cell is epitaxial GaAs with 28.8% AM1.5 efficiency [43.12]. The III-V semiconductors also offer the flexibility of alloying to change the bandgap and tune the response of the photovoltaic junction. In addition, heterojunctions can be formed and multijunction solar cells can be produced to convert more of the solar spectrum into electricity and thus exceed the theoretical limits set by the Shockley–Queisser limit. For example, in the laboratory a GaAs/GaSb tandem solar cell was reported with 35.6% efficiency as long ago as 1990 [43.13]. Higher efficiencies are possible with three and even four junctions to capture more of the infrared that would otherwise not be absorbed. The Ge substrate, which is closely lattice matched with GaAs, is commonly used as an infrared absorber to capture the radiation that passes through the $GaInP_2$/GaAs cell [43.14, 15]. The top cell has an absorber of $GaInP_2$, which has a bandgap of 1.9–2 eV, and will capture the visible part of the spectrum without too much voltage loss. The next cell is GaAs, which has a bandgap of 1.42 eV, and will capture the red to infrared part of the spectrum. The bottom junction is formed in the Ge substrate, which has a bandgap of 0.67 eV and will capture the light further into the infrared. The best triple-junction solar cell has achieved 37.9% efficiency under AM1.5 illumination [43.16]. The triple-junction cell, in production for space applications, can yield 29.5% (AM 0) conversion efficiency [43.17].

The layers of these multijunction cells are grown epitaxially onto single-crystal Ge substrates. As with crystalline silicon, high crystalline quality is needed to obtain high efficiency cells; polycrystalline GaAs does not work as well due to grain-boundary conduction that reduces the available photocurrent [43.18]. This restricts the photovoltaic applications to high-quality epitaxial material as GaAs is not suitable for cheap thin-film photovoltaics on glass substrates. The topic of thin-film polycrystalline photovoltaics will be considered in the next section on CdTe solar cells. Epitaxial growth avoids minority-carrier recombination at grain boundaries, by avoiding their formation, but the cost of the substrates is inherently higher than the glass or ceramic substrates used for thin-film devices. However, defects must also be avoided at the junction and interfaces between layers of different composition and this imposes the constraint that the heterostructures should be lattice matched. This restricts the choice of alloys to those that are either lattice matched, or so that the thickness and mismatch are such that the film is strained. The simplest structures use AlGaAs, which has a close lattice match to GaAs over the entire composition range, as a window and passivation

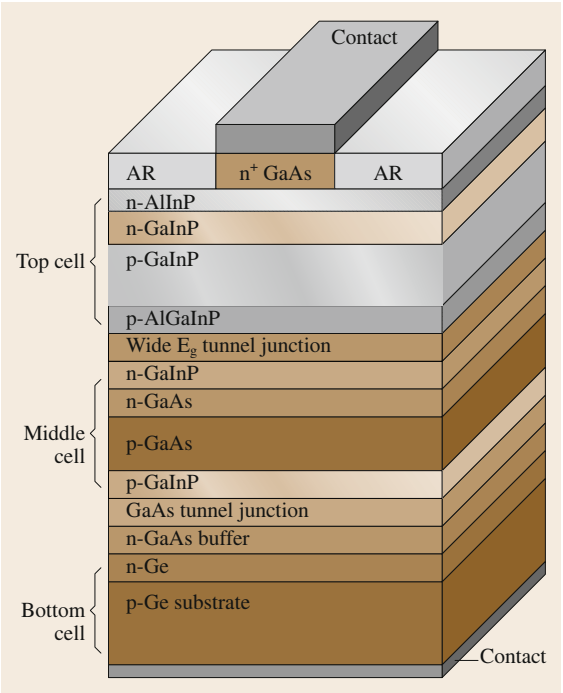


Fig. 43.7 Schematic of a triple-junction GaAs solar cell, showing the positions of three junctions. (After [43.13])

layer. GaAs is also closely matched to Ge, offering a choice of substrates. The GaAs or Ge substrate is the narrowest bandgap part of the structure, so the device has to be front-side illuminated with a grid of metal contacts on the top surface to contact the cells. The design of the cell structure needs to have the wider-bandgap layers further from the substrate (last to be grown) to allow the longer wavelengths to penetrate to the narrow-bandgap absorber layer. This places further constraints on the design of these epitaxial structures.

The Ge and GaAs substrates have a similar lattice parameter to AlGaAs, which is used as a passivation layer for both n-type and p-type GaAs layers. Other III-V compounds that can be used include GaSb, which is sensitive to the near infrared out to 1.8 μm and will therefore absorb the radiation that is transmitted through the GaAs cell, as an alternative to the Ge-bottom cell. The alloy GaInP also lattice matches for the 50% Ga/In mix and is useful as a wider-bandgap junction. GaSb is also attractive for thermo-photovoltaics where the solar radiation is converted into heat, which is then absorbed by the narrow-bandgap GaSb device. The narrow-bandgap quaternary GaInNAs has been investigated [43.19] as a narrow-bandgap absorber layer and can be inserted between the GaAs layer and the Ge substrate to give a small boost in voltage and overcome the

disadvantage of the excess current that would be produced by the Ge bottom cell.

The push to achieving further increases in multijunction solar cell efficiency is driving new approaches to increasing the number of cells without being constrained by lattice mismatch. One of these approaches is the inverted metamorphic structure (IMM) where lattice mismatch is graded out using a buffer layer but to avoid compromising the top GaInP junction with threading dislocations this is grown as an inverted structure onto lattice matched GaAs and then the substrate is removed [43.20]. By this way the highly mismatched bottom junction of $In_{0.27}Ga_{0.73}$ As is grown last but can provide better a bandgap match, giving a boost in efficiency. This is opening the way for better current matching and optimizing the bandgap for increasing the efficiency of multijunction cells and moving towards achieving greater than 50% conversion efficiency.

The III-V photovoltaic structures are grown by metalorganic vapor-phase epitaxy (MOVPE), which gives excellent control over alloy composition and doping concentration. An example of a multijunction cell is shown in Fig. 43.7. The structures tend to be complex because a tunnel junction is needed for current to flow between the two (series) photovoltaic cells, otherwise the connection between the two would be a high-impedance reverse-bias junction. It is also important for the device design to match the currents between the two (or more) junctions so that they can both operate at near-optimum conditions.

GaAs-based solar cells are now dominating the space market to power satellites. Their efficiency is high and stability is good, but the single-crystal Ge substrates and complex epitaxial layer growth leads to high cost. There may be toxicity issues for the disposal of solar modules if GaAs was used widely for terrestrial applications. The most attractive terrestrial application would be in concentrators where the solar radiation is optically concentrated onto the PV cells, so the collection area can be much greater than the area of expensive PV modules. This is looking like an attractive and cost-effective option for regions of the world with a high proportion of direct sunlight [43.21].

43.5 CdTe Thin-Film Solar Cells

The bandgap of CdTe of 1.45 eV at room temperature makes it another semiconductor that is close to the theoretical optimum for efficient conversion of solar radiation (Fig. 43.5) into electrical power. The absorption coefficient is $> 5 \times 10^4\,\mathrm{cm}^{-1}$ for photon energy greater than the bandgap, allowing efficient collection for only 2 μm-thick films, similar to GaAs. CdTe, like GaAs is also a direct-bandgap semiconductor but is from the II-VI family of semiconductor compounds. The theoretical efficiency for the CdS/CdTe photovoltaic cell is ≈ 30% at AM 1.5, the world record efficiency has increased from nearly half this maximum, in 2010 to 21.5% in 2015 [43.22] and for modules in production the efficiency has increased to over 14%. The attraction of CdTe compared with GaAs is that these cells can be made from polycrystalline thin films on glass substrates, thus avoiding the need for expensive single-crystal substrates. A further advantage in using glass substrates is that illumination of the photovoltaic cell occurs through the substrate rather than from the top face, so the substrate becomes the window for the cell. The front contact is made with a transparent conducting oxide (TCO) such as ITO, as shown in Fig. 43.8, and this avoids the need for an opaque grid of metal contacts. This approach of using a TCO on glass substrate as the window is called a superstrate. The TCO has to have high optical transmission as well as a metal-like electrical conductivity. A typical thin-film resistivity would be 10 Ω per square.

The CdTe solar cell is made from a heterojunction between the wider-bandgap CdS and the CdTe absorber layer. CdS is often referred to as the window layer;

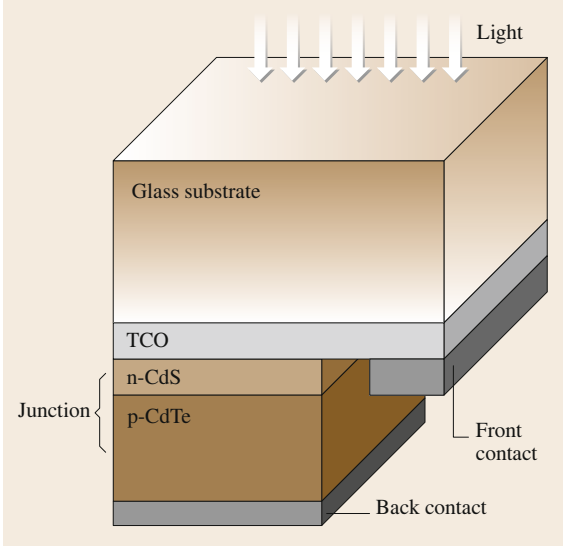


Fig. 43.8 Schematic of a CdTe device structure based on a glass superstrate

it has a bandgap of 2.4 eV allowing most of the visible spectrum to pass with very little absorption. The most common method for depositing CdS is by chemical bath deposition (CBD). The layer is n-type, forming a junction with the p-type CdTe. The thickness is kept to a minimum (around 100–200 nm) in order to minimize absorption at the blue end of the spectrum. A schematic of the CdTe PV structure is shown in Fig. 43.8.

Deposition techniques for CdTe (electrodeposition or close-space sublimation (CSS)) require postdeposition treatment for doping, grain growth and resistivity reduction, at temperatures above 400 °C. The exact processes are not clear but have a role in providing grain boundary passivation, improving efficiency from a few % to > 10% if this anneal is carried out in air in the presence of $CdCl_2$ [43.23]. There are variations in this process in terms of temperature, time and ambient gas but all involve the diffusion of Cl into the CdTe and some interdiffusion of the CdS/CdTe interface. This process improves grain size to $\approx 1\,\mu m$ if the starting material has a much smaller grain size. This is further complicated by the variations in grain size from the CdS interface up to the back surface, where the junction region displays smaller grains. The best efficiency in the laboratory, which stood for ten years since 2001, was 16.5% reported by *Wu* [43.24] and in 2015 rose to 21.5% [43.22].

A typical CdTe module with an area of $0.72\,m^2$ has a peak output power of 110 W; this compares favorably with other thin-film photovoltaics and approaches that of multicrystalline modules [43.25]. The advantage for CdTe thin-film PVs over the maturer crystalline silicon PVs is the smaller temperature coefficient where conversion efficiency decreases with increasing module temperature by −0.5% for crystalline silicon and −0.25% for CdTe modules [43.26]. Future prospects for CdTe can be seen in current research activities with encouraging results for CdTe solar cells on flexible plastic substrates [43.27] and more recently both superstrate on thin polyimide and substrate devices on flexible metal foils [43.28]. Using these substrates places tighter constraints on deposition and annealing temperatures and alternatives to the $CdCl_2$ anneal would be attractive, particularly considering the toxic nature of $CdCl_2$. A nontoxic approach has been proposed by *Major* et al. [43.29] using $MgCl_2$, which achieves the same device improvement as $CdCl_2$ but is nontoxic and costs less than 1% of the cost of $CdCl_2$.

Another area for improvement in CdTe solar cells is in the p-type doping, which has relied on native defects or a controlled Cu diffusion [43.30]. An alternative approach is in situ doping of CdTe with arsenic using metalorganic chemical vapor deposition (MOCVD) [43.31, 32]. Active acceptor doping concentrations as high as $5\times10^{15}\,cm^{-3}$ have been demonstrated and these devices show excellent stability. MOCVD offers greater flexibility over deposition conditions and this has been utilized with CdTe solar cell structures on ultra-thin glass superstrates for ultra-lightweight PV solar cells for space [43.33].

A concern with CdTe is the abundance of tellurium and the volume of available supply to feed a large expansion in the manufacturing of CdTe PV modules. The abundance figures look low but up to 5000 t of Te could be produced each year from copper ore and 20 000 t of Cd is produced each year as a by-product of zinc production. An important environmental and resource issue may be the recycling of these materials. It is important to note that although Cd is a very toxic metal it is relatively benign in the form of CdTe as it is a very stable compound and is not soluble in water. The use of Cd in solar cells has the added benefit of making use of the Cd that is produced as a by-product of zinc production in an environmentally friendly way. *Jager-Waldau* [43.10] states that *Every energy source or product may present some environmental, health and safety hazard, and those of CdTe should by no means be considered barriers to technology scaling*. This technology is less well developed than those in the other case studies but the thin-film technology on glass combined with good stability makes this a very attractive proposition for both terrestrial and space applications.

43.6 $CuInGaSe_2$ ($CIGS_2$) Thin-Film Solar Cells

This is an alternative material system for achieving an optimum bandgap for solar energy conversion and one that has attracted a lot of research effort in the past 20 years with impressive improvements in conversion efficiency. The highest efficiency reported for laboratory solar cells back in 2003 was 19.2%, which then was the best of the thin-film PV technologies [43.34]. This has risen to the current world record in 2015 of 22.3% by the leading CIGS module manufacturer, Solar Frontier [43.35]. In common with other thin-film PV technologies, these films can be deposited onto cheap substrates, at relatively low temperature, and the potential for processing in large volumes. The absorber layer is based on the Chalcopyrite-phase $CuInSe_2$ (CIS). This has a bandgap of 1.04 eV, which is lower than the bandgap for optimum efficiency in

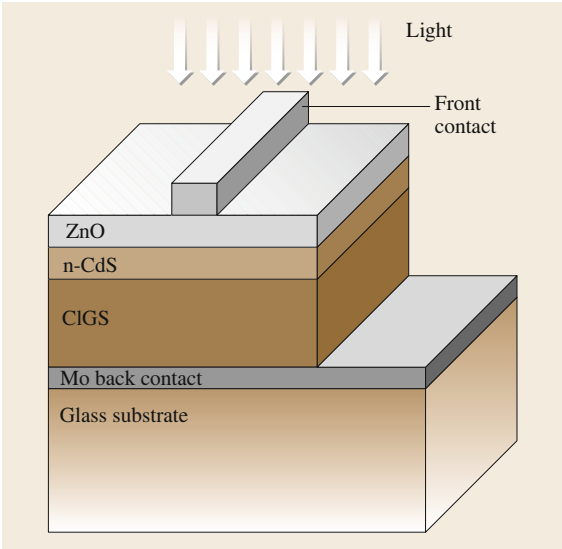


Fig. 43.9 Schematic of CIGS device structure based on a glass substrate

the Shockley–Queisser limit. However, cells made from CIS have achieved > 10% efficiency. Alloying with Ga to form $CuIn_{1-x}Ga_xSe_2$ (CIGS) increases the bandgap to 1.7 eV for $x = 1$. This gives a range of alloy composition across the important optimum range for efficient conversion. An increase in bandgap will increase the voltage of the cell (V_{oc}) but decrease the number of absorbed photons, thus decreasing J_{sc}. In practice the alloy is not uniform and thus a greater proportion of the photon flux is absorbed by the single-junction device than would occur for a single wide-bandgap absorber layer [43.36]. The other benefit of grading of the alloy composition is that the resultant bandgap grading creates a built-in electric field that can drift the electrons towards the junction.

Unlike a-Si and CdTe, the preferred arrangement is the substrate configuration where the cell is illuminated from the top surface as with the crystalline GaAs cells. The typical layer structure is shown in Fig. 43.9. The substrate is soda-lime glass with a sputtered coating of molybdenum, which acts as the back contact. The next layer is the $CuIn_{1-x}Ga_xSe_2$ alloy, which is the p-type absorber layer. The junction is formed with a thin layer of CdS (as with CdTe). The front contact is formed by aluminum-doped ZnO, which is highly conducting but transparent, allowing solar radiation to pass without significant absorption. One disadvantage of the substrate approach is that the top surface is protected with another glass sheet, or other suitable transparent material, when it is fabricated into a solar PV module. This step is not necessary in the superstrate approach. However, the substrate configuration for CIGS has shown much higher maximum efficiency than the superstrate approach (favored by CdTe and a-Si) and provides the opportunity to deposit the CIGS layers onto flexible metal foils. The structure for CIGS on metal foil substrates is similar to those on the Mo/glass substrates except for a barrier layer of SiO_2 between the metal foil and the Mo back contact. This is needed to reduce the diffusion of impurities into the active semiconductor layer. However, one impurity, sodium, is needed to give the desired doping properties for the CIGS absorber layer [43.37]. This impurity occurs naturally when cheap soda-lime substrates are used, as the sodium will diffuse from the glass substrate into the CIGS layer during film deposition and annealing. Using metal foil substrates makes the PV cells flexible and enables their manufacture in a continuous reel-to-reel process.

There are various deposition methods that can be used for the deposition of CIGS and this can lead to low-cost production routes. The early results were obtained by co-evaporation from separate elemental sources. More recently a range of techniques have been used from e-beam evaporation of the metal sources followed by selenisation and annealing to electrodeposition. The characteristic of CIGS synthesis approaches is a separation of the deposition process and alloy formation, allowing cheap and potentially high-throughput techniques to be used. This also enables control of the alloy composition and stoichiometry that will affect the doping [43.38]. The subsequently deposited junction layer, CdS, can be deposited by either chemical bath deposition, sputtering or CVD. The transparent contact layer, ZnO, is typically deposited by sputtering. All these techniques can be operated on a large industrial scale and lend themselves to cheap production methodology.

CIGS solar cells have moved rapidly into production and a number of manufacturers are offering modules with efficiencies in the region of of 14%. For example, Avancis produces a CIGS panel with a peak output power of 140 W for a module area of 1.05 m^2 and Soler Frontier produces modules with a peak power output of 170 W for 1.228 m^2 modules. The potential for further price reduction is similar to other thin-film technologies and competitive with crystalline silicon.

As with CdTe solar cells, the abundance of the elements In and Se appears low, but for a thin-film photovoltaic, where each module only needs a few grams of material, the total amount needed is not huge. However, the current price of In is high as it is not readily produced as a by-product of other mining processes. Price volatility with indium supply could significantly affect the future price of these modules and the cost of the active materials could contribute nearly a half of the

overall module production cost [43.39]. Concerns about materials cost and abundance of materials has stimulated research into alternative chalcogenide absorber materials to replace CIGS in the longer term.

One candidate is the Kesterites copper zinc tin sulfide/selenide (CZTS) with the formula unit Cu_2ZnSnS_4 [43.40]. With a bandgap of 1.45 eV, CZTS has the right characteristics for an efficient single junction absorber material. The synthesis routes are similar to those for CIGS with a preference for formation of the metal precursor through evaporation followed by sulphidisation [43.41]. To achieve the highest conversion efficiencies it has been necessary to alloy with selenium, which reduces the bandgap towards 1.2 eV. The best laboratory efficiency was reported by *Wang* et al. [43.42] with 12.6%, which is well below the CIGS record but shows the potential for high efficiency solar cells using Kesterites. This was achieved through alloying with selenium but it would be advantageous to avoid the less abundance selenium altogether. The best result for selenium-free CZTS is 8.4% reported by *Shin* et al. [43.43] and shows the potential for future development of this material towards low-cost thin-film PV module production.

43.7 Excitonic PV

The drive to find cheaper solutions to the conversion of solar energy has taken research down some interesting and unusual avenues. One of the more successful of the alternative approaches is the dye-sensitized cell, known as the Grätzel cell after its inventor *Michael Grätzel* [43.44]. This uses organic dyes to absorb the solar radiation and transfer electrons to a porous TiO_2 surface, which is effectively the junction. Other approaches have looked at conducting polymers and polymer blends to form photovoltaic junctions. The generic term, excitonic PV, refers to the relatively strong excitonic binding energy of the excited state following photon absorption that binds the electron to the hole. For polymer blend solar cells the dissociation of the exciton is achieved at the interface with a high electron affinity material such as fullerene. The hole is then free to be transported via a hole collector, which is a doped conducting polymer such as PEDOT:PSS [43.45]. The conversion efficiency of this type of solar cell has risen sharply in recent years with the best laboratory efficiency over 10% [43.46]. Organic PV offers the potential for very low-cost production where low temperature processing will also lead to the benefit of a very low carbon footprint.

An exciting recent development with excitonic PV has been the rise of the perovskite absorbers, which were initially used to replace the organic dye in dye-sensitized solar cells [43.14]. A characteristic of excitonic PV is the short excitonic diffusion lengths, so efficient cells require a nanostructured material where an electron collecting interface is never far away. A remarkable property of the methylammonia lead trihalogen perovskites was demonstrated by *Liu* et al. who showed that planar structures could be made, similar to conventional thin-film PV, and achieve high conversion efficiency of over 15% [43.47]. The world record (2015) currently stands at 21%, which was achieved by Michael Grätzel's EPFL team [43.48]. There is still much work to be done on the perovskite solar cells, one of which is to overcome the moisture sensitivity and long term stability. Some of the research is aimed at replacing Pb as the metal center with Sn, which removes any toxicity issues. The potential is for the perovskite solar cells to be used in new applications such as color neutral, partially transparent PV glazing in building-integrated PV [43.49] and many other applications that can take advantage of their unique properties.

43.8 Conclusions

This chapter has introduced some basic concepts about photovoltaic solar cells and examples of some of the more common materials being used for solar-cell production. The area of research is huge with a number of specialist journals dedicated to solar energy and large international conferences held annually. Recent developments with crystalline silicon, III-V heterojunctions, thin-film PV and exciton solar cells have been summarized. These examples show the diversity of materials that can be used for solar PV conversion and different stages of maturity. These approaches will enable very cheap solar PV modules to become a reality in the next five to ten years and also diversify the applications of solar PV.

The largest and most mature production facilities are based on crystalline or multicrystalline silicon.

These modules are also the most efficient for single-junction cells. The most efficient cells are made from multijunction GaAs on Ge substrates and are the most complex and most expensive. Most of the multijunction solar cell production is for the space market but increasingly is showing potential for use with concentrators for terrestrial applications. This is acceptable for standalone solar power generation but the largest challenge for large-scale terrestrial power generation is to integrate photovoltaic modules into buildings. The equivalent of a gigawatt power station would require, at least, of the order of $10^7\,m^2$ of solar modules. In reality it would have to be much larger to account for the average power production being much less than the peak production. A modest installation would have approximately $100\,m^2$ of solar panels so we would need more than 10^5 buildings. So, the goal has to be that every house and building has a PV façade or roof.

Large-scale implementation will require the modules to be architecturally acceptable, in appearance and size, and probably serve multiple functions such as keeping the rain out and heat in, and so on. This need is opening the opportunity for a range of materials and designs and ultimately flexible panels. On top of this the cost per watt of electricity produced needs to come down, probably by a factor of three based on current installed costs. The thin-film technologies are inherently cheaper but much of this cost benefit is not realized at present because the efficiencies and production volumes are lower than crystalline silicon. The current growth in PV solar energy after decades of research has become very rapid and this expansion is set to continue well into the future.

References

43.1 S. Nowak: Report IEA-PVPS T1-27: (International Energy Agency, Paris 2015)
43.2 C.H. Henry: J. Appl. Phys. B **51**, 4494 (1980)
43.3 W. Shockley, H.J. Queisser: J. Appl. Phys. **32**, 510 (1961)
43.4 J. Zhao, A. Wang, M.A. Green: Appl. Phys. Lett. **73**, 1991 (1998)
43.5 L. Pirozzi, G. Arabito, F. Artuso, V. Barbarossa, U. Besi-Vetrella, S. Loreti, P. Mangiapane, E. Salza: Sol. Energy Mater. Sol. Cells **65**, 287 (2001)
43.6 A.G. Aberle: Sol. Energy Mater. Sol. Cells **65**, 239 (2001)
43.7 T. Matsui, H. Sai, T. Suezaki, M. Matsumoto, K. Saito, I. Yoshida, M. Kondo: Development of highly stable and efficient amorphous silicon based solar cells, Proc. 28th Eur. Photovolt. Solar Energy Conf., 2213, Paris (2013)
43.8 S.W. Ahn, S.E. Lee, H.M. Lee: Toward commercialization of triple-junction thin-film silicon solar panel with $>12\%$ efficiency, Proc. 27th Eur. Photovolt. Solar Energy Conf., 3A05.1, Frankfurt (2012)
43.9 N. Bernhard, G.H. Bauer, W.H. Bloss: Prog. Photovolt. Res. Appl. **3**, 3 (1995)
43.10 A. Jager-Waldau: Sol. Energy **77**, 667 (2004)
43.11 A. Goodrich, P. Hacke, Q. Wang, B. Sopori, R. Marg, M. Woodhouse: Solar Energy Mater. Solar Cells **114**, 110 (2013)
43.12 B.M. Kayes, H. Nie, R. Twist, S.G. Spruytte, F. Reinhardt, I.C. Kizilyalli, G.S. Higashi: 27.6% conversion efficiency, a new record for single-junction solar cells under sun illumination, Proc. 37th IEEE Photovolt. Specialists Conf. (2011)
43.13 L.M. Fraas, J.E. Avery: Optoelectron. Dev. Technol. **5**, 297 (1990)
43.14 A. Kojima, K. Teshima, Y. Shirai, T. Miyasaka: J. Am. Chem. Soc. **131**, 6050 (2009)
43.15 N.H. Karam, R.R. King, M. Haddad, J.H. Ermer, H. Yoon, H.L. Cotal, R. Sudharsanan, J.W. Eldredge, K. Edmondson, D.E. Joslin, D.D. Krut, M. Takahashi, W. Nishikawa, M. Gillanders, J. Granata, P. Hebert, B.T. Cavicchi, D.R. Lillington: Sol. Energy Mater. Sol. Cells **66**, 453 (2001)
43.16 K. Sasaki, T. Agui, K. Nakaido, N. Takahashi, R. Onitsuka, T. Takamoto: AIP Procs. **1556**, 22 (2013)
43.17 B.C. Richards, Y. Lin, P. Patel, D. Chumney, P.I.R. Sharps, C. Kerestes, D. Forbes, K. Driscoll, A. Podell, S. Hubbard: Performance and radiation resistance of quantum dot multi-junction solar cells, Proc. 38th IEEE Photovolt. Specialists Conf. (PVSC) (2012) p. 158
43.18 J.R. Sites, J.E. Granata, J.F. Hiltner: Solar Energy Mater. Solar Cells **55**, 43 (1998)
43.19 J.F. Geisz, D.J. Friedman, J.M. Olson, S.R. Kurtz, B.M. Keyes: J. Cryst. Growth **195**, 401 (1998)
43.20 J.F. Geisz, D.J. Friedman, J.S. Ward, A. Duda, W.J. Olavarria, T.E. Moriarty, J.T. Kiehl, M.J. Romero, A.G. Norman, K.M. Jones: Appl. Phys. Lett. **93**, 123505 (2008)
43.21 P. Perez-Higueras, E. Munoz, G. Almonacid, P.G. Vidal: Renew. Sustain. Energy Rev. **15**, 1810 (2011)
43.22 M. Osborne: (PVTECH, 2015) http://www.pv-tech.org/news/first_solar_pushes_verified_cdte_cell_efficiency_to_record_21.5 accessed April 2016
43.23 X. Wu, J.C. Keane, R.G. Dhere, C. Dehert, D.S. Albin, A. Dude, T.A. Gessert, S. Asher, D.H. Levi, P. Sheldon: Proc. 17th Eur. Photovolt. Sol. Energy Conf. II (2001) p. 995
43.24 X. Wu: Sol. Energy **77**, 803 (2004)
43.25 S. Hegedus: WIREs: Energy Environ. **2**, 218 (2012)
43.26 N. Strevel, M. Gloeckler: Photovoltaics Int. **17**, 1 (2012)
43.27 X. Mathew, J.P. Enriquez, A. Romeo, A.N. Tiwari: Sol. Energy **77**, 831 (2004)

43.28 L. Kranz, S. Buecheler, A.N. Tiwari: Solar Energy Mater. Solar Cells **119**, 278 (2013)
43.29 J.D. Major, R.E. Treharne, L.J. Phillips, K. Durose: Nature **511**, 334 (2014)
43.30 L. Kranz, C. Gretner, J. Perrenoud, R. Schmitt, F. Pianezzi, F. La Mattina, P. Blosch, E. Cheah, A. Chirila, C.M. Fella, H. Hagendorfer, T. Jager, S. Nishiwaki, A.R. Uhl, S. Nuecheler, A.N. Tiwari: Nature Commun. **4**, 2306 (2013)
43.31 S.J.C. Irvine, V. Barrioz, D. Lamb, E.W. Jones, R.L. Rowlands-Jones: J. Crystal Growth **310**, 5198 (2008)
43.32 G. Kartopu, L.J. Phillips, V. Barrioz, S.J.C. Irvine, S.D. Hodgson, E. Tejedor, D. Dupin, A.J. Clayton, S.L. Rugen-Hankey, K. Durose: Prog. Photovolt. (2015), doi:10.1002/pip.2668
43.33 S.J.C. Irvine, D.A. Lamb, A.J. Clayton, G. Kartopu, V. Barrioz: J. Electron. Mater. **43**, 2818 (2014)
43.34 K. Ramanathan, M.A. Contreras, C.L. Perkins, S. Asher, F.S. Hasoon, J. Keane, D. Young, M. Romero, W. Metzger, R. Noufi, J. Ward, A. Duda: Prog. Photovolt. Res. Appl. **11**, 225 (2003)
43.35 T. Kamada, T. Yagioka, S. Adachi, A. Handa, K. Fai Tai, T. Kato, H. Sugimoto: Proc. Photovoltaic Specialists Conference (PVSC), 2016 IEEE 43rd, 1287 (2016)
43.36 I.M. Kotschau, G. Bilger, H.W. Schock: Mater. Res. Soc. Symp. Proc. **763**, 263 (2003)
43.37 D. Hermann, F. Kessler, K. Hertz, M. Powalla, A. Schulz, J. Schneider, U. Schumacher: Mater. Res. Soc. Symp. Proc. **763**, 287 (2003)
43.38 S. Hishikawa, T. Satoh, S. Hayashi, Y. Hashimoto, S. Shimakawa, T. Megami, T. Wada: Sol. Energy Mater. Sol. Cells **67**, 217 (2001)
43.39 C. Candelise, M. Winskel, R. Gross: Prog. Photovolt. **20**, 816 (2012)
43.40 K. Ito, T. Nakazawa: J. Appl. Phys. **27**, 4 (1988)
43.41 H. Katagiri, K. Jimbo, S. Yamada, T. Kamimura, W.S. Maw, T. Fukano, T. Ito, T. Motohiro: Appl. Phys. Express. **1**, 2 (2008)
43.42 W. Wang, M.T. Winkler, O. Gunawan, T. Gokman, T.K. Todorov, Y. Zhu, D.B. Mitzi: Adv. Energy Mater. **4**, 1301465 (2014)
43.43 B. Shin, O. Gunawan, Y. Zhu, N.A. Bojarczuk, S. Jay Chey, S. Guha: Prog. Photovolt. Res. Appl. **21**, 72 (2011)
43.44 M. Gratzel: MRS Bull. **30**, 23 (2005)
43.45 J. Nelson: Mater. Today **14**, 462 (2011)
43.46 T. Ameri, P. Khoram, J. Min, C.J. Brabec: Adv. Mater. **25**, 4245 (2013)
43.47 M. Liu, M.B. Johnston, H.J. Snaith: Nature **501**, 395 (2013)
43.48 http://www.solarserver.com/solar-magazine/solar-news/current/2015/kw50/epfl-achieves-21-world-record-efficiency-for-perovskite-solar-pv-cells.html, Accessed December 2015
43.49 G.E. Eperon, V.M. Burlakov, A. Goriely, H.J. Snaith: ACS Nano **8**, 591 (2013)

44. Disordered Semiconductors on Mechanically Flexible Substrates for Large-Area Electronics

Peyman Servati, Arokia Nathan

Low-temperature-thin-film semiconductors and dielectrics are critical for large-area flexible electronics, including displays, smart skins and imagers. Despite the presence of structural disorder, these materials show promising electronic transport properties that are vital for devices such as thin-film transistors (TFTs) and sensors. This chapter presents an overview of material and transport properties pertinent to large-area electronics on mechanically flexible substrates. We begin with a summary of process challenges for low-temperature fabrication of a-Si:H TFTs on plastic substrates, followed by a description of transport properties of amorphous semiconducting films, along with their influence on TFT characteristics. The TFTs must maintain electrical integrity under mechanical stress induced by bending of the substrates. Bending-induced changes are not limited to alteration of device dimensions and involve modulation of electronic transport of the active semiconducting layer.

44.1 **a-Si:H TFTs on Flexible Substrates** 1112

44.2 **Field-Effect Transport in Amorphous Films** 1112
44.2.1 Localized and Extended States 1113
44.2.2 Density of States (DOS) 1114
44.2.3 Effective Carrier Mobility 1114

44.3 **Electronic Transport Under Mechanical Stress** 1118
44.3.1 Thin-Film Strain Gauges 1118
44.3.2 Strained Amorphous-Silicon Transistors 1120

References 1123

There is significant demand for electronic-grade thin-films for a variety of applications such as organic light-emitting diode (OLED) displays and lighting modules [44.1, 2], solar cells [44.3], digital imagers [44.4, 5], and wearable sensors [44.6]. Interest in these materials is driven by the promise of low-cost roll-to-roll manufacturing [44.7] and amenability to novel flexible and stretchable product designs. They enable high-performance electronic devices including thin-film transistors (TFTs), sensors, diodes and p-i-n photodetectors on large-area flexible plastic, paper and fabric substrates. For example, OLEDs have become highly attractive as thin emissive devices due to their fast response, high conversion efficiency, wide viewing angle, thin profile, and compatibility with plastic substrates [44.8]. In high-information-content displays, the OLED must be integrated with a TFT circuit that provides stable drive currents and allows for fast content switching [44.9]. Here, TFTs based on low-temperature a-Si:H (solution-processed/vacuum-deposited) organic semiconductors or metal-oxides [44.10–12] are needed for building the switches and circuits.

In addition to advances in material deposition and integration technologies, significant research efforts aim to improve the lifetime of TFTs and OLEDs in view of instabilities induced by bias stress and ambient exposure. The instability of TFTs stems from structural and interfacial disorder in the semiconductor and insulator layers, and interfaces. This leads to a shift in the threshold voltage ΔV_T under prolonged bias stress in a-Si:H [44.13] and organic [44.14] TFTs as well as photo-bias instability for metal oxide TFTs [44.10, 11]. To deal with the shift in device parameters, smart pixel circuits and driving schemes have been employed for compensation [44.15]. The lifetime of the OLED, on the other hand, can be enhanced by efficient thin-film encapsulation layers [44.16], which are imperative for active matrix organic light-emitting diode (AMOLED) displays on plastic substrates.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_44

This chapter presents an overview of material and transport properties pertinent to mechanically flexible electronics, in which amorphous-silicon is used as the example. We begin with an overview of low-temperature fabrication of high-performance a-Si:H transistors and associated processing challenges. The transport properties of disordered semiconducting films are then presented, along with their influence on TFT characteristics. We also examine the impact of external mechanical stress, induced by bending of the flexible substrate, on the characteristics of thin-film devices, including strain gauges and TFTs. The shifts are not only limited to changes in device dimensions but also to a modulation of electronic transport of the active semiconducting layer.

44.1 a-Si:H TFTs on Flexible Substrates

Compared to conventional (rigid) glass substrates, mechanically flexible substrates are attractive because of their reduced fragility and weight [44.17]. Polyethylene terephthalate (PET) [44.18] and polyimide [44.19] are examples of such substrates. Most plastic substrates have a low glass-transition temperature T_g ($< 250\,^\circ\mathrm{C}$). Consequently, the device processing temperatures must be reduced without compromising the electronic properties of the deposited materials. Despite the adverse effects of reducing the processing temperature, high-performance low-temperature ($< 150\,^\circ\mathrm{C}$) processes have been developed for fabrication of a-Si:H TFTs on flexible substrates [44.18–20], and even lower temperature processes ($75\,^\circ\mathrm{C}$) are under investigation [44.21].

One of the key challenges in low-temperature fabrication is to achieve a high-quality gate and/or passivation dielectric films, including a-SiO_x [44.21] or a-SiN_x:H [44.20] gate and/or passivation dielectric with low defect density, leakage, and interface state density. Nitrogen rich ($x = 1.6$ to 1.7) a-SiN_x:H films with high resistivity ($\rho > 10^{15}\,\Omega\,\mathrm{cm}$) and breakdown field ($> 5\,\mathrm{MV/cm}$) have been deposited using conventional plasma-enhanced chemical vapor deposition (PECVD) at $120\,^\circ\mathrm{C}$ [44.20].

The fabrication challenges are not limited to the electrical quality of the deposited materials. Most plastic substrates have a large coefficient of thermal expansion (CTE) ($> 50\,\mathrm{ppm/^\circ C}$), which is much higher than that of the TFT layers ($\approx 4\,\mathrm{ppm/^\circ C}$) [44.22]. As the deposited films go through a temperature drop ΔT upon cooling from the process temperature [44.2], thermal stress $\sigma_{th} = Y_f \Delta C \Delta T$ develops in the film [44.23]. Here, ΔC is the difference between CTEs of the substrate and the film, and Y_f denotes the biaxial elastic modulus of the deposited film. The thermal stress bends the composite structure, giving rise to handling problems. Thus, the substrate needs to be held flat in a frame or glued to a rigid substrate [44.2]. In addition, since most layers are patterned during processing, the thermal stress exhibits a local pattern-dependent distribution, which leads to problems in alignment of consecutive masks. Lowering the process temperature and choosing a substrate with a smaller CTE, e.g., polyethylene naphthalate (PEN) with a CTE of $13\,\mathrm{ppm/^\circ C}$ [44.22], will help alleviate the thermal stress, and associated alignment issues. It is important to note that thin rigid (steel, aluminum, or glass) substrates are also mechanically flexible and are being considered for device fabrication [44.24]. Although these substrates can withstand higher temperatures, handling issues still remain and depend on the desired degree of mechanical flexibility.

Research on TFT fabrication on flexible substrates have demonstrated successful handling of material processing issues. For instance, TFTs fabricated on plastic substrate at $150\,^\circ\mathrm{C}$ have been reported to have parameter values very similar to the high-temperature counterparts in terms of low reverse and high on current (with switching ratio of $> 10^8$), reasonable field-effect mobility $> 0.8\,\mathrm{cm^2/(V\,s)}$, threshold voltage of $\approx 1{-}2\,\mathrm{V}$, and low subthreshold slope $< 0.3\,\mathrm{V/dec}$ [44.20].

44.2 Field-Effect Transport in Amorphous Films

The electronic properties of a semiconducting film are primarily determined by its structural disorder and grain size, which strongly depend on deposition conditions and postprocessing treatment [44.25].

The structure of disordered semiconductors such as a-Si:H lacks the long-range periodicity of crystalline silicon. However, the atoms of the materials bond to their neighbors according to the 8-N rule [44.26]

for both amorphous and crystalline silicon, leading to short-range structural order. The extent to which this short-range order is conserved is dependent on processing conditions and leads to different degrees of disorder. This range can be as small as a few nanometers in amorphous semiconductors or as large as a few tens of microns in polycrystalline material. Nanocrystalline (nc) and microcrystalline (μc) materials reside between these extremes depending on grain size and stoichiometry.

The observation of an absorption gap and thermally activated conductivity in amorphous semiconductors stems from the short-range order [44.27] and resembles the band-gap properties of an ideal crystalline semiconductor, which has no states in the gap. The edges of the gap are well defined and represent discontinuities in the density of states (DOS). In contrast, the lack of long-range order in the amorphous semiconductor leads to the presence of states in the gap, and thus, tailing of the band edges.

44.2.1 Localized and Extended States

In the ideal crystalline semiconductor, the band states are all extended with a corresponding band mobility μ_{band}. In the case of an amorphous semiconductor, however, the structural disorder of the material leads to the presence of localized states for which the mobility vanishes at $T = 0$. *Anderson* [44.28] demonstrated that, for a sufficiently large degree of disorder, all states in the band are localized. In his work, the disorder in the material is modeled by a random but uniform distribution U_0 for the energy of the wells at lattice sites. He found that all of the states of the energy band become localized when $U_0/B > 5$, where B is the band width and is given by $B = 2ZI$ [44.26]. Here, Z is the coordination number, and I the overlap energy for the states in neighboring wells, which decreases exponentially with distance based on the tight-binding approximation.

For a small degree of disorder, Anderson's criteria may not hold for states with large overlap energies (i. e., the extended states). *Mott* and *Davis* [44.26] stipulated that, for a given value of U_0, the states can be localized or extended, depending on their overlap energy. Consequently, for some materials there are distinct energies which separate the localized states from the extended ones and can be determined from Anderson's criteria $U_0 \approx 5B$. These energies are referred to as mobility edges (or mobility shoulders) due to the significant change in the mobility expected at these edges. Figure 44.1 illustrates the DOS $g(E)$ for the mobility gap of a-Si:H indicating the extended and localized states. The mobility shoulder E_C of the conduction band separates the localized states from the extended ones. A similar band structure and mobility shoulder is present for the valence band. This graph is compiled from the work of *Cohen* et al. [44.29], who first demonstrated that the change in mobility at these edges can be of the order of 10^3–10^4, signifying the different transport mechanisms on the two sides of the edge.

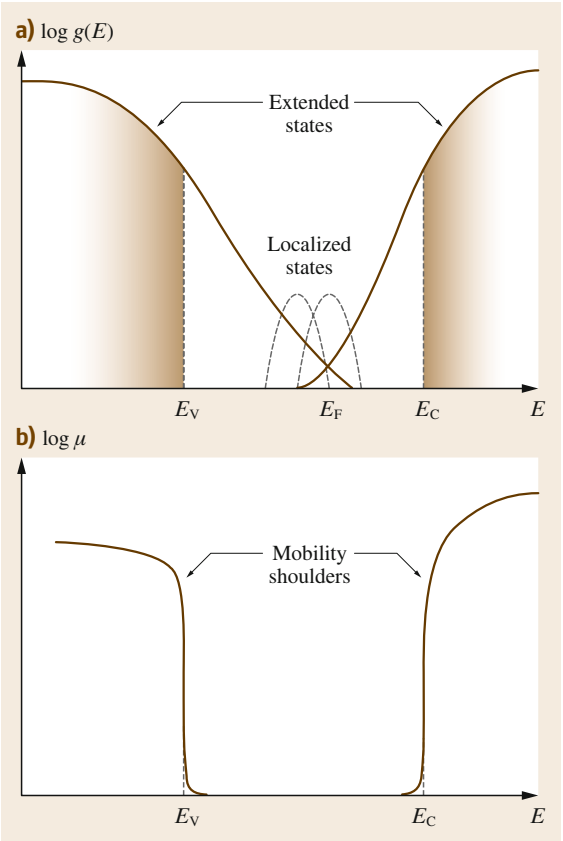


Fig. 44.1 Sketch of (a) DOS in valence and conduction bands, and (b) hole and electron mobilities (After [44.29])

In the extended states, the carriers have a finite mobility, and thus take part in conduction. This prevails at room temperature for a-Si:H and takes the form $\sigma_0 = 0.06q^2/\hbar r(E)$ [44.26], where q is the elementary charge and $r(E)$ the mean distance between the wells that contribute to states with energy E. In the conduction band, $r(E)$ will be close to the lattice constant, which yields $\sigma_0 = 350\,\Omega^{-1}\,\text{cm}^{-1}$. For Fermi energies $E_F < E_C$, the conductivity depends on the Fermi energy and reads

$$\sigma = \sigma_0 \exp\left(\frac{E_F}{k_B T}\right), \tag{44.1}$$

where k_B is Boltzmann's constant and T the temperature, predicting Arrhenius behavior for the conductivity.

Here, E_F is defined as negative for the traps with respect to the conduction band edge E_C. This change in conductivity is due to a change in the concentration of carriers, excited to the extended states n_{band} with a band mobility μ_{band} such that

$$\sigma = q\mu_{band}n_{band}\,, \tag{44.2}$$

where

$$n_{band} = N_b \exp\left(\frac{E_F}{k_B T}\right), \tag{44.3}$$

with $N_b = \sigma_0/q\mu_{band}$. From the given value for σ_0, the band mobility can be found as $\mu_{band} \approx 12\,cm^2/(V\,s)$ for a-Si:H.

In the localized states, the conduction is only possible by hopping of the carriers to the neighboring localized states [44.30]. This method of conduction prevails at low temperatures ($T < 200\,K$) and for highly disordered materials.

The picture of the mobility gap can be completed by adding defect states in the middle of the gap, as indicated by the dashed lines in Fig. 44.1a. These states are attributed to the broken or unsaturated bonds such as dangling bonds in the network structure. The density of dangling bonds is reduced to 10^{15}–$10^{16}\,cm^{-3}$ in a-Si:H by hydrogen atoms, which passivate some of these unsaturated bonds [44.27, 31].

44.2.2 Density of States (DOS)

The density of states (DOS) $g(E)$, including the density of band tail and defect states, as schematically represented in Fig. 44.1a, determines the transport properties of the disordered semiconductor [44.32]. The DOS in a-Si:H has been extensively studied by using different experimental techniques such as field-effect measurements [44.33, 34], photoconductivity measurements [44.35], deep-level transient capacitance spectroscopy (DLTS) [44.36], and capacitance–voltage–frequency (C–V–f) characteristics [44.37, 38]. It is found that an exponential distribution of the deep and tail states [44.39] can efficiently describe the distribution of the localized states and hold for a wide range of materials with rapidly changing DOS. In the case of a-Si:H, the localized states in the upper half of the mobility gap, closer to E_C, behave as acceptor-like states, while the states in the lower part of the gap, closer to E_V, behave as donor-like states. Acceptor-like states are neutral when they are empty, and negatively charged after capturing an electron, whereas the donor-like states are positively charged when they are empty and neutral after capturing an electron.

In a-Si:H, the number of donor-like states closer to the valence band is much higher than the number of acceptor-like states. As a result, following the neutrality condition, the position of the Fermi energy E_F in an intrinsic a-Si:H sample in the dark lies closer to E_C due to the asymmetrical DOS distribution [44.31]. This results in a much stronger electron conduction in a-Si:H, which signifies the role of the conduction band tail in dispersive electron transport. The density of states for the conduction band tail $g(E)$ is written as

$$g(E) = \frac{N_t}{k_B T_t} \exp\left(\frac{E}{k_B T_t}\right), \tag{44.4}$$

where N_t is the total acceptor-like states in the conduction band tail, and T_t the associated slope of the exponential state distribution.

44.2.3 Effective Carrier Mobility

Due to the high density of tail and deep states, only a small number of carriers are thermally excited to the extended states and contribute to conduction as described in (44.2). The total carrier density n is the sum of the excited and trapped carriers such that

$$n = n_{band} + n_t\,. \tag{44.5}$$

To obtain the trapped carrier density n_t as a function of the Fermi energy E_F, we integrate the product of the DOS and the probability of occupation of a state $f(E)$ over the mobility gap

$$n_t = \int_{E_V}^{0} g(E) f(E) \mathrm{d}E\,. \tag{44.6}$$

In equilibrium, the probability of occupation can be described by the Fermi–Dirac function,

$$f(E) = \frac{1}{1 + \exp\left[(E - E_F)/k_B T\right]}\,. \tag{44.7}$$

Equation (44.6) can be numerically solved and approximated along the lines given by *Shaw* and *Hack* [44.40] as

$$n_t(E_F) = N_t \exp\frac{E_F}{k_B T_t} u\left(\frac{T_t}{T}, E_F\right). \tag{44.8}$$

The underlying assumption is that E_F moves no closer than a few $k_B T$ to the mobility edge. This is true because of the high density of tail states, which tends to pin the movement of the Fermi energy. Here,

$u(T_t/T, E_F)$ represents the changes in the trapped-carrier density with normalized ambient temperature T_t/T. For $T \ll T_t$, $u(T_t/T, E_F)$ is often approximated by $[\sin(\pi T/T_t)/\pi T/T_t]$ [44.41] with a value close to 1. The ratio T_t/T, referred to as the dispersion parameter, characterizes the dispersive transport of electrons in the conduction band tail, and can be obtained from the time dependence of the electron drift mobility in time-of-flight experiments [44.42]. The presence of a high trapped-carrier density in amorphous semiconductors leads to an effective trapped-carrier mobility that is lower than the band mobility. *Street* [44.31] has defined the drift mobility μ_D of the carriers as the band mobility reduced by the fraction of time that the carrier is trapped,

$$\mu_D = \mu_{band} \frac{\tau_{band}}{\tau_{band} + \tau_{trap}} , \qquad (44.9)$$

where τ_{band} and τ_{trap} are the times that carriers spend in the extended and localized states, respectively.

Carriers accumulated by the field effect in TFTs also demonstrate similar transport properties. The field-effect mobility μ_{FE} in TFTs is conventionally defined as

$$\mu_{FE} = \frac{g}{qn} , \qquad (44.10)$$

where g is the conductance of the film, and qn is the field-induced charge in the semiconducting film. The terms g and n are averaged over the device active area and are related to the number of band carriers n_{band} and trapped carriers n_t. The μ_{FE} is conventionally retrieved from measurement of the transistor current ($I_{DS,lin}$) in the linear regime ($V_{DS} = 0.1$ V), by using the following

$$\mu_{FE} = \frac{L}{WC_iV_{DS}} \frac{\partial I_{DS,lin}}{\partial V_{GS}} , \qquad (44.11)$$

where W and L are the channel width and length, respectively, C_i the gate capacitance, and V_{GS} and V_{DS} are the gate–source and drain–source biases, respectively.

According to (44.3) and (44.8), the densities of the band and trapped carriers increase differently with increasing E_F. Consequently, when n_{band} and n_t are averaged over the volume of the semiconducting film to obtain the μ_{FE} described by (44.10), the mobility becomes a function of device parameters (e.g., layer thicknesses and bias conditions).

The bias dependence is also evident in temperature-dependence measurements of mobility and conductivity in a-Si:H and organic TFTs. Figure 44.2 illustrates the temperature dependence of μ_{FE} for a-Si:H, pentacene and polythienylene vinylene (PTV) [44.43],

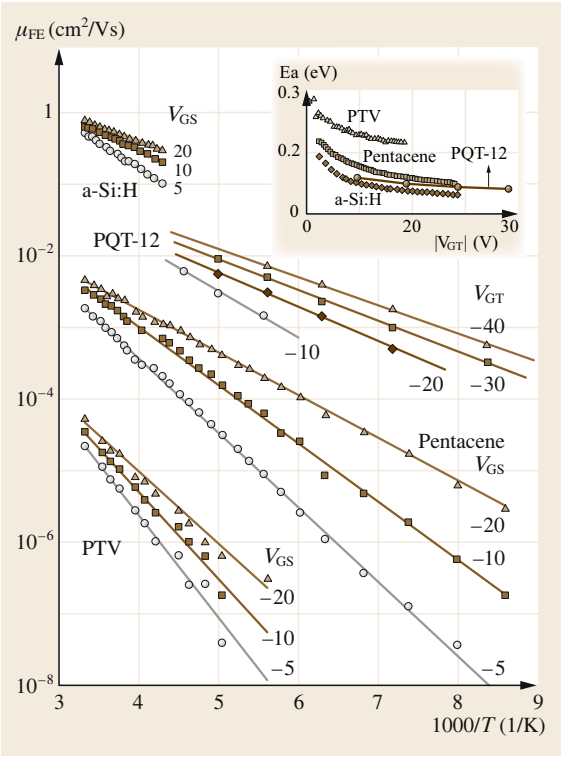


Fig. 44.2 Temperature dependence of μ_{FE} for a-Si:H, PQT-12 [44.44], pentacene and PTV [44.43]. *Inset* shows the bias dependence of the activation energy E_a of μ_{FE} in these materials

and poly[5,5′-bis(3-alkyl-2-thienyl)-2,2′-bithiophene)] (PQT-12) [44.44, 45] TFTs at different gate biases. Here, the activation energy E_a turns out to be bias-dependent (inset of Fig. 44.2). More importantly, an anomaly arises in which the μ_{FE} and E_a become dependent on the gate capacitance C_i, which is solely a geometrical capacitance. A higher C_i implies higher carrier accumulation and consequently higher μ_{FE} for the same bias.

To remove this anomaly, we need to identify the effective carrier mobility such that $\mu_{FE} = \mu_{eff} \times f(\phi)$, where μ_{eff} is the effective physical mobility and $f(\phi)$ describes the device attributes such as bias and geometry.

We reconsider the densities of trapped and free carriers according to (44.3) and (44.8). Since both densities vary exponentially with the Fermi energy, the density of band carriers in terms of trapped carriers can be written as

$$n_{band} = \theta n_t^{T_t/T} , \qquad (44.12)$$

where $\theta = N_b/(N_t u)^{T_t/T}$. Equation (44.12) has been found to hold empirically for a wide range of disor-

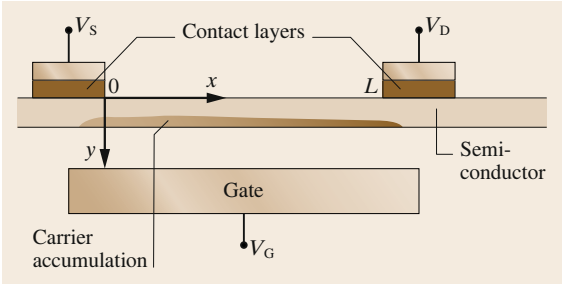


Fig. 44.3 Simplified schematic of a TFT

dered semiconductor systems and operating conditions, including organic semiconductors [44.46, 47].

According to (44.2) and (44.12), the conductivity $\sigma(n)$ as a function of carrier concentration can be written as

$$\sigma(n) = \frac{q}{\mu_{\text{eff}} N_0} \left(\frac{n}{N_0}\right)^{T_t/T} , \tag{44.13}$$

where μ_{eff} is the effective mobility defined at a reference concentration as

$$\mu_{\text{eff}} \equiv \frac{\sigma(n = N_0)}{qN_0} = \mu_{\text{band}} \frac{N_b}{N_0} \left(\frac{N_0}{N_t u}\right)^{T_t/T} . \tag{44.14}$$

Consequently, the conductivity at any carrier concentration is given by μ_{eff} and T_t/T. Here, μ_{eff} represents the effective carrier mobility at concentration N_0 and T_t/T describes the change in conductivity with carrier concentration.

The significance of this representation becomes clear when we use it to obtain the current–voltage characteristics of a TFT (Fig. 44.3). Using the gradual channel approximation, we can write

$$I_{\text{DS,lin}} = \frac{W}{L} V_{\text{DS}} \int_0^{\delta} \sigma(y) \mathrm{d}y , \tag{44.15}$$

where y denotes the location across the channel and δ the channel depth. Using (44.13) for σ and changing the integral parameter from y to n, we find after mathematical manipulation [44.48]

$$I_{\text{DS,lin}} = \mu_{\text{eff}} \zeta \frac{W}{L} (C_i V_{\text{GT}})^{\alpha - 1} V_{\text{DS}} , \tag{44.16}$$

where $V_{\text{GT}} = V_{\text{GS}} - V_{\text{thr}}$, V_{thr} the threshold voltage, $\alpha = 2T_t/T$ the saturation current–voltage characteristics power parameter, and

$$\zeta = \frac{(2\epsilon k_B T_t N_0)^{1-\alpha/2}}{\alpha - 1} . \tag{44.17}$$

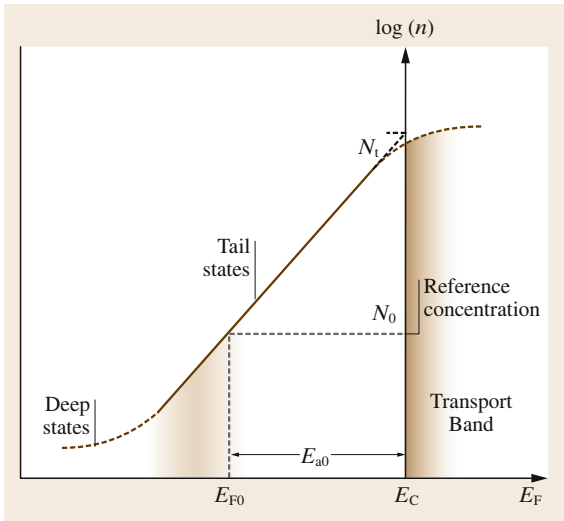


Fig. 44.4 Density of trapped carriers as a function of the Fermi energy, showing the reference concentration N_0 and transport (conduction) band

Here, ζ is just a function of T_t/T and accounts for the carrier distribution across the film. Revisiting the definition of the field-effect mobility as given by (44.11), but this time employing a more meaningful representation of the current, i. e., (44.16), we find

$$f(\phi) \equiv \frac{\mu_{\text{FE}}}{\mu_{\text{eff}}} = \zeta(\alpha - 1)(C_i V_{\text{GT}})^{\alpha - 2} . \tag{44.18}$$

Representation of μ_{eff} according to (44.14) is valid only when there is an exponential relationship between the carrier concentration n and the Fermi energy E_F as given by (44.8). This is true for Fermi energy locations below the transport band edge and above the energies of the deep states (the region shown by the solid slope in Fig. 44.4). Since the deep states mostly contribute to the threshold voltage V_{thr} and are filled before the device turns on, the mobility definition of (44.14) is valid for the above-threshold regime. Consequently, the value of the reference concentration N_0 must be selected such that the Fermi energy E_{F0} associated with N_0 resides well above the deep states. This requires the charge accumulated in the channel Q_{channel} to be higher than the charge $C_i V_{\text{thr}}$ needed to turn on the device. If the carrier concentration at the semiconductor interface is N_0, $Q_{\text{channel}} = C_i V_0$, where $V_0 = (2\epsilon k_B T_t N_0)^{1/2}/C_i$. Thus, for $Q_{\text{channel}} > C_i V_{\text{thr}}$, we conclude $V_0 > V_{\text{thr}}$ or $N_0 > C_i^2 V_{\text{thr}}^2/(2\epsilon k_B T_t)$, indicating the lower limit for N_0. Using typical values, $V_{\text{thr}} = 2\,\text{V}$ and $C_i = 20\,\text{nF/cm}^2$, we have $N_0 > 6 \times 10^{16}\,\text{cm}^{-3}$. For instance, with $N_0 = 10^{17}\,\text{cm}^{-3}$, we have for the bias dependence factor $\zeta = \left(4.1 \times 10^{-16} \alpha\right)^{(1-\alpha/2)}/(\alpha - 1)$.

Table 44.1 Extracted transport parameters (μ_{eff}, T_t, and E_{a0}) for a selection of disordered semiconductors at $N_0 = 10^{17}\,\text{cm}^{-3}$

Semiconductor	μ_{eff} at 300 K ($\text{cm}^2/(\text{V s})$)	T_t (K)	E_{a0} (eV)	Reference
a-Si:H	0.98	350	0.23	Our measurements
PQT-12	3.9×10^{-2}	326	0.24	[44.44]
Pentacene	9.0×10^{-4}	385	0.31	[44.43]
PTV	1.1×10^{-5}	380	0.42	[44.43]

We now revisit the temperature dependence of μ_{eff} by recasting (44.14) in the following manner

$$\mu_{\text{eff}} = \mu_{\text{eff0}} \exp\left(\frac{-E_{a0}}{k_B T}\right), \tag{44.19}$$

where

$$\mu_{\text{eff0}} \equiv \mu_{\text{band}} \frac{N_b}{N_0} \quad \text{and}$$
$$E_{a0} \equiv k_B T_t \ln\left(\frac{N_t u}{N_0}\right) = -E_{F0}\,. \tag{44.20}$$

Equation (44.19) predicts Arrhenius behavior for the effective mobility with an activation energy of $E_{a0} = -E_{F0}$ (Fig. 44.4). The activation energy of μ_{eff} is not bias-dependent and corresponds to the energy needed for carriers to thermalize from E_{F0} to the mobility edge.

We now examine the relation between the two activation energies E_a and E_{a0}. To do so, we look at the temperature dependence of μ_{FE} and μ_{eff}, viz.,

$$\frac{\partial \mu_{\text{FE}}}{\mu_{\text{FE}} \partial T} = \frac{\partial \mu_{\text{eff}}}{\mu_{\text{eff}} \partial T} + \frac{\partial f(\phi)}{f(\phi) \partial T}$$
$$= -\frac{E_{a0}}{k_B T^2} + \frac{2T_t}{T^2} \ln\left|\frac{V_{GT}}{V_0}\right|, \tag{44.21}$$

which yields

$$E_a = E_{a0} - 2k_B T_t \ln\left|\frac{V_{GT}}{V_0}\right|. \tag{44.22}$$

Equation (44.22) describes the bias dependence observed for E_a in the inset of Fig. 44.2.

Amorphous Organic Semiconductors

The concept of effective mobility and transport band may be generalized to accommodate different amorphous organic semiconductors despite differences in the underlying transport mechanism. This generalization follows from the relation $n_{\text{band}} = \theta n^{T_t/T}$ (44.12) for organic semiconductors for a wide range of temperatures and carrier concentrations [44.46]. For this relation to hold, the distribution of trapped and band carriers must be exponential. Although evidence of a Gaussian trap distribution has been reported for organic materials [44.49], the Gaussian distribution is effectively seen as an equivalent exponential distribution due to the small variation in the Fermi energy because of the large tail state distribution. In addition, *Shapiro* and *Adler* [44.50] have demonstrated that a transport band is present in which hopping conduction dominates irrespective of the position of Fermi energy. Similar to the mobility edge, the trapped carriers are thermalized to the hopping band [44.30]. Relative to E_{F0}, the hopping band is located at

$$E_{0\text{hopping}} = k_B T_t \ln\left[\frac{\gamma^3}{N_0}\left(\frac{2T}{3T_t}\right)^3\right], \tag{44.23}$$

where γ is the effective overlap parameter for electronic states in the band tail. *Baranovskii* et al. [44.51] have generalized this concept of transport band beyond the exponential DOS assumption and to a broader range of disordered materials with Gaussian or similar rapidly changing distributions. The generalized band concept can also accommodate the percolation-based hopping transport described by *Vissenberg* and *Matters* [44.43] for amorphous organic semiconductors, which predicts a hopping band that is just $k_B T_t \ln\left(B_C/3\pi^3\right)$ higher than that predicted by *Monroe* [44.30]. Here, $B_C \approx 2.8$ is the critical number for percolation in three-dimensional amorphous systems. Table 44.1 summarizes the values for μ_{eff} at room temperature T_t, and E_{a0} at $N_0 = 10^{17}\,\text{cm}^{-3}$, determined from the results presented for different disordered materials reported in literature.

Amorphous Oxide Semiconductors

Amorphous oxide semiconductors, such as indium zinc oxide (IZO) or the ternary indium gallium zinc oxide (IGZO), have high band gap and strong ionic electronic structure that can exhibit high mobility ($> 5\,\text{cm}^2/(\text{V s})$) even when fabricated at low temperatures. The high mobility and reduced sensitivity to disorder is attributed to the unique role of the spherical 5s orbitals in the conduction of these materials. While chemical bonds in covalent semiconductors such as amorphous silicon are made of sp^3 or p orbitals, strained bonds form deep and high density localized states below the conduction band and above the valence band resulting in formation of tail states [44.52, 53]. On the other hand the electronic structure in oxides benefits from its strong

ionicity. The conduction band minima are made up of spherically extended s orbitals of metal cations and their overlaps with neighboring metal s orbitals are not significantly altered in the disordered oxide. Therefore the electronic levels are not very affected by structural randomness. Carrier transport benefits from the strong ionicity in which charge transfer occurs from metal to oxygen atom. The conduction band minima is almost entirely localized on the metal s orbital and the valence band consists of oxygen 2p states. However, non-bonding state of the metal cation or oxygen defect can be formed in or at conduction band minima in the non-compensated system, allowing vacancies, for example, to act as shallow donors but not as an effective electron trap [44.52]. But these states are not stable; their level relaxes to form either shallow donors or deep occupied states. The presence of oxygen defects leads to persistent conductivity [44.54, 55] behavior, in which the post-stress relaxation of the threshold voltage in an oxide TFT shows an initial rapid recovery followed by a slow-decaying component. This is indicative of a slow- decaying recovery process. The relaxation behavior of the sub-threshold slope is also similar. Here, the slow-decaying component points towards an increased density of shallow states near or above Fermi level. To restore its initial state a positive gate pulse can be applied. This has demonstrated to maintain high frame rates in displays and imagers [44.56].

In multi-component oxides, trap-limited conduction (TLC) and percolation are observed to play a role at low and high gate voltages, respectively [44.57, 58]. The percolation arises from the presence of potential barriers above the conduction band, which lead to spatially distributed perturbations in the minima of conduction band having an average potential barrier of ϕ_{B0} and variance of σ_B. As a result, based on the effective mobility model, the activation energy of TLC conduction for these materials can be written as

$$E_{0TLC} = -E_{F0} - q\phi_{B0} + \frac{(q\sigma_B)^2}{2}\,, \qquad (44.24)$$

At high voltage, as the Fermi energy gets close to the mobility edge, the mobility approaches saturation and percolation dominates the characteristics. A new power law dependence for current voltage characteristics sets in. In contrast, at low temperatures variable range hopping plays a critical role in determining the transport in these materials [44.59].

44.3 Electronic Transport Under Mechanical Stress

Mechanical stress deforms the structure of the thin film leading to modulation in carrier mobility and density of states, and consequently, modulation of resistance. The change in resistance of a solid with elastic strain or stress is commonly referred to as the *elastoresistance* or *piezoresistance* effect, respectively [44.60]. The magnitude of the change is a function of the electronic properties of the material, the dimensions of the solid, and the direction of current flow.

Structural order in the material influences the elastoresistance effect. In the case of crystalline silicon, anisotropic scattering of electrons in the n-type material leads to a strong orientation dependence of the elastoresistive behavior [44.61]. In polycrystalline silicon, the crystallite size and orientation, and material texture play a critical role in determining the magnitude of the effect [44.62, 63]. In amorphous semiconductors, the random network behaves like an isotropic medium, and the anisotropy found in the crystalline material is less visible. However, the elastoresistance coefficients still depend on the relative orientation of the current and applied strain [44.64].

In sputtered amorphous silicon, *Welber* and *Brodsky* [44.65] have reported a decrease in the absorption gap with hydrostatic pressure with a coefficient of about $-1\,\mathrm{meV/kbar}$ ($-1\times10^{-11}\,\mathrm{eV/Pa}$) observed from the shift in the absorption edge. *Weinstein* [44.66] has also reported similar results drawn from photoluminescence experiments. The change in the optical gap is similar in sign and magnitude to that measured for crystalline silicon [44.67]. *Lazarus* [44.68], however, has reported an exponential increase in the resistivity of a-Si:H with increasing hydrostatic pressure at room temperature. The increase in resistivity is ascribed to either a decrease in the number of carriers or a reduction of the mobility with compressive strain [44.64].

In this section, we investigate the impact of mechanical stress on electrical properties of thin-film devices, insight into which is critical for design of mechanically flexible electronics. We begin with metallic and semiconductor thin films for strain-gauge applications and continue with a-Si:H TFTs.

44.3.1 Thin-Film Strain Gauges

Strain gauges are thin-film transducer elements embedded on a substrate Fig. 44.5. The strain in the substrate

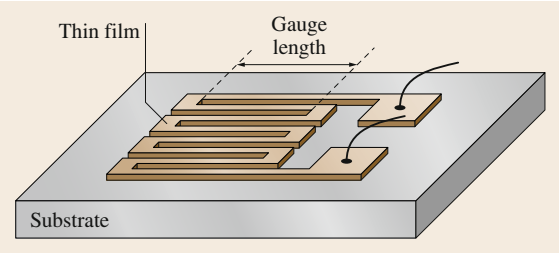


Fig. 44.5 Schematic of a thin-film strain gauge

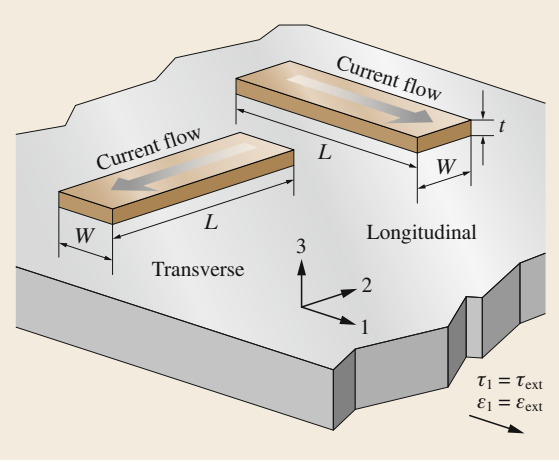


Fig. 44.6 Longitudinal and transverse wires under uniaxial stress and strain components in the 1-direction

leads to a change in the geometry, and therefore, the resistance of the gauge, which is detected and measured by external circuitry. The sensitivity of the resistance R to strain ϵ is referred to as the gauge factor k, which can be written as

$$k \equiv \frac{\Delta R}{\epsilon R} = \frac{\Delta L}{\epsilon L} - \frac{\Delta W}{\epsilon W} - \frac{\Delta t}{\epsilon t} + \frac{\Delta \rho}{\epsilon \rho} , \qquad (44.25)$$

where ρ is the resistivity of the material, and W, L, and t are the width, length, and thickness of the strain gauge, respectively. The last term on the right-hand side of (44.25) reflects the strain-induced change in resistivity of the sample, whereas the first three terms refer to geometrical changes only.

Usually, the gauge length is oriented in a direction where strain is largest to achieve the highest gauge factor. This is referred to as the longitudinal orientation, where the current flows parallel to the strain. Similarly, the transverse orientation is defined when the gauge length is oriented in a direction perpendicular to the maximum strain (Fig. 44.6). The modulation of resistivity in the different orientations with the strain and correspondingly stress can be summarized

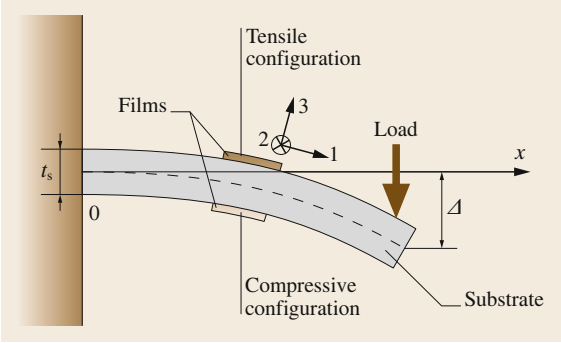


Fig. 44.7 Schematic of a beam-deflection experiment

by the following expression

$$\frac{\Delta \rho_i}{\rho} = \gamma_{ij}\epsilon_j = \pi_{ij}\tau_j$$
$$\text{with } i = 1, 2, 3 \text{ and } j = 1, 2, \ldots, 6 , \qquad (44.26)$$

where γ_{ij} and π_{ij} are the elements of the compact matrix of the elastoresistance and piezoresistance coefficients, respectively, and ϵ_j and τ_j denote the strain and stress components, respectively. (Here, we have used compact notation for these tensors [44.60].)

In metal gauges, the sensitivity of ρ to strain is assumed to be negligible, leading to the well-known longitudinal gauge factor of $k_l = 1 - 2\nu$, where ν is the Poisson's ratio. However, *Arlt* [44.69] has shown that the term $\Delta\rho/\rho$ also includes geometrical attributes. This is due to the change in the volume of the wire and the resulting change in carrier density. According to *Arlt* [44.69], we have for the longitudinal gauge factor

$$k_l = 2 - \frac{\Delta(\gamma\mu)_l}{\epsilon\gamma\mu} , \qquad (44.27)$$

where γ denotes the number of free electrons per atom and μ the electron mobility. Since the last term on the right-hand side of (44.27) is relatively small for metallic gauges, the gauge factor mainly stems from geometrical changes; it is close to 2 and temperature independent [44.70].

These results can be validated using beam-deflection (cantilever) experiments performed on strain gauges integrated on glass and silicon substrates. Figure 44.7 illustrates the schematic of a beam-deflection system. One end of the sample is clamped ($x = 0$) and the free end is deflected by a displacement Δ. *Rajanna* and *Mohan* [44.71] used this method in measurements of both tensile and compressive configurations by simply placing the sample with the films on the top or bottom, respectively (see Fig. 44.7). The strain at lo-

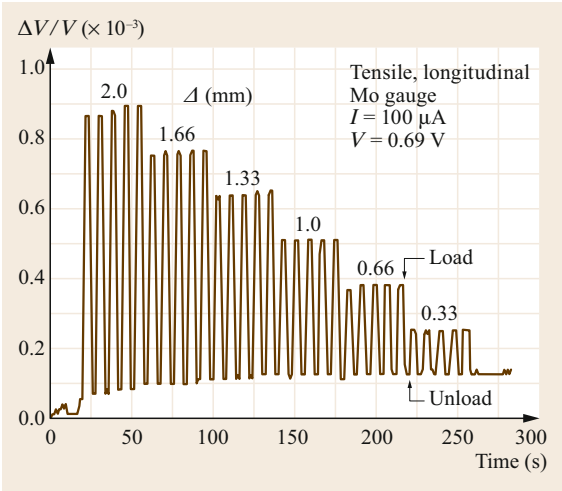


Fig. 44.8 Normalized change in output voltage of a longitudinal Mo strain gauge

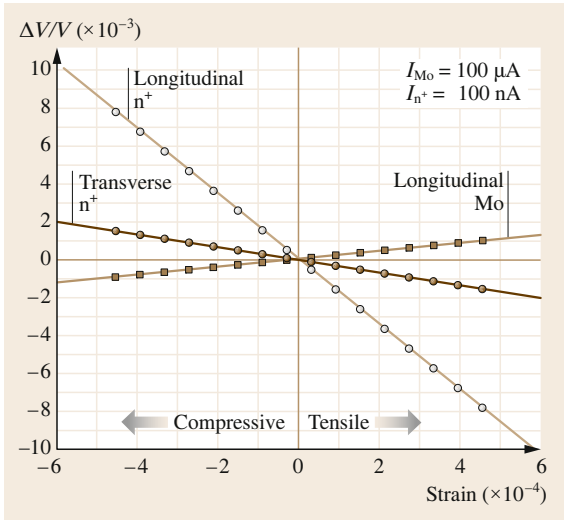


Fig. 44.9 Change in voltage of longitudinal and transverse n^+ μc-Si:H and longitudinal Mo gauges under tensile and compressive strains

cation x along a sample with a length L_0 reads

$$\epsilon(x) = \pm\frac{3t_s\Delta}{2L_0^2}\frac{1-x}{L_0}\,, \tag{44.28}$$

with the positive and negative signs denoting tensile and compressive configurations, respectively.

Figure 44.8 illustrates the result of the beam-deflection experiment on a 150 nm-thick molybdenum strain gauge with 30 turns, line width of 20 μm, and length of 1100 μm. The strain gauges are biased with a constant current of 100 μA and the voltage drop across its terminals is measured. The gauges are subjected to a sequence of loading/unloading steps in tensile configuration to eliminate systematic errors associated with slowly varying transients. By averaging the value of the voltage modulation, we retrieve a gauge factor of approximately $k_l \approx 2.06$ for the longitudinal strain gauges.

The gauge factor for semiconductors is much higher due to the higher elastoresistance coefficients. *Dössel* [44.70] has related the gauge factor to the elastoresistance coefficients as follows

$$k_l = \gamma_1 - (\nu_s + \nu_h)\gamma_2 \quad \text{and}$$
$$k_t = -\nu_s\gamma_1 + (1-\nu_h)\gamma_2\,,$$

where

$$\nu_h = \nu_f\frac{1-\nu_s}{1-\nu_f}\,, \tag{44.29}$$

and $\gamma_1 = 2 - \Delta(\gamma\mu)_l/\gamma\mu\epsilon$ and $\gamma_2 = \Delta(\gamma\mu)_t/\gamma\mu\epsilon$ are the elastoresistance coefficients, and ν_s and ν_f denote the Poisson's ratios for the substrate and the film, respectively.

The same beam-deflection experiment can also be performed on n^+ μc-Si:H strain gauges. Figure 44.9 displays the results of tensile and compressive tests on metal and semiconductor gauges, indicating a higher gauge factor. The values are $k_l = -17.0$ and $k_t = -3.41$ for longitudinal and transverse semiconductor gauges, respectively. Assuming $\nu_s = \nu_f = 0.23$, we find that $\gamma_1 = -22$ and $\gamma_2 = -10.9$ for n^+ μc-Si:H films.

The longitudinal gauge factor obtained for the n^+ μc-Si:H gauges corroborate the results of *Germer* [44.72] for phosphorus-doped microcrystalline thin-film samples with a doping density of $\approx 10^{20}\,\mathrm{cm}^{-3}$. For the transverse gauge factor, Germer has observed a small (negative or positive) gauge factor for doping densities in the range 5×10^{19}–10^{20}. This shows that the transverse gauge factor is highly sensitive to the doping density and other process conditions, which explains the slight difference between our values and that reported by Germer.

44.3.2 Strained Amorphous-Silicon Transistors

Figure 44.10 illustrates measurement results for the transient drain current of a longitudinal TFT subject to a sequence of tensile loading/unloading steps. We see immediate changes in current superimposed on the intrinsic transient response of the TFT. The measured change in current ΔI_D decreases with decreasing dis-

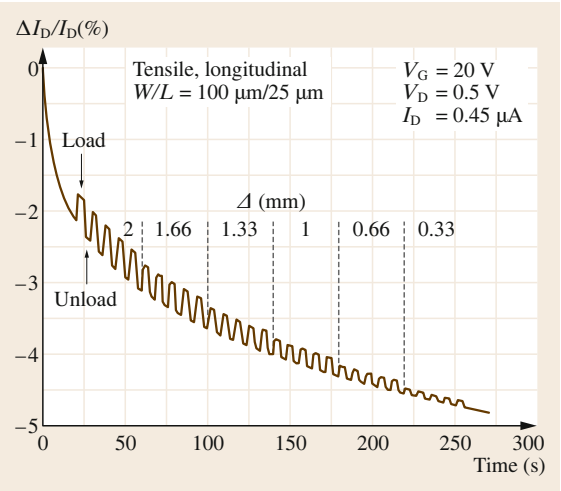


Fig. 44.10 Change in TFT current in deflection experiments

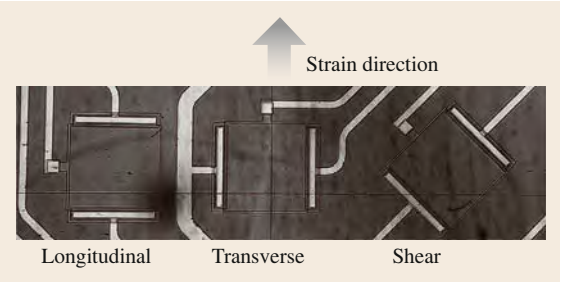


Fig. 44.11 Normalized change in TFT current as a function of tensile and compressive strain

placement. Correspondingly, we define the sensitivity

$$S_\epsilon(I_D) = \frac{1}{I_D}\left.\frac{\partial \Delta I_D}{\partial \epsilon}\right|_{\epsilon=\epsilon_0}, \qquad (44.30)$$

where I_D is the unstrained value of current and ϵ_0 a reference strain value.

Figure 44.11 depicts measurement results for the change in drain current of the longitudinal, shear, and transverse TFTs (Fig. 44.12) under tensile and compressive strains. The TFTs have $W/L = 400\,\mu m/400\,\mu m$ and are biased in the linear regime with constant $V_{GS} = 20\,V$ and $V_{DS} = 0.5\,V$. As seen in the figure, the results for compressive and tensile strain are similar but opposite in sign. For the longitudinal TFT, the current increases with tensile strain with a sensitivity $S_\epsilon(I_{D,l}) = 12.5$. In contrast, for the transverse TFT, this is small and negative $S_\epsilon(I_{D,t}) = -1.1$, clearly signifying an orientation dependence. The value of $S_\epsilon(I_{D,s}) = 4.5$ for the shear TFT can be explained from the linear superposition of

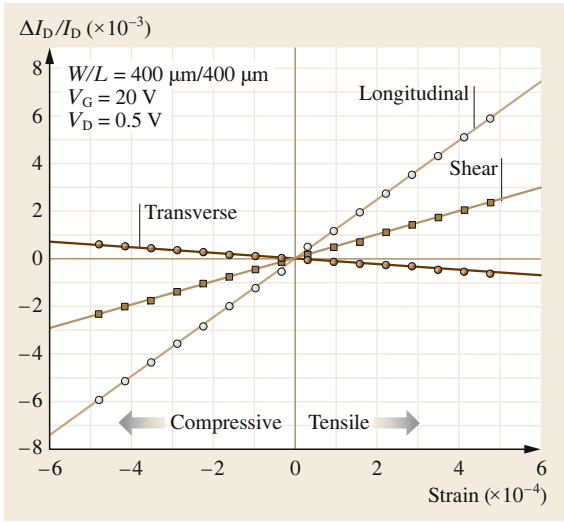


Fig. 44.12 Longitudinal, transverse, and shear TFTs

the effects of longitudinal and transverse strain components.

The measured value for $S_\epsilon(I_{D,l})$ is close to the value of 15 ± 3 reported by *Spear* and *Heintze* [44.64] for intrinsic a-Si:H at room temperature. *Gleskova* et al. [44.73] have found a higher longitudinal sensitivity $S_\epsilon(I_{D,l}) = 26$. For the transverse sensitivity, Spear and Heintze have reported a positive value of $S_\epsilon(I_{D,t}) = 7$ for intrinsic a-Si:H samples.

In addition to TFT orientation, we observe that the gate bias alters the magnitude of strain-induced change in current ΔI_D. To examine the impact of bias, deflection experiments were performed for different values of the gate bias V_{GS} in the range 4–20 V in 1 V steps. Figure 44.13 illustrates the measured $S_\epsilon(I_D)$ as a function of V_{GS} for TFTs of different orientations. Solid symbols denote measurement data for the linear regime ($V_{DS} = 0.5\,V$) while the open symbols are those for the saturation regime where the gate and drain terminals are shorted. Interestingly, the modulation in the current shifts toward positive values as the gate bias decreases. This is true for TFTs of all orientations, and independent of whether the devices were integrated on glass or silicon substrates.

However, the sensitivity $S_\epsilon(I_{D,t})$ for the transverse TFT undergoes a sign change. At high biases, the $S_\epsilon(I_{D,t})$ is generally small and negative (i. e., ΔI_D is positive for tensile strain). As the gate bias decreases, the $S_\epsilon(I_{D,t})$ virtually vanishes at approximately 7 V, and subsequently increases to a sizable positive value (i. e., ΔI_D is negative for tensile strain) at lower voltages ($V_{GS} < 7\,V$).

From the results of Fig. 44.13, we identify two distinct components underlying the strain-induced modu-

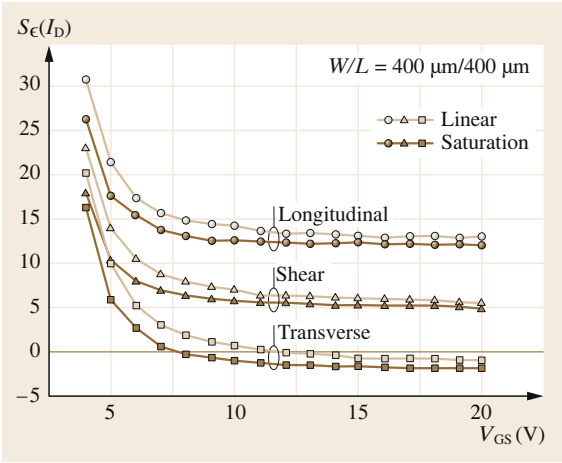


Fig. 44.13 Bias dependence of the TFT current sensitivity

lation of current: the high-bias ($V_{GS} > 7$ V) $S_\epsilon^H(I_D)$ and low-bias ($V_{GS} < 7$ V) $S_\epsilon^L(I_D)$ components such that

$$S_\epsilon(I_D) = S_\epsilon^H(I_D) + S_\epsilon^L(I_D) \,. \tag{44.31}$$

At high biases, $S_\epsilon(I_D)$ of longitudinal, transverse, and shear TFTs gradually approach constant values. As seen from Table 44.2, the extracted values for $S_\epsilon^H(I_D)$ are strongly orientation dependent, suggesting the presence of strain-induced modulation in carrier mobility, whose sensitivity we define as $S_\epsilon(\mu_{eff}) = \Delta\mu_{eff}/\mu_{eff}\epsilon$. The mobility change in the longitudinal orientation is higher than that in the transverse orientation.

Superimposed on the high-bias component is the low-bias component $S_\epsilon^L(I_D)$ which manifests itself as a bias-dependent positive shift in $S_\epsilon(I_D)$. This component can be attributed to the modulation in threshold voltage. Correspondingly, we define the threshold-voltage sensitivity as $S_\epsilon(V_{thr}) = \Delta V_{thr}/V_{thr}\epsilon$. The change in V_{thr} can be attributed to a strain-induced change in the density of deep states, which is orientation independent [44.64]. The modulation in threshold voltage leads to a significant change in current at low biases, and can be quantitatively explained by looking at the current–voltage relation in the linear regime. From (44.16) and using partial differentiation with respect to strain, we can write $S_\epsilon(I_{DS,lin})$ as

$$S_\epsilon(I_{DS,lin}) = S_G + S_\epsilon(\mu_{eff}) - S_\epsilon(V_{thr})\frac{\alpha V_{thr}}{V_{GT}} + S_\epsilon(\alpha)\left(\alpha \ln\left|\frac{V_{GT}}{V_0}\right| - \frac{\alpha}{2} - \frac{1}{\alpha - 1}\right), \tag{44.32}$$

where $V_0 = (2\epsilon k_B T_t N_0)^{1/2}/C_i$, and

$$S_G = S_\epsilon(W) - S_\epsilon(L) + (\alpha - 1)S_\epsilon(C_i) \,.$$

Table 44.2 Values for $S_\epsilon^H(I_D)$, $S_\epsilon(\mu_{eff})$, and $S_\epsilon(V_{thr})$ for different orientations

Parameter	Longitudinal	Transverse	Shear
$S_\epsilon^H(I_D) = \Delta I_D/I_D\epsilon$	12.1	−1.1	4.5
$S_\epsilon(\mu_{eff}) = \Delta\mu_{eff}/\mu_{eff}\epsilon$	11	−1.1	4.0
$S_\epsilon(V_{thr}) = \Delta V_{thr}/V_{thr}\epsilon$	5	4.5	4.7

Here, S_G includes the effect of change in device dimensions. The other terms on the right-hand side of (44.32), in order from the left, describe the dependence of $S_\epsilon(I_{DS,lin})$ on $S_\epsilon(\mu_{eff})$, $S_\epsilon(V_{thr})$, and $S_\epsilon(\alpha)$, respectively, which represent strain-induced modulation of the different TFT parameters (μ_{eff}, V_{thr}, and α).

As seen in (44.32), the modulation in mobility $S_\epsilon(\mu_{eff})$ is directly reflected in the change of current. In contrast, the strain-induced change in threshold voltage $\Delta V_{thr}/V_{thr}\epsilon$ is scaled, and by a factor of $\alpha V_{thr}/V_{GT}$. This is particularly visible at low biases and its effect decreases with increasing gate bias V_{GT}. Thus, at high biases, the impact of threshold-voltage modulation is minimal, which yields

$$S_\epsilon^H(I_D) = S_G + S_\epsilon(\mu_{eff}) \,. \tag{44.33}$$

In contrast, the low-bias component can be written as

$$S_\epsilon^L(I_D) = -S_\epsilon(V_{thr})\frac{\alpha V_{thr}}{V_{GT}} \,. \tag{44.34}$$

It is important to note that the last term on the right-hand side of (44.32) represents the impact of strain-induced modulation of α. This term contains a bias-dependent scaling factor of $\alpha \ln|V_{GT}/V_0|$, which does not correlate with the observed bias dependence of $S_\epsilon(I_D)$ seen in Fig. 44.13. This implies that modulation in V_{thr} and μ_{eff} are the dominant contributors to the observed changes in current.

The activation energy for the temperature dependence of $S_\epsilon^L(I_D)$ and $S_\epsilon^H(I_D)$ is defined as

$$E_i = \frac{\partial S_\epsilon^i(I_D)}{\partial(1/(k_B T))} \quad \text{with } i = \text{H or L} \,. \tag{44.35}$$

The values for E_H and E_L are found to be 140 meV and 0.58 eV, respectively. The much lower activation energy (140 meV) at high biases indicates that the $S_\epsilon^H(I_D)$ stems from the shallow states in the conduction-band tail. Again, this corroborates our previous findings that the sensitivity of the current at high biases is associated with the mobility modulation that is principally determined by the tail states. *Spear* and *Heintze* [44.64] have found an activation energy of 0.52 eV for intrinsic and doped a-Si:H layers. This corroborates with our low-voltage sensitivity data. The high activation energy

for $S_\epsilon^{\mathrm{L}}(I_\mathrm{D})$ identifies the role of deep states in the gap, which is in agreement with our previous finding that $S_\epsilon(I_\mathrm{D}) \propto S_\epsilon(V_\mathrm{thr})$ at low biases. Here, the strain is believed to modify the energy of the deep states [44.64].

From (44.33) and (44.34), the values of $S_\epsilon(\mu_\mathrm{eff})$ and $S_\epsilon(V_\mathrm{thr})$ for different orientations can be determined (Table 44.2). The values can be incorporated in a compact model for the different TFT orientations, which can be used for computer-aided design (CAD) of mechanically flexible TFT circuits [44.74].

Acknowledgments. The authors would like to acknowledge Natural Sciences and Engineering Research Council (NSERC) of Canada for their financial support.

References

44.1 S.R. Forrest: Nature **428**, 911 (2004)
44.2 S. Wagner, H. Gleskova, J. C. Sturm, Z. Suo: In: *Technology and Applications of Amorphous Silicon*, ed. by R.A. Street (Springer, Berlin 2000) p. 222
44.3 S.E. Shaheen, R. Radspinner, N. Peyghambarian, G.E. Jabbour: Appl. Phys. Lett. **79**, 2996 (2001)
44.4 R.A. Street, M. Mulato, R. Lau, J. Ho, J. Graham, Z. Popovic, J. Hor: Appl. Phys. Lett. **78**, 4193 (2001)
44.5 P. Servati, Y. Vygranenko, A. Nathan, S. Morrison, A. Madan: J. Appl. Phys. **96**, 7575 (2004)
44.6 S. Soltanian, A. Servati, R. Rahmanian, F. Ko, P. Servati: J. Mater. Res. **30**, 121 (2015)
44.7 L. Collins: IEE Rev. **49**(3), 42 (2003)
44.8 A.B. Chwang, M.A. Rothman, S.Y. Mao, R.H. Hewitt, M.S. Weaver, J.A. Silvernail, K. Rajan, M. Hack, J.J. Brown, X. Chu, L. Moro, T. Krajewsky, N. Rutherford: Appl. Phys. Lett. **83**(3), 413 (2003)
44.9 A. Nathan, A. Kumar, K. Sakariya, P. Servati, S. Sambandan, D. Striakhilev: IEEE J. Solid-State Circuits **39**(9), 1477 (2004)
44.10 J.K. Jeong: J. Mater. Res. **28**(16), 2071 (2013)
44.11 H.Y. Jung, Y. Kang, A.Y. Hwang, C.K. Lee, S. Han, D.-H. Kim, J.-U. Bae, W.-S. Shin, J.K. Jeong: Sci. Rep. **4**, 3765 (2014)
44.12 A. Nathan, S. Lee, S. Jeon, J. Robertson: J. Disp. Technol. **10**(11), 917 (2014)
44.13 M.J. Powell, C. van Berkel, A.R. Franklin, S.C. Deane, W.I. Milne: Phys. Rev. B **45**(8), 4160 (1992)
44.14 D. Knipp, R.A. Street, A. Volkel, J. Ho: J. Appl. Phys. **93**, 347 (2003)
44.15 S. Alexander, P. Servati, G.R. Chaji, S. Ashtiani, R. Huang, D. Striakhilev, K. Sakariya, A. Kumar, A. Nathan, C. Church, J. Wzorek, P. Arsenault: J. Soc. Info. Disp. **13**(7), 587 (2005)
44.16 G.L. Graff, R.E. Williford, P.E. Burrows: J. Appl. Phys. **96**(4), 1840 (2004)
44.17 J.A. Rogers, Z. Bao, A. Dodabalapur, A. Makhija: IEEE Electron Device Lett. **21**(3), 100 (2000)
44.18 C.-S. Yang, L.L. Smith, C.B. Arthur, G.N. Parsons: J. Vac. Sci. Technol. B **18**(2), 683 (2000)
44.19 H. Gleskova, S. Wagner, V. Gašparík, P. Kováč: Appl. Surf. Sci. **175/176**, 12 (2001)
44.20 D. Stryahilev, A. Sazonov, A. Nathan: J. Vac. Sci. Technol. A **20**(3), 1087 (2002)
44.21 M. Meitine, A. Sazonov: Mater. Res. Soc. Symp. Proc. **769**, H6.6.1 (2003)
44.22 W.A. MacDonald: J. Mater. Chem. **14**, 4 (2004)
44.23 K.L. Chopra: *Thin Film Phenomena* (McGraw–Hill, Toronto 1969)
44.24 E.Y. Ma, S. Wagner: Appl. Phys. Lett. **74**, 2661 (1999)
44.25 A. Madan, P.G. Le Comber, W.E. Spear: J. Non-Cryst. Solids **20**, 239 (1976)
44.26 N.F. Mott, E.A. Davis: *Electronic Processes in Non-Crystalline Materials* (Oxford Univ. Press, Oxford 1971)
44.27 C. Popescu, T. Stoica: In: *Thin Film Resistive Sensors*, ed. by P. Ciureanu, S. Middelhoek (IOPP, New York 1992) p. 37
44.28 P.W. Anderson: Phys. Rev. **109**(5), 1492 (1958)
44.29 M.H. Cohen, H. Fritzsche, S.R. Ovshinsky: Phys. Rev. Lett. **22**(20), 1066 (1969)
44.30 D. Monroe: Phys. Rev. Lett. **54**(2), 146 (1985)
44.31 R.A. Street: *Hydrogenated Amorphous Silicon* (Cambridge Univ. Press, New York 1992)
44.32 P. Servati, A. Nathan, G.A.J. Amaratunga: Phys. Rev. B **74**, 245210 (2006)
44.33 W.E. Spear, P.G. Le Comber: J. Non-Cryst. Solids **8–10**, 727 (1972)
44.34 M.J. Powell: Philos. Mag. B **43**(1), 93 (1981)
44.35 C.-Y. Huang, S. Guha, S.J. Hudgens: Phys. Rev. B **27**(12), 7460 (1983)
44.36 J.D. Cohen, D.V. Lang, J.P. Harbison: Phys. Rev. Lett. **45**(3), 197 (1980)
44.37 M. Hirose, T. Suzuki, G.H. Döhler: Appl. Phys. Lett. **34**(3), 234 (1979)
44.38 P. Viktorovitch, G. Moddel: J. Appl. Phys. **51**(9), 4847 (1980)
44.39 M. Shur, M. Hack: J. Appl. Phys. **55**(10), 3831 (1984)
44.40 J.G. Shaw, M. Hack: J. Appl. Phys. **64**(9), 4562 (1988)
44.41 C. Tanase, E.J. Meijer, P.W.M. Blom, D.M. de Leeuw: Phys. Rev. Lett. **91**(21), 216 (2003)
44.42 T. Tiedje, A. Rose: Solid State Commun. **37**, 49 (1980)
44.43 M.J.C.M. Vissenberg, M. Matters: Phys. Rev. B **57**(20), 12964 (1998)
44.44 A. Salleo, T.W. Chen, A.R. Völkel, Y. Wu, P. Liu, B.S. Ong, R.A. Street: Phys. Rev. B **70**, 115 (2004)
44.45 B.S. Ong, Y. Wu, P. Liu, S. Gardner: J. Am. Chem. Soc. **126**, 3378 (2004)
44.46 A.J. Campbell, M.S. Weaver, D.G. Lidzey, D.D.C. Bradley: J. Appl. Phys. **84**(12), 6737 (1998)
44.47 D. Natali, M. Sampietro: J. Appl. Phys. **92**(9), 5310 (2002)
44.48 P. Servati, D. Striakhilev, A. Nathan: IEEE Trans. Electron Devices **50**(11), 2227 (2003)
44.49 C. Tanase, E.J. Meijer, P.W.M. Blom, D.M. de Leeuw: Org. Electron. **4**, 33 (2003)
44.50 F.R. Shapiro, D. Adler: J. Non-Cryst. Solids **74**(2/3), 189 (1985)

44.51 S.D. Baranovskii, T. Faber, F. Hensel, P. Thomas: J. Phys.: Condens. Matter **9**, 2699 (1997)
44.52 S. Lee, A. Nathan, Y. Ye, Y. Guo, J. Robertson: Sci. Rep. **5**, 13467 (2015)
44.53 S. Lee, A. Nathan: Appl. Phys. Lett. **101**, 11 (2012)
44.54 K. Ghaffarzadeh, A. Nathan, J. Robertson, S. Kim, S. Jeon, C. Kim, U.-I. Chung, J.-H. Lee: Appl. Phys. Lett. **97**, 143510 (2010)
44.55 S. Jeon, S.-E. Ahn, I. Song, C.J. Kim, U.-I. Chung, E. Lee, I. Yoo, A. Nathan, S. Lee, J. Robertson, K. Kim: Nat. Mater. **11**, 301 (2012)
44.56 S. Lee, S. Jeon, R. Chaji, A. Nathan: Proc. IEEE **103**, 644 (2015)
44.57 S. Lee, K. Ghaffarzadeh, A. Nathan, J. Robertson, S. Jeon, C. Kim, I.-H. Song, U.-I. Chung: Appl. Phys. Lett. **98**, 20 (2011)
44.58 S. Lee, S. Jeon, A. Nathan: J. Disp. Technol. **9**, 883 (2013)
44.59 S. Lee, A. Nathan, J. Robertson, K. Ghaffarzadeh, M. Pepper, S. Jeon, C. Kim, I.-H. Song, U.-I. Chung, K. Kim: Temperature dependent electron transport in amorphous oxide semiconductor thin film transistors, Technical Digest, IEEE Int. Electron Device Meet. (IEDM) 2011 (2011), 14.6.1
44.60 A. Nathan, H. Baltes: *Microtransducer CAD* (Springer, Wien 1999)
44.61 C. Herring, E. Vogt: Phys. Rev. **101**(3), 944 (1956)
44.62 V.A. Gridchin, V.M. Lubimsky, M.P. Sarina: Sens. Actuators A **49**(1/2), 67 (1995)
44.63 A. Dévényi, A. Belu, G. Korony: J. Non-Cryst. Solids **4**, 380 (1970)
44.64 W.E. Spear, M. Heintze: Philos. Mag. B **54**(5), 343 (1986)
44.65 B. Welber, M.H. Brodsky: Phys. Rev. B **16**(8), 3660 (1977)
44.66 B.A. Weinstein: Phys. Rev. B **23**(2), 787 (1981)
44.67 R. Zallen, W. Paul: Phys. Rev. **155**(3), 703 (1967)
44.68 D. Lazarus: Phys. Rev. B **24**(4), 2282 (1981)
44.69 G. Arlt: J. Appl. Phys. **49**(7), 4273 (1978)
44.70 O. Dössel: Sens. Actuators **6**(3), 169 (1984)
44.71 K. Rajanna, S. Mohan: Phys. Status Solidi (a) **105**(2), K181 (1988)
44.72 W. Germer: Sens. Actuators **7**(2), 135 (1985)
44.73 H. Gleskova, S. Wagner, W. Soboyejo, Z. Suo: J. Appl. Phys. **92**(10), 6224 (2002)
44.74 P. Servati, A. Nathan: Appl. Phys. Lett. **86**(7), 033504 (2005)

45. Photoconductors for X-Ray Image Detectors

M. Zahangir Kabir, Safa Kasap

Modern flat-panel x-ray imaging detectors have played an important role in the transition from analog to digital x-ray imaging. They capture an x-ray image electronically and hence enable a clinical transition to digital radiography. This chapter critically discusses the material, transport and imaging detector properties (e.g., dark current) of several potential x-ray photoconductors and compares them with an ideal photoconductor for use in direct-conversion imaging detectors. The present chapter also considers various metrics of detector performances including sensitivity, detective quantum efficiency, resolution in terms of the modulation transfer function, image lag and ghosting; and examines how these metrics depend on the photoconductor material, and detector structure and design.

45.1 **X-Ray Photoconductors** 1127
45.1.1 Ideal Photoconductor Properties 1127
45.1.2 Potential Photoconductors 1128
45.1.3 Summary and the Future 1135
45.2 **Dark Current Limitations** 1136
45.2.1 Charge Carrier Depletion 1137
45.2.2 Steady-State Thermal Generation 1137
45.2.3 Carrier Injection 1138
45.3 **Metrics of Detector Performance** 1139
45.3.1 X-Ray Sensitivity 1140
45.3.2 Detective Quantum Efficiency 1141
45.3.3 Modulation Transfer Function (MTF) 1143
45.3.4 Image Lag and Ghosting 1144
45.4 **Summary** 1145
References 1145

The flat-panel x-ray image detector (sensor) concept is illustrated in Fig. 45.1 where x-rays passing through the patient's hand are incident on the sensor. The sensor converts the incident image to a digital image that can be sent to a computer and viewed on a monitor. Research over the past two decades [45.1] has indicated that the most practical flat-panel medical digital radiographic systems (mammography, chest radiography and fluoroscopy) are those based on a large-area integrated circuit or active matrix array (AMA). Flat-panel imagers incorporating active matrix arrays are called active matrix flat-panel imagers or AMFPIs. Active matrix arrays using hydrogenated amorphous silicon (a-Si:H) thin-film transistors (TFTs) have been shown to be practical pixel-addressing systems for displays [45.2]. These a-Si:H arrays can be converted into x-ray-sensitive imaging devices by adding a thick (0.1−1 mm) x-ray-detecting medium, either a phosphor or a photoconductor. The physical form of the x-ray AMFPI is similar to a film/screen cassette and thus it will easily fit into current medical film/screen-based x-ray imaging systems. The x-ray image is stored and displayed on the computer almost immediately after the x-ray exposure. The stored image can be rapidly transmitted to remote locations for consultation and analysis. The dynamic range of recently developed AMFPI systems is much higher than film/screen-based imaging systems [45.3]. AMFPIs are able to read out an entire image in $1/30$ s, sufficiently rapid to perform fluoroscopy (real-time imaging) [45.4].

The active matrix array used for pixel addressing and readout in an AMFPI consists of many single pixels, each of which represents a corresponding pixel of the image. Each pixel collects charge proportional to the x-ray radiation that it receives. To generate this signal charge, either a phosphor converts the x-rays to visible light, which in turn is detected with a pin photodiode at the pixel (indirect) or an x-ray photoconductor converts the incident x-rays to charge (direct) in one step. For both the indirect and direct conversion approaches, the latent image is a charge distribution residing on the array's pixels. The charges are read out by scanning the arrays row by row using peripheral electronics and multiplexing the parallel columns into a serial digital signal. This signal is then transmitted to a computer system for storage and display. Several manufacturers and academic researchers have used the indirect approach [45.5, 6]. However, we believe that

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_45

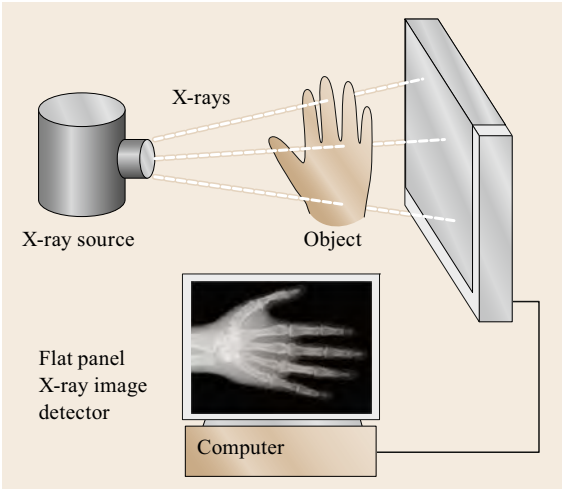


Fig. 45.1 Schematic illustration of an active matrix flat-panel imaging (AMFPI) system used for x-ray imaging

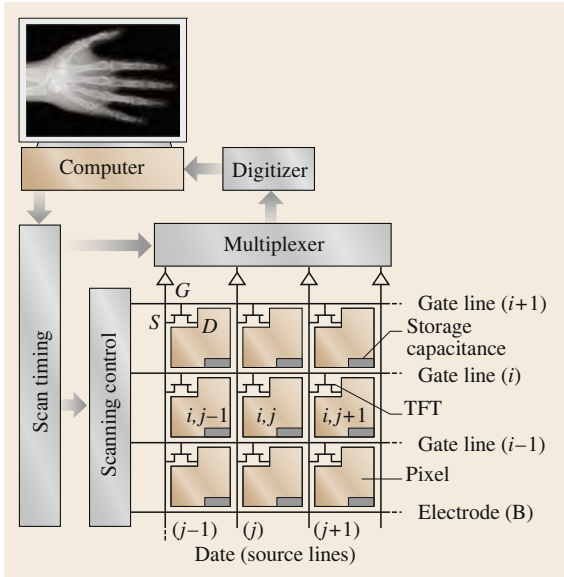


Fig. 45.2 Schematic diagram that shows some pixels of the active matrix array for use in x-ray AMFPI with self-scanned electronic readout. The charge distribution residing on the panel's pixels are read out by scanning the arrays row by row using peripheral electronics and multiplexing the parallel columns into a serial digital signal

the direct approach, due to its higher resolution, has the potential to produce systems superior in image quality to indirect conversion sensors and be both easier and cheaper to manufacture due to their simpler structure. This chapter considers only the direct-conversion x-ray imager and how its dark current, sensitivity, resolution and detective quantum efficiency (DQE) depend on the photoconductor material and detector structure. This chapter also discusses essential photoconductor properties, charge transport and imaging properties of promising photoconductors.

Active matrix arrays allow a monolithic imaging system of large area (e.g., 40×40 cm) to be constructed. As for conventional integrated circuits, planar processing of the array through deposition and doping of lithographically masked individual layers of metals, insulators and semiconductors implement the design of active matrix arrays. In the future, even larger areas should become feasible if required. Millions of individual pixel electrodes in the matrix are connected, as shown in Fig. 45.2. Each pixel has its own thin-film transistor (TFT) switch and storage capacitor to store image charges. The TFT switches control the image charge so that one line of pixels is activated electronically at a time. Normally, all the TFTs are turned off, permitting the latent image charge to accumulate on the array. The readout is achieved by external electronics and software controlling the state of the TFT switches. The active matrix array consists of $M \times N$ (e.g., 2480×3072) storage capacitors C_{ij} whose stored image charge can be read through properly addressing the TFT (i,j) via the gate (i) and source (j) lines. The charges read on each C_{ij} are converted to a digital image. The readout is self-scanning in that no external means such as a laser are used. The scanning is part of the AMFPI electronics and software, thus permitting a truly compact device.

To construct a direct-conversion x-ray AMFPI, a large-bandgap (> 2 eV) high-atomic-number semiconductor or x-ray photoconductor (e.g., stabilized amorphous selenium, a-Se) layer is coated onto the active matrix array. An electrode (A) is subsequently deposited onto the photoconductor layer to facilitate the application of a biasing potential and, hence, an electric field F in the photoconductor layer as shown in Fig. 45.3. The biasing potential applied to the radiation-receiving electrode A (top electrode) may be positive or negative, the selection of which depends on many factors discussed later in this chapter. The applied bias varies from a few hundred to several thousand V. The capacitance C_{pc} of the photoconductor layer over the pixel is in series with, and much smaller than, the pixel capacitance C_{ij} attached to the pixel electrode, so that most of the applied potential is dropped across the photoconductor and not across the pixel capacitance. The electron–hole pairs (EHPs) generated in the photoconductor by the absorption of x-ray photons travel along the field lines and are collected by the electrodes. If the applied bias voltage is positive, then electrons collect at the positive bias electrode and holes accumulate on the storage capacitor C_{ij}. Each pixel electrode car-

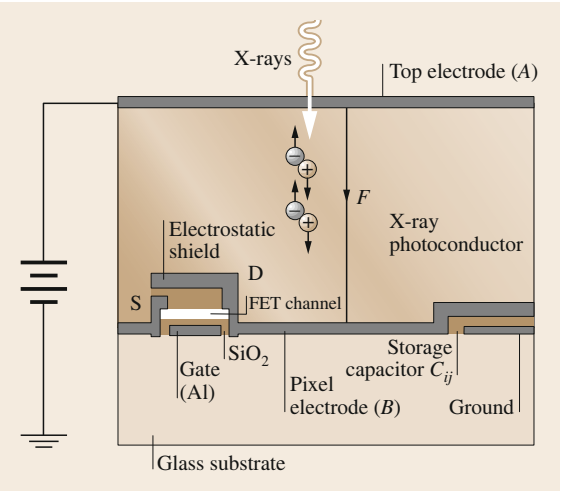


Fig. 45.3 Cross section of a single pixel (i,j) with a TFT switch. The top electrode (A) on the photoconductor is a vacuum-coated metal (e.g., Al). The bottom electrode (B) is the pixel electrode that is one of the plates of the storage capacitance (C_{ij}). (Not to scale – the TFT height is highly exaggerated)

ries an amount of charge Q_{ij} proportional to the amount of incident x-ray radiation in the photoconductor layer over that pixel, which can be read during self-scanning.

For a historical perspective on the development of a-Se based flat panel detectors, the reader is referred to a number of early papers by *Rowlands* and coworkers [45.7–13] and *Lee* et al. [45.14, 15].

The selection of the photoconductor material for use in direct-conversion x-ray image detectors is currently an important research field in electronic materials. There are various competing semiconductors, such as amorphous (a-)Se, HgI_2, CdZnTe, PbI_2, PbO and organic perovskites. Detectors based on a-Se have already been developed and successfully commercialized for mammography [45.16]; the majority of modern direct conversion mammographic detectors now use an a-Se alloy as a photoconductor. This chapter discusses charge-transport and imaging detector properties (e.g., dark current and image lag) of these photoconductors and compares them with an ideal photoconductor for x-ray imaging detectors. The present chapter also examines various imaging characteristics of photoconductor-based x-ray AMFPIs, including dark current, sensitivity (S), detective quantum efficiency (DQE), resolution in terms of the modulation transfer function (MTF), image lag and ghosting. We examine how these characteristics depend not only on the photoconductor's charge-transport properties but also on the detector structure, i. e., the size of the pixel and the thickness of the photoconductor.

45.1 X-Ray Photoconductors

45.1.1 Ideal Photoconductor Properties

The performance of direct-conversion x-ray detectors depends critically on the selection and design of the photoconductor. It is therefore instructive to identify what constitutes a nearly ideal x-ray photoconductor to guide a search for improved performance or a better material. Ideally, the photoconductive layer should possess the following material properties:

1. Most of the incident x-ray radiation should be absorbed within a practical photoconductor thickness to avoid unnecessary patient exposure. This means that, over the energy range of interest, the absorption depth δ of the x-rays must be substantially less than the photoconductor layer thickness L. In other words, the quantum efficiency (η) should be high.
2. The photoconductor should have high intrinsic x-ray sensitivity, i. e., it must be able to generate as many collectable (free) electron–hole pairs (EHPs) as possible per unit of incident radiation. This means the amount of radiation energy required, denoted by $W_{\pm}$, to create a single free electron–hole pair must be as low as possible. Typically, $W_{\pm}$ increases with the bandgap E_g of the photoconductor [45.17] and thus a low E_g is desired for maximum x-ray sensitivity.
3. There should be little bulk recombination of electrons and holes as they drift to the collection electrodes; EHPs are generated in the bulk of the photoconductor. Bulk recombination is proportional to the product of the concentration of holes and electrons, and typically it is negligible for clinical exposure rates (i. e., provided the instantaneous x-ray exposure is not too high) [45.18].
4. There should be negligible deep trapping of electrons and holes, which means that, for both electrons and holes, the schubweg $\mu\tau F$ must be much greater than L, where μ is the drift mobility, τ is the deep-trapping time (lifetime), F is the electric field and L is the photoconductor layer thickness. The schubweg is the mean distance a carrier drifts before it is trapped and unavailable for conduction. The temporal responses of the x-ray image detector, such as lag and ghosting, depend on the rate of carrier trapping.

5. The diffusion of carriers should be negligible compared with their drift. This property ensures less time for lateral carrier diffusion and leads to better spatial resolution.
6. The dark current should be as small as possible, because it is a source of noise. The applied field causes a current to flow through the detector in the absence of irradiation, which is called the dark current. This means the contacts to the photoconductor should be noninjecting and the rate of thermal generation of carriers from various defects or states in the bandgap should be negligibly small (i. e., dark conductivity is practically zero). Small dark conductivity generally requires a wide-bandgap semiconductor, which conflicts with condition (2) above. The dark current should preferably not exceed $10-1000\,\mathrm{pA/cm^2}$, depending on the clinical application [45.19].
7. The longest carrier-transit time, which depends on the smallest drift mobility, must be shorter than the image readout time and interframe time in fluoroscopy.
8. The properties of the photoconductor should not change with time because of repeated exposure to x-rays, i. e., x-ray fatigue and x-ray damage should be negligible.
9. The photoconductor should be easily coated onto the AMA panel (typically $30 \times 30\,\mathrm{cm}$ and larger), for example by conventional vacuum techniques without raising the temperature of the AMA to damaging levels (e.g., $\approx 300\,^{\circ}\mathrm{C}$ for a-Si panels). This eliminates the possibility of using single-crystal materials that would require extended exposure to much higher temperature if they were to be grown directly onto the panel.
10. The photoconductor should have uniform characteristics over its entire area.
11. The temporal artifacts such as image lag and ghosting should be as small as possible. Lag is the carryover of image charge generated by previous x-ray exposures into subsequent image frames. The residual signal fractions following a pulsed x-ray irradiation are referred to as *image lag*. Ghosting is the change of x-ray sensitivity of the x-ray image detector as a result of previous exposure to radiation. In the presence of ghosting, a shadow impression of a previously acquired image is visible in subsequent uniform exposures.

The large-area-coating requirement in (9) rules out the use of crystalline semiconductors, whose only practical production process is to grow large boules, which are subsequently sliced. Thus, only amorphous or polycrystalline (poly)photoconductors are currently practical for use in large-area x-ray imaging detectors. Amorphous selenium (a-Se) is one of the most highly developed photoconductors for large-area detectors due to its commercial use in photocopiers and laser printers as an electrophotographic photoreceptor [45.20]. In fact, the direct-conversion flat-panel imaging technology has been made possible by the use of two key elemental amorphous semiconductors: a-Si:H (used for TFTs) and a-Se (used for photoconductor layers). Although their properties are different, both can be readily prepared in large areas, which is essential for an x-ray imaging detector. Stabilized a-Se (a-Se alloyed with 0.2−0.5% As and doped with 5−20 ppm Cl) is currently the preferred photoconductor for clinical x-ray image sensors, because it can be quickly and easily deposited as a uniform film over large areas and it has an acceptable x-ray absorption coefficient, good charge-transport properties for both holes and electrons, and lower dark current than many competing polycrystalline layers [45.17, 19]. Flat-panel x-ray image detectors with an a-Se photoconductor have been shown to provide excellent images.

There has been active research to find potential x-ray photoconductors to replace a-Se in flat-panel image detectors because of the substantially higher $W_{\pm}$ and operating electric field of a-Se compared to other potential x-ray photoconductors [45.19, 21]. For example, the typical value of the electric field used in a-Se devices is $10\,\mathrm{V/\mu m}$ where the value of $W_{\pm}$ is about 45 eV; the value of $W_{\pm}$ is 5−6 eV for polycrystalline mercuric iodide (poly-HgI_2) and poly-CdZnTe. The main drawback of polycrystalline materials is the adverse effects of grain boundaries in limiting charge transport and the nonuniform response of the sensor due to large grain sizes. Grain boundaries in the polycrystalline material are expected to create trapping levels within the bandgap and introduce potential barriers between neighboring grains [45.22, 23]. Another disadvantage of these polycrystalline detectors is the higher dark current compared to a-Se detectors. However, there have been efforts to improve the material properties and reduce the dark currents of poly-HgI_2 and poly-CdZnTe-based image detectors [45.24, 25]. Indeed, experiments on large-area HgI_2, PbI_2, CdZnTe (< 10% Zn), and PbO polycrystalline x-ray photoconductive layers deposited on active matrix arrays have shown encouraging results [45.26–28]. A more detailed description of these three potential photoconductors for direct-conversion AMFPIs is presented below.

45.1.2 Potential Photoconductors

The properties of an ideal photoconductor for x-ray image detectors have been discussed in the section above.

In this section, important properties of potential photoconductors for x-ray image detectors are discussed and compared with the ideal case.

Amorphous Selenium (a-Se)

Stabilized a-Se can be easily coated as thick films (e.g., 100−1000 μm) onto suitable substrates by conventional vacuum deposition techniques and without the need to raise the substrate temperature beyond 60–70 °C (much below the damage threshold of the AMA, e.g., ≈ 300 °C for a-Si:H panels). Its amorphous state maintains uniform characteristics to a very fine scale over large areas. Stabilized a-Se, not pure a-Se, is used in x-ray sensors, because pure a-Se is thermally unstable and crystallizes over several weeks or months following manufacture. Crystalline Se is unsuitable as an x-ray photoconductor because it has a much lower dark resistivity and hence orders of magnitude larger dark current than a-Se. Alloying pure a-Se with As (0.2−0.5% As) greatly improves the stability of the composite film and helps to prevent crystallization. However, it has been found that the addition of As has an adverse effect on the hole lifetime because the arsenic introduces deep hole traps. If the alloy is doped with 5−20 parts per million (ppm) of a halogen (such as Cl), the hole lifetime is restored to its initial value. Thus, an a-Se film that has been alloyed with 0.2−0.5% As (nominal 0.3% As) and doped with ppm-levels of Cl is called stabilized a-Se. The density of a-Se is 4.3 g/cm^3 with a bandgap (usually called the mobility gap in amorphous semiconductors) of $E_g \approx 2.1-2.2$ eV.

There are localized states within the mobility gap of a-Se. Some of these are located near the band edges (*shallow traps*), while some are located deep in the energy band (*deep traps*). Localized states are simply traps and are not extended throughout the material, but are localized in space (both shallow and deep) due to various structural defects that are stable at room temperature. Drift of both electrons and holes involves interactions with shallow and deep traps, as shown in Fig. 45.4. Shallow traps reduce the drift mobility; deep traps prevent the carriers from crossing the photoconductor. The effective drift mobility μ of carriers is the mobility μ_0 in the extended states reduced by the trapping and release events due to the presence of shallow traps,

$$\mu = \frac{\tau_c}{\tau_c + \tau_r}\mu_0 \,, \tag{45.1}$$

where τ_c and τ_r are the average capture and release times in the shallow-trap centers. The capture time represents the mean time that a mobile carrier drifts in the extended states before becoming trapped in a shallow-trap center. The release time is the mean time that

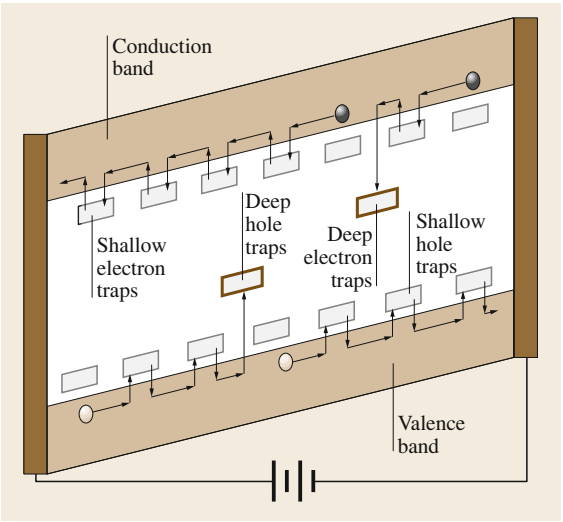


Fig. 45.4 Diagram illustrating the bandgap of a photoconductor with an applied electric field that tilts the bands, encouraging drift of holes in the direction of the field and electrons counter to the field. Drift of both electrons and holes involves interactions with *shallow* and *deep traps*. *Shallow traps* reduce the drift mobility and *deep traps* prevent the carriers from crossing the photoconductor

a carrier remains in a trap before being released back into the extended states. Re-emission from a shallow trap is mostly dominated by thermally activated processes. The shallow-trap release time is relatively short and a typical carrier may experience many shallow capture and release events while traversing the detector thickness (100−1000 μm).

Although the nature of the shallow traps in a-Se has not been fully established, the drift mobilities of both holes and electrons are quite reproducible. The room-temperature effective hole mobility μ_h up to moderate fields ($F \leq 10\,\mathrm{V/\mu m}$) is independent of the preparation of the sample and has a value of $\approx 0.12\,\mathrm{cm^2/Vs}$, whereas the effective electron mobility μ_e is in the range $0.003-0.006\,\mathrm{cm^2/Vs}$ [45.21]. The hole drift mobility does not change with the addition of As or Cl. The value of μ_e decreases with the addition of As to a-Se (e.g., in stabilized a-Se) but Cl doping does not affect it.

Once a carrier is caught in a deep trap, it will remain immobile until a lattice vibration gives it enough energy to be excited back into the extended states, where it can drift once again. The deep-trap release time is very long (minutes to hours), and a deeply trapped carrier is essentially permanently removed from conduction. Therefore, the carrier lifetime depends on the concentration of deep rather than shallow traps. The charge-carrier lifetimes vary substantially between

different samples and depend on factors such as the source of a-Se material, impurities, and the preparation method. The electron lifetime τ_e is particularly sensitive to impurities in the a-Se source material. The hole lifetime τ_h drops rapidly with decreasing substrate temperature (temperature of the a-Se substrate during the evaporation process) whereas τ_e is unaffected. Increasing the As concentration in a-Se decreases τ_h and increases τ_e [45.29]. On the other hand, Cl doping increases τ_h and decreases τ_e. The typical lifetimes in stabilized a-Se are in the range 10–500 μs for holes and 100–1500 μs for electrons [45.21].

The fractional increase in the τ_e with As addition is greater than the drop in μ_e. Thus the electron range ($\mu_e\tau_e$) increases with As content. The effect of Cl doping on the carrier ranges ($\mu\tau$ products) is more pronounced than that of As doping. Thus, we can control both electron and hole ranges by appropriately choosing the relative amounts of As and Cl in a-Se.

The electron–hole-pair creation energy $W_\pm$ in a-Se has a strong dependence on electric field F but only a weak dependence on the x-ray photon energy E [45.30, 31]. The quantity $W_\pm$ is decreased by increasing either F or E. $W_\pm$ at a given E in a-Se follows an empirical relation given by

$$W_\pm \approx W_\pm^0 + \frac{B(E)}{F^n} , \qquad (45.2)$$

where $B(E)$ is a constant that depends on E, $W_\pm^0$ is the saturated EHP creation energy (at infinite F), and n is typically in the range 0.7–1 [45.32]. The value of $W_\pm^0$ should be $2.2E_g + E_{phonon}$ [45.33], where E_{phonon} is a phonon energy term. With $E_g \approx 2.1$ eV and $E_{phonon} < 0.5$ eV, we would expect that $W_\pm^0 \approx 5$ eV. The energetic primary electron generates many EHPs but only a certain fraction of these are free to drift and the rest recombine before they can contribute to the photocurrent. There are various possible explanations for the F dependence of the EHP creation energy. First, the simultaneously generated electron and its hole twin are attracted to each other by their mutual Coulombic force and may eventually recombine. This type of recombination is called geminate recombination (from Gemini – the twins). Another possible mechanism is columnar recombination that involves the recombination of nongeminate electrons and holes generated close to each other in the columnar track of the single high-energy electron (primary) created by the absorption of an x-ray. In both the geminate and columnar cases, the number of carriers escaping recombination should increase with increasing F, which acts to separate the oppositely charged carriers. The question of whether the F dependence of $W_\pm$ in a-Se is dominated by geminate or columnar recombination has not been fully resolved and is currently an area of active research [45.33, 34]. However, the energy dependence of $W_\pm$ is better understood. It decreases slowly with increasing photon energy in the diagnostic [45.30] and megavolt range [45.35]. The total change in $W_\pm$ from 20 keV to 6 MeV is of the factor of 3. This appears to be due to a reduction of recombination with increase in energy. The rate of deposition of energy per unit distance traveled by a primary electron decreases as a function of energy, decreasing the density of EHPs in the column around it. This is expected to reduce columnar recombination – as is seen. Thus, it appears that at low energies the contribution from columnar recombination is approximately twice that from geminate, but at high enough energy the columnar effect is reduced to zero. The typical field value used in a-Se devices is 10 V/μm where the value $W_\pm$ is 35–55 eV over the diagnostic beam energy (12–120 keV) and ≈ 15 eV at megavolt energies.

The dark resistivity of a-Se is often quoted roughly as 10^{14} Ωcm. However, experimental results in the literature have shown that the main source of the dark current is the injection of holes from the metal contacts and the bulk thermal generation current is negligible [45.36, 37]. It was the development of multilayer p-i-n type device structures that eventually reduced the dark current to very low levels (< 0.1 nA/cm^2 at F as high as 20 V/μm [45.38]) and made an a-Se commercially viable detector [45.39]. A pin (p-like-intrinsic-n-like) a-Se structure is shown in Fig. 45.5, where the thickness of both the p-like and n-like blocking contact layers is a few microns. These p- and n-like-layers are appropriately doped to serve as unipolar conducting layers that easily trap electrons and holes respectively, but allow the transport of oppositely charged carriers. The rate of emission of these deeply trapped carriers is so small that there is no significant current injection into the bulk a-Se layer. The commercial a-Se detector (in this example, the Analogic detector) is n-i-p or p-i-n type where the p-like layer is usually a-As_2Se_3 and the n-like layer is alkali metal doped a-Se [45.36]. Recent work has

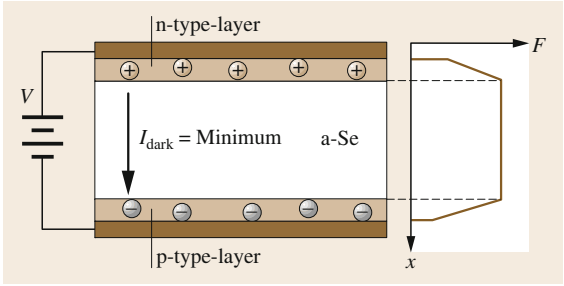


Fig. 45.5 A multilayer pin-type a-Se device structure for blocking the dark current

shown that an n-i structure with a cold deposited n-like layer also has a very low dark current [45.39]. With these n- and p-like layers, the field at the metal electrodes is sufficiently small to minimize charge injection from the contacts, which substantially reduces the dark current.

The image lag in a-Se detectors is under 2% after 33 ms and $< 1\%$ after 0.5 s in the fluoroscopic mode of operation [45.40]. Therefore, image lag in a-Se detectors is considered to be negligible. The pixel-to-pixel sensitivity variation is also negligible in a-Se detectors. The presampling MTF of these detectors is almost equal to the theoretical MTF (sinc function) determined by the pixel aperture [45.40].

Avalanche Gain a-Se

Amorphous selenium is one of the exceptional amorphous semiconductors, which exhibits usable impact ionization, i. e., at a very high field F (above $\approx 70\,\mathrm{V/\mu m}$) holes in a-Se can gain enough energy to create additional EHPs through impact ionization with a useful avalanche gain of 1000 or more [45.41]. Based on this hole impact ionization, a practical vacuum device, the high-gain avalanche rushing photoconductor (HARP) optical image sensor, has been developed [45.42]. Recently, there has been an intense interest to replace the electron beam of HARP by a two-dimensional array of pixel electrodes utilizing avalanche selenium detectors for low-dose medical x-ray imaging [45.43]. The avalanche multiplication may increase the signal strength and improve the signal to noise ratio in low dose x-ray imaging applications. However, the dark current in avalanche detectors can be high and very critical because of extremely high fields and the avalanche nature of the dark current. Electrons start impact ionization at a field of $\approx 100\,\mathrm{V/\mu m}$ in a 30 μm thick a-Se detector structure (it is $110\,\mathrm{V/\mu m}$ for the thickness of 10 μm) [45.44]. Once electron impact ionization starts, a self-sustaining avalanche breakdown process is initiated and the large number of multiplied charge carriers act to screen the external bias. Therefore, $110\,\mathrm{V/\mu m}$ field and 30 μm thickness are considered as the practical limit for an avalanche a-Se detector for x-ray imaging applications.

The potential avalanche selenium detector structures for solid state flat-panel digital x-ray imaging are shown in Fig. 45.6. The prospective detector structures are classified as; (a) type 1: cerium dioxide (CeO_2) hole blocking and resistive interface layer (RIL) electron blocking layers, and Au bottom electrode, and (b) type 2: indium gallium zinc oxide (IGZO) hole blocking layer and Au bottom electrode. The top electrode for all the structures is positively biased indium-tin-oxide (ITO). The thicknesses of the CeO_2 and IGZO layers are ≈ 10–30 and 375 nm respectively. Both CeO_2 and IGZO are n-type wide bandgap (bandgap $E_g = 3.3$ and 3.7 eV for CeO_2 and IGZO respectively) materials [45.42, 45]. Their bandgaps and electron affinities are such that they both block hole injection from electrodes but don't block the flow of photogenerated electrons from the a-Se layer to the positive electrode [45.45, 46]. The RIL layer is 1 μm thick and it is a semi-insulating polymer, namely cellulose acetate. RIL blocks electron injection and prevents gold diffusion into the a-Se structure [45.47]. Among the reported avalanche detector structures, the type 2 structure has shown the minimum steady-state dark current ($1\,\mathrm{pA/mm^2}$ at $F = 60\,\mathrm{V/\mu m}$). Another advantage of the type 2 structure is that it doesn't block the flow of photogenerated electrons from the a-Se layer to the positive electrode.

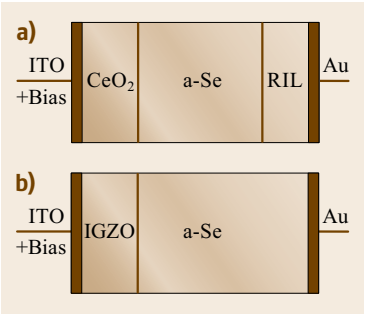


Fig. 45.6a,b Schematic diagrams of a-Se avalanche detector structures with ITO and Au top and bottom electrodes respectively. **(a)** Type 1: Cerium dioxide (CeO_2) hole blocking layer and resistive interface layer (RIL) as electron blocking layer. **(b)** Type 2: Indium gallium zinc oxide (IGZO) hole blocking layer. (After [45.48])

The avalanche multiplication gain g depends on the photoconductor layer thickness L and the impact ionization coefficient of the carrier. If only holes undergo impact ionization, the carrier multiplication is $\exp(\alpha_h L)$, where α_h is the impact ionization coefficient for holes. The field dependence of the α_h, at least over the limited fields where avalanche is observed, can be modeled by the following empirical relation

$$\alpha_h = \alpha_0 \exp\left(-\frac{F_0}{F}\right), \tag{45.3}$$

where α_0 and F_0 are the fitting parameters. *Reznik* et al. [45.41] carefully measured α_h for holes in various samples at room temperature and plotted as α_h versus $1/F$. The experimental results for different samples are surprisingly quite close as shown in Fig. 45.7. The symbols represent the experimental results and the solid line represents (45.3) with $\alpha_0 = 1.1 \times 10^4\,\mu\mathrm{m}^{-1}$ and $F_0 = 1.09 \times 10^3\,\mathrm{V/\mu m}$.

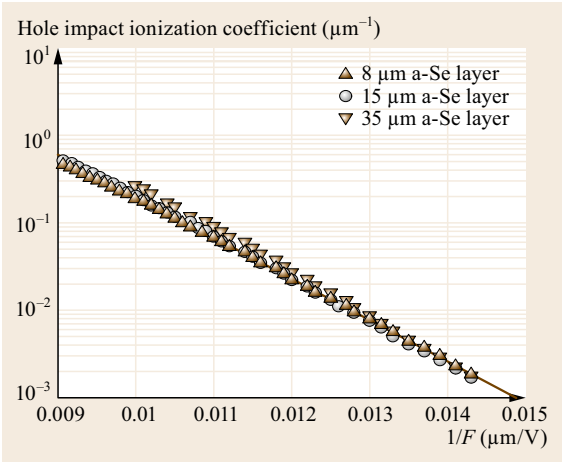


Fig. 45.7 The hole impact ionization coefficient as a function of electric field. *Symbols*: experimental data. (After [45.41]). *Solid line*: empirical equation (45.3)

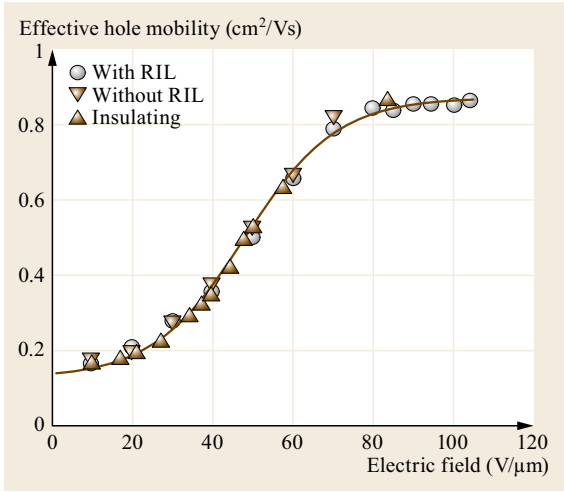


Fig. 45.8 The effective hole drift mobility as a function of the applied electric field at room temperature. *Symbols*: experimental data. (After [45.47]). *Solid line*: empirical equation (45.4)

The effective mobility of holes at extremely high fields increases with increasing field. The effective hole mobility at room temperature T_r is well defined and almost independent of the preparation conditions of the sample [45.47, 49]. An empirical relation for the effective hole mobility in a-Se at T_r can be expressed by fitting the experimental results, which is [45.50]

$$\mu_h\,(F, T_r) \approx 0.127 + \frac{0.745}{1 + \exp\left[-\frac{(F-48)}{11.5}\right]}\,, \qquad (45.4)$$

where F is the electric field in V/μm and μ_h is in cm^2/Vs. Figure 45.8 shows the effective hole drift mobility as a function of applied electric field at room temperature. The symbols and solid line represent the experimental results and the empirical relation (45.4) respectively. The experimental field-dependent effective mobility data have been extracted from Fig. 3 of [45.47]. The drift mobility was measured in ITO/CeO_2/a-Se/Sb_2S_3/Au and ITO/CeO_2/a-Se/Sb_2S_3/RIL/Au detector structures. As evident from Fig. 45.8, the above empirical expression gives a close fit to the experimental results. The physical models for the field-dependent mobility and impact ionization coefficients are also described in the literature [45.50].

Mercuric Iodide (HgI_2)

Polycrystalline HgI_2 (poly-HgI_2) has been used as a photoconductor layer in x-ray image detectors. It has been prepared by either physical vapor deposition (PVD) or screen printing (SP) from a slurry of HgI_2 crystals using a wet particle-in-binder process [45.26]. There appears to be no technological barrier to preparing large-area layers, and direct-conversion x-ray AMFPI of $20 \times 25\,cm^2$ (1536×1920 pixels) and $5 \times 5\,cm^2$ (512×512 pixels) size have been demonstrated using PVD [45.26] and SP poly-HgI_2 layers. The prototype HgI_2 sensors can potentially be used for radiation therapy [45.51], mammographic [45.52], fluoroscopic or radiographic imaging. There has been active research to improve the material properties of poly-HgI_2-based image detectors including improving the nonuniformity by reducing the grain size [45.53]. The bandgap energy $E_g = 2.1\,eV$, the ionization energy $W_\pm \approx 5\,eV$ and the density of poly-HgI_2 is $6.3\,g/cm^3$. The resistivity of this material is $\approx 4 \times 10^{13}\,\Omega cm$ [45.54].

The HgI_2 layer is deposited onto either conductive (indium-tin-oxide-(ITO) or gold-coated) glass plates or a-Si TFT arrays. HgI_2 tends to react chemically with most metals; hence a thin blocking layer (typically, $\approx 1\,\mu m$ layer of insulating polymer) is used between the HgI_2 layer and the pixel electrodes to prevent the reaction and also to reduce the dark current. The HgI_2 layer thickness varies in the range 100–500 μm and the grain size is 20–60 μm. Several hundred Å of palladium (Pd) or Au are deposited (by direct evaporation) on top of the HgI_2 layer followed by a polymer encapsulation layer to form a bias electrode.

The dark current of HgI_2 imagers increases superlinearly with the applied bias voltage [45.55]. For PVD HgI_2 detectors the dark current depends strongly on the operating temperature. It increases by a factor of about two for each 6 °C of temperature rise. It is reported that the dark current varies from $\approx 2\,pA/mm^2$ at 10 °C to $\approx 180\,pA/mm^2$ at 35 °C at $F = 0.95\,V/\mu m$, which is not desirable for medical imagers (the maximum dark

current for medical imaging should be $< 10\,\mathrm{pA/mm^2}$). Therefore, PVD HgI_2 imagers should be operated at relatively low bias (preferably less than $\approx 0.5\,\mathrm{V/\mu m}$) and relatively low temperature ($< 25\,^\circ\mathrm{C}$). The dark current in the SP sample is an order of magnitude smaller than in the PVD sample and more stable against temperature variation. It is possible to keep the dark current below $10\,\mathrm{pA/mm^2}$ at temperatures up to $35\,^\circ\mathrm{C}$ and $F = 1\,\mathrm{V/\mu m}$ [45.55]. A major disadvantage of SP detectors is that they show 2–4 times less sensitivity compared to PVD detectors.

Electrons have much longer range than holes in HgI_2 and thus the radiation-receiving electrode is negatively biased to obtain a higher sensitivity. The $\mu_e\tau_e$ for electrons in the SP HgI_2 is in the range 10^{-6}–$10^{-5}\,\mathrm{cm^2/V}$ [45.26, 56], and $\mu_e\tau_e$ in the PVD sample is about an order of magnitude larger. Recently it was reported that $\mu_e\tau_e$ in PVD HgI_2 is in the range 10^{-5}–$10^{-4}\,\mathrm{cm^2/V}$, which is almost equal to that of single-crystal HgI_2 [45.24, 57]. The reason is that the PVD HgI_2 layer grows in a columnar structure perpendicular to the substrate. Thus, a charge carrier may drift along a column without having to pass through grain boundaries, where it would encounter excess trapping and/or recombination. Samples with larger grain sizes may have fewer grain-boundary defects. Hence there is a trend of increasing $\mu_e\tau_e$ with grain size in the PVD samples. However, this trend is not observed in the SP sample, which indicates another mechanism is responsible for its low $\mu_e\tau_e$. Larger grain sizes may cause nonuniform sensor response. The grain sizes must be much smaller than the pixel size to get a uniform response. $\mu_h\tau_h$ in poly-HgI_2 is of the order $\approx 10^{-7}\,\mathrm{cm^2/V}$ [45.56].

Two important drawbacks of polycrystalline sensors are the image lag and the pixel-to-pixel sensitivity variation or nonuniform response. The lowest image-lag characteristics reported are $\approx 7\%$ first-frame lag, $\approx 0.8\%$ after 1 s and $\approx 0.1\%$ at 3 s in fluoroscopic mode (15 frames/s). The pixel-to-pixel sensitivity variation reduces the dynamic range of the imagers. The relative standard deviation of the sensitivity (standard deviation/average value) in the latest HgI_2 AMFPI is $\approx 10\%$. It is reported that HgI_2 image detectors with smaller grain sizes show good sensitivity and also an acceptably uniform response. The presampling MTF of these detectors is almost equal to the theoretical MTF (sinc function) determined by the pixel aperture.

Cadmium Zinc Telluride (CdZnTe)

CdZnTe ($< 10\%$ Zn) polycrystalline film has been used as a photoconductor layer in x-ray AMFPI. CdZnTe is commonly called CZT. Although CZT can be deposited on large areas, direct-conversion AMFPI of only $7.7 \times 7.7\,\mathrm{cm^2}$ (512×512 pixels) have been demonstrated using polycrystalline CZT (poly-CZT). The CZT layer thickness varies in the range $200-500\,\mu\mathrm{m}$ [45.58]. Temporal lag and nonuniform response were noticeable in early CZT sensors. Large and nonuniform grain sizes are believed to be responsible for the temporal lag and nonuniform response of the sensor. Recent studies show that Cl doping makes a finer and more uniform grain structure [45.25]. The ionization energy $W_\pm \approx 5\,\mathrm{eV}$, and the density of $Cd_{0.95}Zn_{0.05}Te$ is $5.8\,\mathrm{g/cm^3}$. The bandgap energy E_g of $Cd_{0.95}Zn_{0.05}Te$ is 1.7 eV and the resistivity of this material is $\approx 10^{11}\,\Omega\mathrm{cm}$ [45.59].

Typically CZT is first coated onto an inert substrate and then attached to the active matrix array. The advantage of this technique is that the electrical properties and hence imaging performances of the detector can be optimized without causing any thermal or chemical damage to the TFT array. The CZT film is deposited by the close-spaced sublimation method [45.60] onto an alumina (Al_2O_3) substrate coated with ITO, which forms the bias (top) electrode. A cross section of the poly-CZT detector is shown in Fig. 45.9. The several-microns-thick ZnTe layer acts as a barrier to electron injection and hence suppresses dark current under the negative bias. This ZnTe layer is omitted for the positive bias. Conducting resin bumps on the pixel electrodes connect each pixel to the CZT layer.

Introduction of Zn into the CdTe lattice increases the bandgap, decreases the conductivity and hence reduces the dark current. In CZT μ_h decreases with increasing Zn concentration whereas μ_e remains nearly constant. Again, addition of Zn to CdTe increases lattice defects and hence reduces carrier lifetimes. The poly-CZT has a lower density, which results in a lower quantum efficiency η than its single-crystal counterpart. Although (for a detector of given thickness) x-ray sensitivity in CZT detectors is lower than in CdTe detectors, the CZT detectors show a better signal-to-noise ratio and hence give a better DQE. The measured sensitivities are higher than a-Se.

The dark current of $Cd_{0.95}Zn_{0.05}Te$ imagers increases almost linearly with F and is $\approx 70\,\mathrm{pA/mm^2}$ at

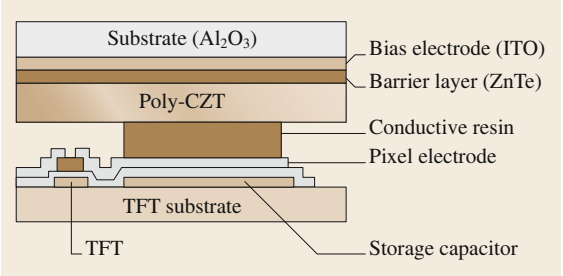


Fig. 45.9 Cross section of the polycristalline CZT detector structure. (After [45.25])

$F = 0.25\,\mathrm{V/\mu m}$ [45.27], which makes it unsuitable for applications requiring long exposure times. The dark current would be expected to decrease with increasing Zn concentration due to the increased E_g. This has in fact been demonstrated [45.25], i. e., the dark current in $Cd_{0.92}Zn_{0.08}Te$ sensors is $40\,\mathrm{pA/mm^2}$ at $F = 0.4\,\mathrm{V/\mu m}$.

The mobility lifetime products of both electrons and holes in poly-CZT are less than that in single-crystal CZT. The values of $\mu\tau$ in single-crystal $Cd_{0.9}Zn_{0.1}Te$ are in the range 10^{-4}–$10^{-3}\,\mathrm{cm^2/V}$ (electrons) and 10^{-6}–$10^{-5}\,\mathrm{cm^2/V}$ (holes) [45.61]. But the values of $\mu\tau$ in poly-$Cd_{0.95}Zn_{0.05}Te$ are $\approx 2\times10^{-4}\,\mathrm{cm^2/V}$ (electrons), and $\approx 3\times10^{-6}\,\mathrm{cm^2/V}$ (holes) [45.62, 63]. Since $\mu_e\tau_e \gg \mu_h\tau_h$ in CZT, negative bias to the radiation-receiving electrode is the preferred choice for better sensitivity and temporal response.

The relative standard deviation of the sensitivity (standard deviation/average value) in the latest CZT AMFPI is $\approx 20\%$ [45.25]. The image-lag characteristics reported are $\approx 70\%$ first-frame lag, $\approx 20\%$ after three frames and 10% at 1 s in fluoroscopic mode (30 frames/s). In single-pulse radiographic mode the first-frame lag is less than 10% [45.25]. The longer image-lag characteristics of CZT sensors in fluoroscopic mode imply that it is not yet suitable for fluoroscopic applications. The presampling MTF of CZT detector is ≈ 0.3(150 µm pixel size) at the Nyquist frequency f_N (theoretical MTF, sinc function, is ≈ 0.64 at $f_N = 3.3$ line pairs per mm), where the MTF of CsI imagers is less than 0.2 [45.27].

Lead Iodide (PbI_2)

Polycrystalline PbI_2 photoconductive layers have been prepared using PVD at a substrate temperature of 200–230 °C. A deposition of several hundred Å of palladium (Pd) is used to form the top electrode. Grains are described as hexagonal platelets with the longest dimensions being 10 µm or less. The platelets grow perpendicular to the substrate, producing films that are less dense (3–5 $\mathrm{g/cm^3}$) than bulk crystalline material (6.2 $\mathrm{g/cm^3}$). There appears to be no technological barrier to preparing large-area layers, and direct-conversion AMFPI of $20\times25\,\mathrm{cm^2}$ size (1536×1920 pixels) have been demonstrated [45.64]. Coating thickness varies in the range 60–250 µm and prototype PbI_2 imagers have been used for radiographic imaging [45.64]. The bandgap energy $E_g = 2.3$ eV, and the ionization energy $W_\pm \approx 5$ eV. The resistivity of this material is in the range 10^{11}–$10^{12}\,\Omega$cm.

Lead iodide detectors have a very long image-lag decay time. The image lag depends on the exposure history. The image-lag characteristics reported are $\approx$ 75% first-frame lag, $\approx 15\%$ after 3 s in fluoroscopic mode (15 frames/s), whereas in single-pulse radiographic mode the first-frame lag is less than 50% and drops below 1% within 1 s [45.64]. The long image-lag characteristics of PbI_2 in fluoroscopic mode imply that it is unsuitable for fluoroscopic applications.

The dark current of PbI_2 imagers increases sublinearly with the applied bias voltage. The dark current is in the range 10–50 $\mathrm{pA/mm^2}$ at $F = 0.5\,\mathrm{V/\mu m}$ [45.64] and 0.1–0.45 $\mathrm{nA/mm^2}$ at $F = 1.0\,\mathrm{V/\mu m}$ [45.65], much higher than PVD HgI_2 detectors, making it unsuitable for long-exposure-time applications. The presampling MTF is ≈ 0.35 (127 µm pixel size) at f_N (theoretical MTF, sinc function, is ≈ 0.64 at $f_N = 3.93$ lp/mm), where the MTF of CsI imagers is less than 0.2 [45.64]. The resolution of PbI_2 imagers is acceptable but slightly worse than that of HgI_2 imagers. Also, the x-ray sensitivity of PbI_2 imagers is lower than that of HgI_2 imagers. The pixel-to-pixel sensitivity variation in PbI_2 imagers is substantially lower.

The $\mu\tau$ product of holes and electrons in PVD PbI_2 are $1.8\times10^{-6}\,\mathrm{cm^2/V}$ and $7\times10^{-8}\,\mathrm{cm^2/V}$ respectively [45.64]. μ_h in poly-PbI_2 is in the range 0.02–0.15 $\mathrm{cm^2/Vs}$ whereas μ_h in single-crystal PbI_2 is 2 $\mathrm{cm^2/Vs}$ [45.66]. This indicates that μ_h in poly-PbI_2 is controlled by shallow traps, probably introduced at the grain boundaries.

Lead Oxide (PbO)

The large-area deposition requirement is compatible with the use of polycrystalline PbO (poly-PbO) film as a photoconductor layer in AMFPI. Direct-conversion flat-panel x-ray imagers of $18\times20\,\mathrm{cm^2}$ (1080×960 pixels) from a poly-PbO with film thicknesses of $\approx$ 300 µm have been demonstrated [45.28]. One advantage of PbO over other x-ray photoconductors is the absence of heavy-element K-edges for the entire diagnostic energy range up to 88 keV, which suppresses additional noise and blurring due to the K-fluorescence. The ionization energy $W_\pm \approx 8$ eV and the density of crystalline PbO is 9.6 $\mathrm{g/cm^3}$ whereas its density is reduced by almost 50% in poly-PbO. The bandgap energy E_g of PbO is 1.9 eV and the resistivity of this material is in the range $(7-10)\times10^{12}\,\Omega$cm [45.28].

Lead oxide photoconductive polycrystalline layers have been prepared by thermal evaporation in a vacuum chamber at a substrate temperature of ≈ 100 °C. An evaporated layer of Al, Au or Pd was used to form a top electrode of thickness 100–200 nm [45.67]. The PbO layer consists of very thin platelets of a few microns thickness and has a density of $\approx 50\%$ of the single-crystal density. PbO slowly degrades if it is exposed to air under normal ambient temperature but a few hours exposure is acceptable. However, in the long term PbO reacts with water and CO_2 causing

an increase in dark current and a decrease in x-ray sensitivity. Therefore, a polymer or a semiconductor (e.g., doped a-Se) passivation layer is used to prevent exposure to the atmosphere. This is usually placed between the photoconductor layer and the top metal electrode. The passivation layer avoids degradation of the PbO layer and reduces the dark current [45.67]. The dark current in PbO sensors is $\approx 40\,\mathrm{pA/mm^2}$ at $F = 3\,\mathrm{V/\mu m}$ [45.28]. The $\mu_e\tau_e$ and $\mu_h\tau_h$ in poly-PbO are 3.5×10^{-7} and $\approx 10^{-8}\,\mathrm{cm^2/V}$ respectively [45.68]. The lag signal in fluoroscopic mode is in the range 3–8% after 1 s. The presampling MTF of a PbO detector is ≈ 0.5 (184 µm pixel size) at f_N (theoretical MTF, sinc function, is ≈ 0.64 at $f_N = 2.72\,\mathrm{lp/mm}$) [45.28].

Thallium Bromide (TlBr)

Polycrystalline thallium bromide (poly-TlBr) has not yet been used in an AMFPI but has been used in a large-area (9 in-diameter) direct-conversion detector called an x-ray-sensitive electron-beam image tube (XEBIT). The operational principle of the XEBIT is similar to the standard light-sensitive vidicon that was utilized extensively in the commercial television industry. The XEBIT can replace an x-ray image intensifier coupled to a video camera using relay lenses with a single direct-conversion device [45.69]. The typical TlBr layer thickness is 300 µm.

The ionization energy $W_\pm \approx 6.5$ eV and the relative dielectric constant of TlBr is 31 [45.70]. The bandgap energy E_g of TlBr is 2.7 eV and its resistivity is $\approx 5\times10^9\,\Omega$cm under ambient conditions. At room temperature the dominant contribution to the dark current is ionic conductivity [45.69]. The ionic conductivity has an exponential dependence on temperature; the conductivity decreases by an order of magnitude for every 19 °C temperature decrease. Therefore, the dark current can be greatly decreased by Peltier cooling. $\mu_h\tau_h$ and $\mu_e\tau_e$ in TlBr are $\approx 1.5\times10^{-6}$ and $\approx 4\times10^{-7}\,\mathrm{cm^2/V}$ respectively [45.69, 70].

Organic Perovskites

Yakunin et al. [45.71] reported that a thick layer ($\approx$ 100 µm) of polycrystalline methylammonium lead iodide (poly-$MAPbI_3$ where MA is CH_3NH_3) perovskite (grain size is larger than 0.25 µm) can be uniformly deposited over large areas using a solution-based synthesis technique without affecting the underlying AMA electronics. This material also shows a reasonable x-ray absorption coefficient and its x-ray sensitivity is comparable to a-Se [45.71]. Therefore, $MAPbI_3$ can be a potential candidate for large-area AMFPI and thus opens the door to a new class of perovskite large-area x-ray sensors. $MAPbI_3$ is a direct bandgap semiconductors of bandgap energy of 1.6 eV and the relative dielectric constant of 28. Its density is $4.3\,\mathrm{g/cm^3}$ [45.72]. The electron and hole mobilities in polycrystalline films are ≈ 8 and $\approx 15\,\mathrm{cm^2/Vs}$ respectively [45.71]. The dark resistivity is $\approx 10^9\,\Omega$cm. The effective masses of electrons and holes are 0.23 m_0 and 0.29 m_0 respectively [45.73], where m_0 is the free electron mass.

Although the reported 60 µm-thick detector structure [45.71] shows a reasonable level of photocurrent for the x-ray fluence of $1.4\times10^7\,\mathrm{photons/mm^2s}$, the dark current is almost half of the photocurrent. Moreover, the dark current increases linearly with the applied field and the detectors shows a photoconductive gain, which indicates that the contacts are ohmic-type. As a result, the dark current could be unacceptably large at high applied field (the high field is needed to achieve acceptable charge collection efficiency and x-ray sensitivity). Though the ohmic contact is required for the photovoltaics, the x-ray detectors need blocking contacts in order to minimize the noise and enhance the DQE. Therefore, the optimization of the structure (for example, the appropriate metal contacts and the blocking layers) is essential for their use in practical AMFPIs.

45.1.3 Summary and the Future

The material and imaging properties of potential photoconductors for x-ray image detectors are summarized in Tables 45.1 and 45.2. Stabilized a-Se is currently the best choice of photoconductor for mammographic x-ray image detectors. The next closest competitor for both mammography and general radiology is the poly-HgI_2 imagers, which show excellent sensitivity, good resolution, and acceptable dark current, homogeneity and lag characteristics. However, the long-term stability of HgI_2 imagers has not been as thoroughly studied as stabilized a-Se detectors. Both the dark current and the image-lag characteristics of CZT, PbI_2 and PbO detectors are worse than those of HgI_2 detectors. However, the x-ray detectors made with CZT photoconductive layers should be mechanically and chemically more stable than HgI_2-based detectors. The main drawback of a-Se detectors are their low conversion gain, which particularly affects the imaging sensor performance at low exposure. This can be overcome by utilizing the avalanche multiplication technique in the a-Se layer [45.74] and/or using on-pixel amplification [45.2]. However, further research is necessary to demonstrate basic operation as well as to examine the long-term stability of detectors utilizing these techniques. The main drawbacks of polycrystalline sensors are the image lag and the nonuniform response. Making smaller, finer and more uniform grain size in polycrystalline sensors may overcome these drawbacks.

Table 45.1 Material properties

Photoconductor, state and preparation	E_g (eV)	$W_{\pm}$ (eV)	Density (g/cm^3)	Resistivity (Ωcm)	Electrons μ_e(cm^2/Vs), $\mu_e\tau_e$(cm^2/V)	Holes μ_h(cm^2/Vs) $\mu_h\tau_h$(cm^2/V)
Stabilized a-Se, vacuum deposition	2.1–2.2	≈ 45 at 10 V/μm	4.3	10^{14}–10^{15}	$\mu_e =$ 0.003–0.006 $\mu_e\tau_e =$ 0.3×10^{-6}–10^{-5}	$\mu_h = 0.12$ $\mu_h\tau_h =$ 10^{-6}–6×10^{-5}
HgI_2, polycrystalline, PVD	2.1	5	6.3	$\approx 4\times10^{13}$	$\mu_e = 88$ $\mu_e\tau_e \approx$ 10^{-5}–10^{-4}	$\mu_h = 3$–4 $\mu_h\tau_h \approx 10^{-6}$
HgI_2, polycrystalline, SP	2.1	5	6.3	$\approx 4\times10^{13}$	$\mu_e\tau_e \approx$ 10^{-6}–10^{-5}	$\mu_h\tau_h \approx 10^{-7}$
$Cd_{0.95}Zn_{0.05}Te$, polycrystalline, vacuum deposition	1.7	5	5.8	$\approx 10^{11}$	$\mu_e\tau_e \approx 2\times10^{-4}$	$\mu_h\tau_h \approx 3\times10^{-6}$
PbI_2, polycrystalline, PVD	2.3	5	3–5	10^{11}–10^{12}	$\mu_e\tau_e = 7\times10^{-8}$	$\mu_h = 0.02$–0.15, $\mu_h\tau_h \approx 2\times10^{-6}$
PbO, polycrystalline, vacuum deposition	1.9	8–20	4.8	7–10×10^{12}	$\mu_e\tau_e \approx 3.5\times10^{-7}$	$\mu_h\tau_h \approx 10^{-8}$
TlBr, polycrystalline	2.7	6.5	7.5	$\approx 5\times10^{9}$ at 20 °C	$\mu_e\tau_e \approx 4\times10^{-7}$	$\mu_h\tau_h \approx 1.5\times10^{-6}$
$MAPbI_3$, polycrystalline	1.6	5	4.3	$\approx 10^{9}$	$\mu_e = 8$	$\mu_h = 15$

Table 45.2 Imaging properties

Photoconductor, state and preparation	Typical operating F (V/μm)	Dark current (pA/mm^2)	Lag (fluoroscopic mode of operation)	Uniformity/ sensitivity variation (standard deviation/ average value)
Stabilized a-Se, single layer	≈ 10	< 10 up to $F = 20$ V/μm	$< 2\%$ after 33 ms	Negligible
Stabilized a-Se, multilayer (pin or nip)	≈ 10	< 1 up to $F = 20$ V/μm	$< 2\%$ after 33 ms	Negligible
HgI_2, polycrystalline, PVD	≈ 0.5	≈ 6 at $F = 0.5$ V/μm	$\approx 7\%$ after 66 ms	$\approx 10\%$
HgI_2, polycrystalline, SP	≈ 1.0	≈ 8 at $F = 1.0$ V/μm	$\approx 7\%$ after 66 ms	$\approx 10\%$
$Cd_{0.95}Zn_{0.05}Te$, polycrystalline	≈ 0.25	≈ 25 at $F = 0.25$ V/μm	$\approx 70\%$ after 33 ms	$\approx 20\%$
PbI_2, polycrystalline, PVD	≈ 0.5	10–50 at $F = 0.5$ V/μm	$\approx 75\%$ after 66 ms	?
PbO, polycrystalline	≈ 1.0	40 at $F = 3$ V/μm field	3–8% after 1 s	?
TlBr, polycrystalline	≈ 0.5	?	?	?

45.2 Dark Current Limitations

As mentioned earlier, a large electric field (e.g., ≈ 10 V/μm in conventional a-Se detectors) is applied across the x-ray detectors in order to collect photogenerated EHPs [45.43]. One of the most important attributes of a photoconductor is the dark current, which should be negligibly small, since it is a source of noise and thus reduces the detectivity and dynamic range of the detector. The dark current in detectors is usually very high right after applying the bias voltage and decays with time. In most of the cases, it reaches a plateau within the time range of 100–10 000 s [45.36, 37, 47].

The dark current in x-ray detectors may have three origins:

1. The depletion of carriers from the defect states within the bandgap after applying the bias
2. Steady-state thermal generation of carriers in the bulk

3. Carrier injections from the metal contacts towards the photoconductor layer.

45.2.1 Charge Carrier Depletion

The depletion of the majority carrier from the bulk and interface defect states within the mobility gap can constitute a transient dark current decay behavior. After applying the bias, majority carriers are depleted and the steady-state quasi-Fermi level E_{FD} becomes slightly different from the Fermi level E_F. Let us consider a-Se as an example. a-Se is slightly p-type and E_F in a-Se at zero bias is slightly below the midgap. After applying the bias, holes are depleted and the steady-state quasi-Fermi level E_{FD} lies above E_F as shown in Fig. 45.10. The temporal behavior of the carrier depletion process is determined by the detrapping time constants. The time-dependent hole depletion rate due to carrier detrapping from the bulk a-Se is [45.48, 75]

$$g_d(t) = \int_{E_v}^{E_c} \frac{N(E)}{\tau_r(E)} [f_{FD}(E) - f_F(E)] \exp\left[-\frac{t}{\tau_r(E)}\right] dE \,, \tag{45.5}$$

where

$$f_F(E) = \frac{1}{1 + \exp\left[\frac{(E-E_F)}{k_B T}\right]}, \tag{45.6}$$

$$f_{FD}(E) = \frac{1}{1 + \exp\left[\frac{(E-E_{FD})}{k_B T}\right]}. \tag{45.7}$$

The mean detrapping or mean release time constant is

$$\tau_r(E) = \nu^{-1} \exp\left[\frac{\left(E - E_V - \beta_{pf}\sqrt{F}\right)}{k_B T}\right], \tag{45.8}$$

where $N(E)$ is the density of states (DOS) of the photoconductor at energy E in the midgap, ν is the attempt-to-escape frequency, k_B is the Boltzmann constant, T is the absolute temperature, t is the instantaneous time (in seconds), $\beta_{pf} = \sqrt{e^3/\pi\varepsilon_s}$ is the Poole–Frenkel coefficient, e is the elementary charge, ε_s is the permittivity of the photoconductor, and E_C and E_V are the conduction and valence band edges respectively. Equation (45.5) represents the rate of hole depletion per unit volume, which decays with time.

The depleted holes drift under the influence of the electric field and induce a current in the detector. Currents resulting from the drifting of carriers in the photoconductive detectors are due entirely to induction, which can conveniently be calculated by using the Shockley–Ramo theorem [45.76]. The transient current density due to the hole depletion from the bulk is [45.77],

$$J_{dpb}(t) = \frac{eL}{2} g_d(t) \,. \tag{45.9}$$

Similarly, the time-dependent hole depletion current density due to carrier detrapping from the interface (e.g., the top interface at positive bias to the top electrode) is [45.48],

$$J_{dpi}(t) = e \int_{E_v}^{E_c} \frac{D_i(E)}{\tau_r(E)} [f_{FD}(E) - f_F(E)] \exp\left[-\frac{t}{\tau_r(E)}\right] dE \,, \tag{45.10}$$

where $D_i(E)$ is the density of interface states ($cm^{-2}\,eV^{-1}$). Note that (45.5)–(45.10) can also applied to slightly n-type photoconductors (e.g., HgI_2) by interchanging f_F and f_{FD} [45.78].

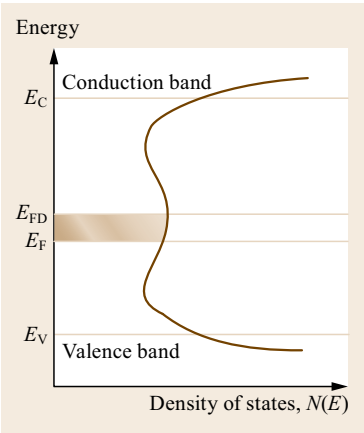


Fig. 45.10 A schematic representing the carrier depletion from the mobility gap in slightly p-type amorphous/polycrystalline materials after applying the bias. The *shaded area* represents the approximate amount of depleted carriers. The figure shows an arbitrary DOS for an amorphous material. (After [45.48])

45.2.2 Steady-State Thermal Generation

The defect states close to the middle of the bandgap of the x-ray photoconductor have a significant probability for thermal excitation of both types of carriers. The bandgap/mobility gap of x-ray photoconductors is quite large and thus the band-to-band thermal carrier generation is expected to be negligible. Therefore, the steady-state thermal generation rate is dominated by the emission from traps within $k_B T$ of E_{FD} as shown in Fig. 45.11. If the excitation rates for electrons and holes are equal, E_{FD} is very close to the middle of the bandgap. The generation rate for a fully depleted sample is determined by the average carrier release time and

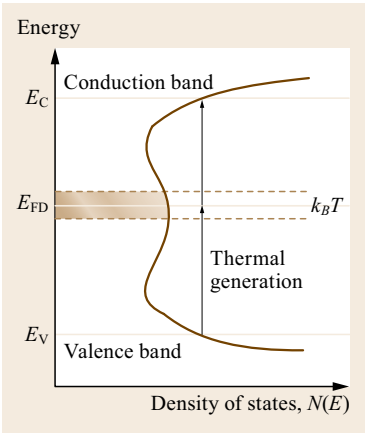


Fig. 45.11 A Schematic representing the thermal generation of carriers from the mobility gap of amorphous/polycrystalline materials near the quasi-Fermi level E_{FD}. (After [45.48])

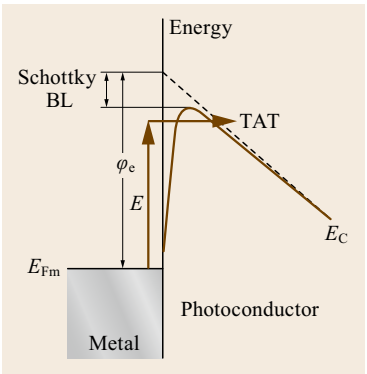


Fig. 45.12 A schematic representing the electron injection process from the metal towards the photoconductor layer. TAT and BL stand for the thermally assisted tunneling and barrier lowering respectively. (After [45.48])

can be written as [45.48],

$$g = N(E_{FD}) k_B T \nu \exp\left[-\frac{\left(E_C - E_{FD} - \beta_{pf}\sqrt{F}\right)}{k_B T}\right], \tag{45.11}$$

where $N(E_{FD})$ is the density of states at energy E_{FD} in the midgap. It is assumed in (45.11) that the density of states is constant over $k_B T$ near E_{FD}. Assuming that the liberated carriers are not lost by trapping or recombination (any loss of carriers can be reflected in the effective value of g), the steady-state thermal generation current is

$$J_{th}(t) = egL\,. \tag{45.12}$$

45.2.3 Carrier Injection

The metal/photoconductor contacts in detectors are generally blocking in nature and the carrier injection is mainly controlled by the Schottky emission. Once the carriers are injected into the photoconductive layer, they move by the drift mechanisms (the diffusion component of current is negligible compared to its drift component because of the very high applied bias) [45.79]. Therefore, the injected current densities through the metal/photoconductor contacts are [45.79],

$$J_e(t) = eN_C v_{de}(t)\left(\frac{v_r}{v_{de}+v_r}\right) \times \exp\left[-\frac{\phi_e - \Delta\phi_e(F)}{k_B T}\right] \tag{45.13}$$

and

$$J_h(t) = eN_V v_{dh}(t)\left(\frac{v_r}{v_{dh}+v_r}\right) \times \exp\left[-\frac{\phi_h - \Delta\phi_h(F)}{k_B T}\right], \tag{45.14}$$

where $v_r = A^* T^2/eN_C$ is the thermal velocity, A^* is the effective Richardson constant, $N_{V(C)}$ is the effective density of states in the valence (conduction) band, $v_d(t) \approx \mu F$ is the instantaneous drift velocity of the injected carriers, μ is the effective mobility of carriers, and ϕ is the effective barrier height for injecting carriers from the metal into the photoconductor layer, considering the effect of surface states. The subscripts e and h stand for electrons and holes respectively. For the potential x-ray photoconductors, $v_r \gg v_{dh}$ or v_{de}, the term inside the square bracket of (45.13) and (45.14) becomes unity, and the injection current is dominated by v_{dh} or v_{de} depending on the particular photoconductor (e.g., $v_{dh} \gg v_{de}$ in a-Se) [45.79].

The quantity $\Delta\phi(F)$ within the exponential in (45.13) and (45.14) is the field-dependent barrier lowering due to other effects such as charge rearrangement in the interfacial layer, the image force barrier lowering, and/or thermally assisted tunneling as shown in Fig. 45.12. For the metal-to-semiconductor carrier injection, the term $\Delta\phi(F)$ can be written as

$$\Delta\phi(F) = \sqrt{\frac{e^3 F}{4\pi\varepsilon_s}} + e\alpha F^2\,. \tag{45.15}$$

The first term in (45.15) is the image force barrier lowering (Schottky barrier lowering), and the second term is the optimum barrier lowering due to the thermally assisted tunneling. The potential barrier for the injecting carriers from the metal contact under a strong applied field is essentially a triangular barrier. Note that the thermally activated tunneling is much more probable than the direct tunneling. In thermally activated tunneling process, as the vibration energy E_p increases, the tunneling barrier decreases and hence the probability increases. On the other hand the population of

the energy level E_p decreases with increasing E_p proportionally to $\exp(-E_p/k_BT)$. Thus, there exists an optimum energy level from where the resultant escape probability is the maximum, which is proportional to $\exp(\alpha F^2/k_BT)$ [45.80]. The expression of α is determined by considering the tunneling probability through a triangular barrier together with Boltzmann occupation probability and given by [45.80]

$$\alpha = \frac{e\hbar^2}{24m^*\,(k_BT)^2}\,, \tag{45.16}$$

where h is the modified Plank constant and m^* is the effective mass of the photoconductor.

Concluding Remarks on Dark Current

One of the advantages of a-Se over competing large-area x-ray photoconductors is the comparatively smaller dark current under typical operating fields [45.36, 81]. The operating field refers to the applied field that ensures the x-ray generated charges are collected; i. e., the sensitivity is not limited by the charge collection efficiency [45.82]. For example, the dark current that has been recently reported for PbI_2 [45.81] is in the range 0.1–0.45 nA/mm^2 at a field of 1 V/μm. Under normal operation, the field needs to be greater than 1 V/μm to ensure good charge collection but this would also mean a dark current that is greater than 1 nA/mm^2, which is much higher than the dark current in typical a-Se detectors even at the operating field of 10 V/μm [45.36].

As mentioned earlier, the conventional a-Se detector can be either n-i-p or n-i type. The experimental results in the literature show that the main source of the dark current is the injection of holes from the metal contacts and the bulk thermal generation current is negligible [45.36, 37]. The above results are consistent with the facts that the bulk thermal generation and electron injection currents are negligible compared to the hole injection current because of a large bandgap and a very low electron mobility in a-Se respectively. *Frey* et al. [45.36] have also shown that the dark current has a strong dependence on the n-like layer thickness; it decreases abruptly with increasing the n-like layer thickness (≈ 4 to 8 μm). The carrier injection, from the metal to the semiconductor, depends on the internal electric field at the metal/a-Se interface. It is believed that the electric field right after applying the bias is almost uniform but quickly decreases at the interface due to high initial current and high carrier trapping in the blocking layers, which reduces carrier injection and hence reduces the dark current. The typical electric field profiles, after applying the bias on the n-i-p and n-i structures, are illustrated in Fig. 45.5. To reduce the contact electric field and dark current, the n-like layer must have a sufficient amount of trap centers and the energy depths of these trap centers should be ≈ 0.75–0.8 eV or more from the valence band mobility edge. The shallower trap levels are unable to retain a sufficient amount of trapped charge and the deeper trap centers create longer transient times to reach a steady state dark current. The detailed descriptions of the dark current mechanisms are given in [45.48].

At the normal operating fields (less than 10 V/μm) of conventional a-Se or poly-HgI_2 x-ray detectors, the thermal generation current is negligible as compared to the injection currents from the metallic electrodes [45.48, 78]. However, the thermal generation rate can be increased exponentially with field because of Poole–Frenkel or thermally assisted tunneling effects [45.83]. Therefore, both the injection and thermal generation currents should be considered to determine the dark current in a-Se avalanche detectors. The excessive dark current has been one of the factors that limits the highest operating field. An acceptable level of dark current up to an electric field as high as 60 V/μm in some a-Se detector structures has recently been reported [45.45]. The injection current is minimized and the dark is slightly higher than the thermal generation current in the type 2 structure in Fig. 45.6.

45.3 Metrics of Detector Performance

X-ray sensitivity, resolution in terms of modulation transfer function (MTF), detective quantum efficiency (DQE), image lag and ghosting are often considered as the metrics of imaging performance. For most practical applications, the spatial-frequency-dependent (f-dependent) detective quantum efficiency, DQE(f), is the appropriate metric of overall system performance and is unity at all f for an ideal detector. The detector performance depends critically on the photoconductor material properties such as the mobility, carrier trapping (both shallow and deep), EHP creation energy, x-ray attenuation and absorption coefficients. The material properties such as carrier mobility, EHP creation energy, x-ray attenuation and absorption coefficients in a well-defined photoconductor are almost constant, but the carrier lifetimes may vary from sample to sample. Shallow and deep trapping are particularly responsible for image lag and ghosting respectively. The effects of charge-transport properties ($\mu\tau$) and the attenuation coefficient of the photoconductor material on the detector

performance depends on L and F through the following normalized parameters

$$\Delta = \text{normalized attenuation depth (attenuation depth/thickness)} = \frac{1}{(\alpha L)},$$

$$x_e = \text{normalized electron schubweg (electron schubweg per unit thickness)} = \frac{\mu_e \tau_e F}{L}, \text{ and },$$

$$x_h = \text{normalized hole schubweg (hole schubweg per unit thickness)} = \frac{\mu_h \tau_h F}{L},$$

where α is the linear attenuation coefficient of the photoconductor, $\mu_{e(h)}$ is the mobility and $\tau_{e(h)}$ is the deep-trapping time (lifetime) of electrons (holes). Equivalently, x_e and x_h are the normalized carrier lifetimes (carrier lifetimes per unit transit time) for electrons and holes respectively. The ranges of these normalized parameters for the three most promising photoconductors (a-Se, poly-HgI_2 and poly-CZT) for use in x-ray image detectors are given in Table 45.3. The combined effects of charge-transport properties (mobility and carrier lifetime), operating conditions (F and E), photoconductor thickness, and the attenuation coefficient of the photoconductor material on the imaging characteristics (x-ray sensitivity, DQE, MTF and ghosting) are examined in the following sections. It must be emphasized that the photoconductor thickness L and the operating field F are as important to the overall performance of the detector as the material properties of the photoconductor itself, a point that will become apparent in the results presented in this chapter.

45.3.1 X-Ray Sensitivity

The x-ray sensitivity (S) of a photoconductive detector is defined as the collected charge per unit area per unit exposure of radiation and is considered an important performance measure for a superior image. High S permits the use of low detector radiation-exposure levels, which also increases the dynamic range of the AMPFI. The selection of the x-ray photoconductor is highly influenced by the value of S.

The value of S can be considered to arise in terms of three controlling factors:

1. The amount of radiation actually attenuated from the incident radiation that is useful for the generation of electron–hole pairs (EHPs), which is characterized by the quantum efficiency η of the detector and depends on the value of α (the linear attenuation coefficient) of the photoconductor and L through $\eta = 1 - e^{-\alpha L}$, where the value of α is x-ray photon-energy-dependent.
2. The generation of EHPs by x-ray interactions, which is characterized by the value of $W_{\pm}$ of the photoconductor and the average absorbed energy E_{ab} per attenuated x-ray photon of energy E, where $W_{\pm}$ depends on the material properties of the photoconductor, and E_{ab} depends on the incident x-ray photon energy [45.84] and the material properties. Note that $E_{ab} = E(\alpha_{en}/\alpha)$, where α_{en} is the energy absorption coefficient of the photoconductor, which depends on the x-ray photon energy E.
3. How much of the x-ray generated charge is actually collected in the external circuit. This is characterized by the charge-carrier drift mobilities (μ) and lifetimes (τ), of the applied F and L.

The S of an x-ray image detector can be normalized with respect to the maximum sensitivity (S_0) that would arise if all the incident radiation were absorbed and all the liberated carriers were collected. Neglecting secondary photon interactions, the expression for S_0 is [45.32]

$$S_0 = \left(\frac{5.45 \times 10^{13}\, e}{\left(\frac{\alpha_{air}}{\rho_{air}}\right) W_{\pm}} \right) \left(\frac{\alpha_{en}}{\alpha}\right), \qquad (45.17)$$

where e is the elementary charge, while α_{air} and ρ_{air} are the energy absorption coefficients of air and its density. If $W_{\pm}$ is expressed in eV, α_{air}/ρ_{air} is in cm^2/g and exposure in (45.17) is in R, then S_0 is in units of C/cm^2R. Thus S_0 is a constant that depends on the x-ray photon

Table 45.3 The values of Δ, x_e and x_h for a-Se, poly-HgI_2 and poly-CZT detectors

Photoconductor	$\mu_e\tau_e$ (cm^2/V) $\mu_h\tau_h$ (cm^2/V)	F (V/μm)	E (keV)	L (mm)	x_e	x_h	Δ
Stabilized a-Se	$\mu_e\tau_e = 0.3\times10^{-6}$–$10^{-5}$ $\mu_h\tau_h = 10^{-6}$–6×10^{-5}	≈ 10	20 60	0.2 1.0	1.5–50 0.3–10	5–300 1–60	0.24 0.98
Poly-HgI_2	$\mu_e\tau_e \approx 10^{-6}$–$10^{-5}$ $\mu_h\tau_h \approx 10^{-7}$	0.5–1	20 60	0.15 0.3	0.7–7 0.35–3.5	≈ 0.1 ≈ 0.05	0.21 0.85
Poly-$Cd_{0.95}Zn_{0.05}Te$	$\mu_e\tau_e \approx 2\times10^{-4}$ $\mu_h\tau_h \approx 3\times10^{-6}$	≈ 0.25	20 60	0.3 0.3	≈ 17 ≈ 17	≈ 0.25 ≈ 0.25	0.26 0.89

energy E and the material properties of the photoconductor, since $W_{\pm}$ is a material property that can usually be taken as constant for a given material. For those materials (e.g., a-Se) that have a significant F- and/or E-dependent $W_{\pm}$, then S_0 depends on F and/or E.

The quantity $s = S/S_0$ takes into account the x-ray absorption and charge-transport effects and is called the charge-collection and absorption-limited normalized sensitivity. It should be emphasized that s is a quantity that is determined by the x-ray absorption profile, photoconductor thickness and the charge-collection efficiency. The s of an x-ray detector considering small signal operation, a constant μ and a single deep-trapping time (lifetime) τ for each type of carrier (holes and electrons) and neglecting carrier diffusion is given by [45.82],

$$\begin{aligned}\frac{S}{S_0} &= x_h\left[\left(1-e^{-\frac{1}{\Delta}}\right)+\frac{1}{\frac{\Delta}{x_h}-1}\left(e^{-\frac{1}{x_h}}-e^{-\frac{1}{\Delta}}\right)\right]\\ &\quad + x_e\left[\left(1-e^{-\frac{1}{\Delta}}\right)-\frac{1}{\frac{\Delta}{x_e+1}}\left(1-e^{-\frac{1}{\Delta}-\frac{1}{x_e}}\right)\right]\\ &= s_h(x_h,\Delta)+s_e(x_e,\Delta)=s(x_h,x_e,\Delta)\,,\end{aligned} \tag{45.18}$$

where subscripts h and e refer to holes and electrons respectively.

The two square brackets on the right-hand side of the normalized sensitivity s expression (45.18) represent the relative contributions of hole and electron transport to the overall sensitivity for a given Δ. It is assumed in (45.18) that the radiation-receiving side of the detector is biased positively. If the bias polarity is reversed, then x_e and x_h must be interchanged. The expression in (45.18) applies for incident radiation that is monoenergetic and has to be appropriately integrated over the radiation spectrum of a practical polyenergic x-ray source by considering the x-ray photon-energy-dependent terms $W_{\pm}$, α and α_{en}. Equation (45.18) applies to an isolated photoconductor sandwiched between two large-area parallel-plate electrodes (small pixel effects are excluded) [45.85] and operating under a constant F (small signal case). An excellent fit of (45.18) to experimental data on poly-HgI_2 is given in [45.56].

The value of s is always less than unity since S for a photoconductor of finite thickness in which carrier collection is not perfect is always less than S_0. Note that $s(x_h, x_e, \Delta) = s_h + s_e = 1$ when all the incident radiation is absorbed and all the charges are collected, i. e., $x_h, x_e \gg 1$ and $\Delta \ll 1$. The sensitivity is then simply S_0 and is controlled by $W_{\pm}$.

The sensitivity is mainly controlled by the charges that have the same polarity as the bias on the radiation-receiving electrode: holes for positive bias and electrons for negative bias [45.55, 81]. The extent of the disparity between s_h and s_e depends on Δ. The disparity is stronger for lower Δ, which can be understood by noting that electron and hole generation do not occur uniformly throughout the thickness of the sample but rather it is more effective closer to the radiation-receiving electrode. An advantage of a-Se detectors is that both x_e and x_h are much greater than one, which is not the case for other photoconductors, as shown in Table 45.3. Therefore, charge collection in a-Se detectors is close to unity and the normalized sensitivity is controlled by the quantum efficiency of the detector. The S and s of different photoconductive detectors using the normalized parameters from Table 45.3 are given in Table 45.4. The values of $W_{\pm}$ for a-Se were taken from the work of *Blevis* et al. [45.30]. At $F = 10\,V/\mu m$, $W_{\pm} = 42.5$ and 46 eV for x-ray photon energies of 60 and 20 keV respectively. The maximum x-ray sensitivity S_0 of chest radiographic detectors ($E = 60$ eV) is much higher than of mammographic detectors ($E = 20$ eV) because of the lower α_{air}/ρ_{air} values in (45.17) at higher x-ray photon energies [45.84]. The value of S_0 in a-Se detectors is much lower than in poly-HgI_2 and poly-CZT detectors because of the higher value of $W_{\pm}$ in a-Se.

45.3.2 Detective Quantum Efficiency

DQE measures the ability of the detector to transfer signal relative to noise from its input to its output. The random nature of image quanta gives rise to random

Table 45.4 X-ray sensitivity of a-Se, poly-HgI_2 and poly-CZT detectors using the normalized parameters from Table 45.3 (E is the x-ray photon energy)

Photoconductor	E (keV)	S_0 ($\mu C/cm^2R$)	$s = S/S_0$		$S(\mu C/cm^2R)$	
			Positive bias	Negative bias	Positive bias	Negative bias
Stabilized a-Se	20	0.244	0.9–0.98	0.8–0.98	0.22–0.24	0.2–0.24
	60	5.37	0.39–0.64	0.35–0.62	2.1–3.38	1.88–3.35
Poly-HgI_2	20	2.75	0.25–0.29	0.53–0.81	0.7–0.81	1.46–2.24
	60	38.54	0.15–0.3	0.21–0.4	6.76–11.21	8.18–15.6
Poly-$Cd_{0.95}Zn_{0.05}Te$	20	3	≈ 0.456	≈ 0.85	≈ 1.37	≈ 2.55
	60	35.87	≈ 0.41	≈ 0.51	≈ 14.69	≈ 18.12

fluctuations in image signals, which contributes to image formation and hence creates random noises. The scattering of image quanta gives rise to image blurring, which is quantified by the MTF(f). Images are partially degraded by various sources of statistical fluctuations that arise along the imaging chain. The relative increase in image noise due to an imaging system as a function of spatial frequency f is expressed quantitatively by DQE(f) which represents the signal-to-noise transfer efficiency for different frequencies of information in an image. DQE(f) is defined as

$$\mathrm{DQE}(f) = \frac{\mathrm{SNR}^2_{\mathrm{out}}(f)}{\mathrm{SNR}^2_{\mathrm{in}}(f)}\,, \tag{45.19}$$

where $\mathrm{SNR}_{\mathrm{in}}$ and $\mathrm{SNR}_{\mathrm{out}}$ are the signal-to-noise ratios (SNR) at the input and output stages of the image detector respectively. DQE(f) is unity for an ideal detector. For simplicity, we are often interested in measuring DQE($f = 0$) of an imaging detector since it represents the signal quality degradation due to the signal and noise transfer characteristics of the system without considering signal spreading.

The random nature of charge-carrier trapping (i. e., incomplete charge collection) in the photoconductor layer creates fluctuations in the collected charge and hence creates additional noise. Thus carrier trapping degrades signal-to-noise performance of the image and reduces the DQE. *Kabir* and *Kasap* [45.86] have examined the effects of charge-carrier trapping on the DQE(0) of an a-Se detector by considering depth-dependent conversion gain and depth-dependent charge-collection efficiency in the cascaded-linear-system model [45.87, 88]. Below we apply the DQE(0) model in [45.86] to potential photoconductive detectors such as a-Se, poly-HgI_2 and poly-$Cd_{0.95}Zn_{0.05}Te$ detectors for fluoroscopic applications to study and compare their DQE(0) performance.

Figure 45.13 shows DQE(0) as a function of x-ray exposure for a-Se, HgI_2, and CZT detectors for a 60 keV x-ray beam. The x-ray exposure (X) is varied from 0.1 μR to 10 μR, which is the range of x-ray exposure for fluoroscopic applications. We assume that the pixel area, $A = 200 \times 200\,\mu\mathrm{m}$, and the effective fill factor is 1.0 for all the photoconductors. The average E is 60 keV and the additive electronic noise (N_e) is assumed to be 2000 electrons per pixel. The following transport and operating parameters are used in Fig. 45.13: for a-Se detectors, $L = 1000\,\mu\mathrm{m}$, $F = 10\,\mathrm{V/\mu m}$, $W_{\pm} \approx 43\,\mathrm{eV}$, $\mu_e\tau_e = 10^{-6}\,\mathrm{cm^2/V}$ and $\mu_h\tau_h = 10^{-5}\,\mathrm{cm^2/V}$; for HgI_2 detectors, $L = 260\,\mu\mathrm{m}$, $F = 0.5\,\mathrm{V/\mu m}$, $W_{\pm} = 5\,\mathrm{eV}$, $\mu_e\tau_e = 6\times10^{-6}\,\mathrm{cm^2/V}$ and $\mu_h\tau_h = 10^{-7}\,\mathrm{cm^2/V}$; and for CZT detectors, $L = 270\,\mu\mathrm{m}$, $F = 0.25\,\mathrm{V/\mu m}$, $W_{\pm} = 5\,\mathrm{eV}$,

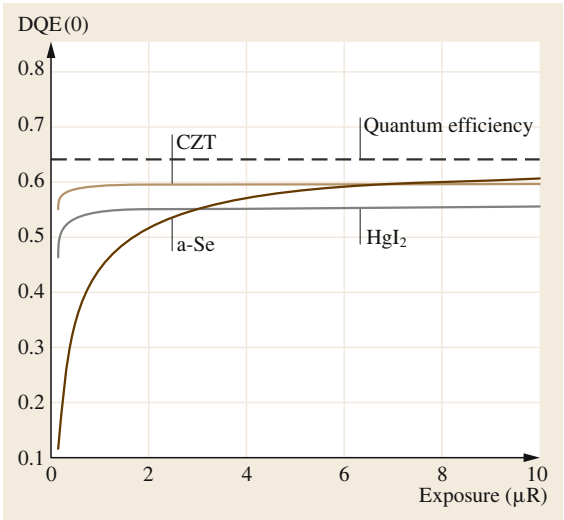


Fig. 45.13 DQE(0) versus x-ray exposure for a-Se, poly-HgI_2, and poly-CZT detectors and for a 60 keV monoenergetic x-ray beam. The electronic noise is 2000 e/pixel. It is assumed that $F = 10\,\mathrm{V/\mu m}$ for a-Se, $0.5\,\mathrm{V/\mu m}$ for HgI_2 and $0.25\,\mathrm{V/\mu m}$ for CZT

$\mu_e\tau_e = 2\times10^{-4}\,\mathrm{cm^2/V}$ and $\mu_h\tau_h = 3\times10^{-6}\,\mathrm{cm^2/V}$. The radiation-receiving electrode is biased positively for a-Se detectors and negatively for HgI_2 and CZT detectors. $\Delta \approx 0.98$ is for all photoconductors. $\eta = 1 - \exp(-1/\Delta) \approx 0.64$ is the maximum achievable DQE(0) if all the liberated charges are collected.

The DQE for the CZT detector is relatively unchanged over the whole exposure range due to a large conversion gain (low $W_{\pm}$) and high charge-collection efficiency (good transport properties). The DQE for a-Se detectors is small at low exposures because of its relatively low conversion gain and is controlled by the added electronic noise. As the x-ray exposure increases, each pixel receives more photons. The relative contribution of electronic noise to the total noise becomes less important and the DQE increases. The DQE for HgI_2 detectors is relatively small even at higher exposures because of its low charge-collection efficiency, which gives rise to considerable gain-fluctuation noise. Therefore, both high conversion gain and high charge-collection efficiency are required to improve the DQE performance of an x-ray image detector [45.89]. The conversion gain depends on $W_{\pm}$, which is a material property of the photoconductor. The charge-collection efficiency can be improved by increasing F and improving the $\mu\tau$ products of the carriers. However, increasing F also increases the dark current dramatically in both HgI_2 and CdZnTe detectors. Thus, there is a practical limitation on F. An F as high as $20\,\mathrm{V/\mu m}$ is achievable in a-Se detectors while keeping the dark current

within an acceptable level for x-ray imaging [45.38]. The charge-collection efficiency of a-Se detectors is relatively high because of the high F that is used (10–20 V/μm) in operating the detector for a reasonable $W_{\pm}$.

45.3.3 Modulation Transfer Function (MTF)

Resolution or resolving power is the ability to record separate images of small objects that are placed very closely together. The overall resolution of a system can be expressed as a convolution of the component resolutions. However, the spatial resolution of an imaging device or a system can also be described in terms of the MTF, which is the relative response of the system as a function of spatial frequency. The MTF of an imaging system can be described as a cascade of several stages where the overall MTF is simply the product of the MTFs of all the individual stages. The MTF(f) is a much more convenient descriptor of spatial response since the resolving power from multiplication is much more easily comprehended than convolution. The overall MTF (or presampling MTF) of an image detector can be expressed as

$$\mathrm{MTF}(f) = \mathrm{MTF_m}(f)\,\mathrm{MTF_a}(f)\,, \qquad (45.20)$$

where $\mathrm{MTF_m}(f)$ is the modulation transfer function of the detector material and $\mathrm{MTF_a}(f)$ is the modulation transfer function associated with the aperture function of the pixel electrodes. $\mathrm{MTF_a}(f)$ arises due to averaging of the signal over a pixel area. If the aperture is square with dimension a, then $\mathrm{MTF_a}(f)$ will be of the form sinc (af). The aperture MTF describes how spatial frequencies are passed through the detector elements.

The spatial resolution in direct-conversion AMFPI is high and closer to the aperture function, as compared to phosphor-based AMFPI. Based on current understanding, charge-carrier trapping and reabsorption of K-fluorescent x-ray photons are the two dominant mechanisms responsible for the loss of resolution [45.89] in direct-conversion AMFPI. Some of the K-fluorescent x-ray photons may be reabsorbed at different points within the detector volume from the primary x-ray-photon interaction point. This creates a lateral spreading of signal and a loss of resolution. The loss of resolution due to fluorescence reabsorption is maximum (although not very substantial [45.89]) just above the K-edge of the photoconductor. This effect can be ignored when: (i) the incident x-ray photon energy is lower than the K-edge of the photoconductor, or (ii) the mean energy of the x-ray beam and the K-edge occur at widely different energies [45.90]. The charge-carrier trapping has a significant effect on the resolution of these direct-conversion x-ray image detectors [45.63], which will be discussed below.

The direct-conversion AMFPI geometry consists of a photoconductor layer sandwiched between two electrodes; the electrode at one side is a continuous metal plate and the electrode on the other side of the photoconductor is segmented into an array of individual square pixels of size $a \times a$, as shown in Fig. 45.14a. The geometric pixel aperture width in a flat-panel detector is smaller than the pixel pitch (center-to-center spacing between two pixels). However, it has been shown that the effective fill factor (the effective fraction of pixel area used for image-charge collection) of a photoconductive flat-panel detector is close to unity [45.91, 92]. Therefore, the effective pixel aperture width a is virtually identical to the pixel pitch. Some of the x-ray-generated carriers are captured by deep traps in the bulk of the photoconductor during their drift across the photoconductor layer. Suppose that a carrier is trapped in the photoconductor above a particular (central) pixel electrode of a pixelated image sensor. This trapped carrier induces charges not only on the central pixel electrode but also on neighboring pixel electrodes, as shown in Fig. 45.14b, and consequently there is a lateral spread of information and hence a loss of image resolution.

An analytical expression for the MTF due to distributed carrier trapping in the bulk of the photoconductor has been developed in [45.63], which examines the effect of charge-carrier trapping on the resolution of direct-conversion AMFPI. The effect of trapping on MTF increases with decreasing normalized carrier lifetime (i. e., normalized schubweg). Trapping of the carriers that move towards the pixel electrodes degrades the MTF performance, whereas trapping of the other type of carriers, which move away from the pixels, improves the sharpness of the x-ray image.

Figure 45.15 shows the experimental MTF and theoretical fit of a 300 μm-thick positively biased CZT detector exposed to an 80 kVp x-ray beam with 26 mm Al filtration. The experimental data have been extracted from [45.93]. The operating $F = 0.25\,\mathrm{V/\mu m}$ and pixel pitch is 150 μm. The Nyquist frequency is 3.3 lp/mm. The dotted line shows the MTF due to the bulk carrier trapping only. The bulk carrier trapping has a significant effect on the MTF of the detector. As is apparent from Fig. 45.15, there is very good agreement between the model and the experimental data. The best-fit $\mu\tau$ products of electrons and holes are $\mu_e\tau_e = 2.4 \times 10^{-4}\,\mathrm{cm^2/V}$ and $\mu_h\tau_h = 3.2 \times 10^{-6}\,\mathrm{cm^2/V}$, which are very close to the $\mu\tau$ values reported previously [45.58, 61]. Although the charge-carrier-trapping-limited MTF model has been applied to the CZT sensors, the model can also be applied to other photoconductive (e.g., a-

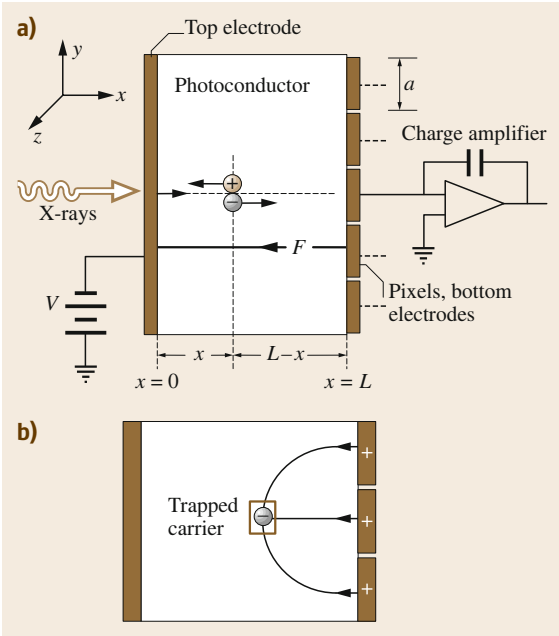


Fig. 45.14 **(a)** A cross section of a direct-conversion pixelated x-ray image detector. **(b)** Trapped carriers in the photoconductor induce charges not only on the central pixel electrode but also on neighboring pixel electrodes, spreading the information and hence reducing spatial resolution

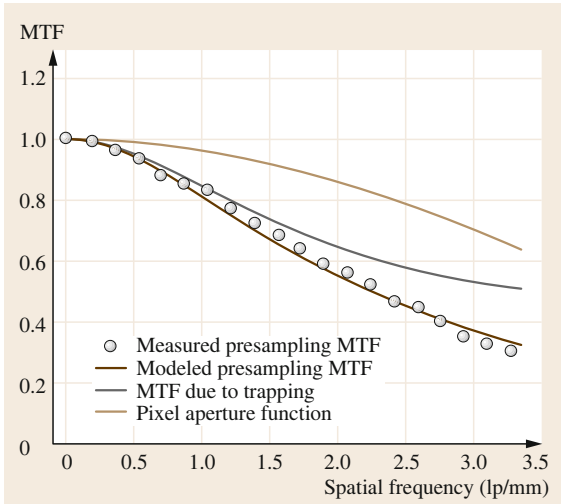


Fig. 45.15 Measured presampling MTF of a polycrystalline CdZnTe detector in comparison with modeled results that included blurring due to charge-carrier trapping in the bulk of the photoconductor. The detector thickness is 300 μm and the pixel pitch is 150 μm. (Experimental data extracted from [45.93, Fig. 12]. Theoretical model from [45.63])

Se, PbO and HgI_2) panel x-ray image detectors. The same model has been also applied to PbO detectors in [45.68].

45.3.4 Image Lag and Ghosting

The detectors should be free of image lag and ghosting effects. When the photoconductor is exposed to x-rays, a transient photocurrent starts to flow that reaches almost a steady value at the carrier transit time [45.94]. Ideally, the current should be present as long as the detector is subjected to x-ray exposure and, once the radiation is stopped, it should diminish to zero. In reality, some of the drifting carriers are trapped in the energy distributed defect states within the bandgap/mobility gap of the photoconductor layer and these trapped carriers are released later. The release time depends on their energy depths from the band/mobility edges [45.95]. As a result, a transient decaying current can be detected for several hundreds/thousands of seconds in some detectors even after the excitation is removed. This is known as the residual current or the lag signal, and has been modeled for a-Se detectors in [45.94]. For most applications, this is an undesirable phenomenon because when part of the current from a previous exposure combines with the next one, the resulting image can be inaccurate. As mentioned in Sect. 45.1.2, the image lag signal was measured in most of the photoconductive detectors. The amount of image lag signal in a-Se detectors has been found to be negligible, whereas it is quite significant in some of other polycrystalline detectors (Sect. 45.1.2).

The x-ray exposure can change the x-ray sensitivity of the exposed area and thus a shadow impression of a previously acquired image is visible in subsequent uniform exposure, which leads to what is called *ghosting* [45.96]. Ghosting can affect the diagnostic value of x-ray images in particular when images are acquired in a fast sequence, for example in fluoroscopy. The causes of change in x-ray sensitivity and resolution, and their recovery mechanisms in a-Se detectors have been systematically investigated in the last decade [45.97, 98]. Though the relative amounts of the sensitivity and resolution changes owing to repeated exposure in a-Se are under an acceptable level, they are due to several mechanisms such as:

1. Recombination of drifting carriers with oppositely charged trapped carriers
2. Creation of x-ray-induced metastable trap centers, and/or
3. Reduction of the free carrier generation due to space charge (i. e., due to a nonuniform electric field) [45.95, 98].

It is expected that the ghosting phenomenon may also be present in other photoconductive (e.g., HgI_2, CZT, PbO and PbI_2) detectors, although it has not yet been systematically measured.

45.4 Summary

The principles of operation of a direct-conversion AMFPI for medical applications have been briefly discussed. The charge-transport and imaging properties of some of the potential photoconductors have been critically discussed and compared with the properties of an ideal photoconductor for x-ray image detectors. The various imaging characteristics of photoconductor-based AMFPIs such as the dark current, sensitivity (S), detective quantum efficiency (DQE), and resolution in terms of the modulation transfer function (MTF) have also been examined. These characteristics depend critically not only on the photoconductor's charge-transport properties but also on the detector structure, i. e., the size of the pixel and the thickness of the photoconductor. It has been shown that the detector structure in terms of the photoconductor thickness and the pixel size is just as important to the overall performance of the detector as the material properties of the photoconductor itself. Long-term stability, and x-ray-induced effects and phenomena should also be examined in the future to obtain more comprehensive detector models.

Acknowledgments. The authors would like to thank Dr. John Rowlands for many fruitful discussions and suggestions during the first edition of this article. The authors acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) through its discovery grants program.

References

45.1 J.A. Rowlands, J. Yorkston: Flat panel detectors for digital radiography. In: *Handbook of Medical Imaging*, Vol. 1, ed. by J. Beutel, H.L. Kundel, R.L. Van Metter (SPIE Press, Washington 2000) p. 223
45.2 K.S. Karim, A. Nathan, J.A. Rowlands, S.O. Kasap: IEE Proc.-CDS **150**, 267 (2003)
45.3 V.N. Cooper III, T. Oshiro, C.H. Cagnon, L.W. Bassett, T.M. McLeod-Stockmann, N.V. Bezrukiy: Med. Phys. **30**, 2614 (2003)
45.4 D.C. Hunt, O. Tousignant, J.A. Rowlands: Med. Phys. **31**, 1166 (2004)
45.5 L.E. Antonuk, J.M. Boudry, Y. El-Mohri, W. Huang, J.H. Siewerdsen, J. Yorkston, R.A. Street: Proc. SPIE **2432**, 216 (1995)
45.6 J. Chabbal, C. Chaussat, T. Ducourant, L. Fritsch, J. Michaïlos, V. Spinnler, G. Vieux, M. Arques, G. Halm, M. Hoheisel, H. Horbaschek, R. Schulz, M. Spahn: Proc. SPIE **2708**, 499 (1996)
45.7 W. Zhao, J.A. Rowlands: Med. Phys. **22**, 1595 (1995)
45.8 W. Zhao, I. Blevis, S. Germann, J.A. Rowlands, D. Waechter, Z. Huang: Med. Phys. **24**, 1834 (1997)
45.9 W. Zhao, J.A. Rowlands: Med. Phys. **24**, 1819 (1997)
45.10 J.A. Rowlands, S.O. Kasap: Phys. Today **50**, 24 (1997)
45.11 J.A. Rowlands, W. Zhao, I.M. Blevis, G. Pang, W.G. Ji, S. Germann, S.O. Kasap, D.F. Waechter, Z.S. Huang: Proc. SPIE **3032**, 97 (1997)
45.12 S.O. Kasap, J.A. Rowlands, B. Fogal, M.Z. Kabir, G. Belev, N. Sidhu, B. Polischuk, R.E. Johanson: J. Non-Cryst. Solids **299**, 988 (2002)
45.13 W. Zhao, W.G. Ji, A. Debrie, J.A. Rowlands: Med. Phys. **30**, 254 (2003)
45.14 D.L. Lee, L.K. Cheung, L.S. Jeromin, E.F. Palecki, B.G. Rodricks: Proc. SPIE **3032**, 88 (1997)
45.15 D.L. Lee, L.K. Cheung, B.G. Rodricks, G.F. Powell: Proc. SPIE **3336**, 14 (1998)
45.16 O. Tousignant, M. Choquette, Y. Demers, L. Laperrière, J. Leboeuf, M. Honda, M. Nishiki, A. Takahashi, A. Tsukamoto: Proc. SPIE **4682**, 503 (2002)
45.17 S.O. Kasap, J.A. Rowlands: J. Mater. Sci. Mater. Electron. **11**, 179 (2000)
45.18 M.Z. Kabir, E.V. Emelianova, V.I. Arkhipov, M. Yunus, G. Adriaenssens, S.O. Kasap: J. Appl. Phys. **99**, 124501 (2006)
45.19 S.O. Kasap, J.A. Rowlands: IEE Proc.-CDS **149**, 85 (2002)
45.20 S.O. Kasap: Photoreceptors: The chacogenides. In: *Handbook of Imaging Materials*, 2nd edn., ed. by A.S. Diamond, D.S. Weiss (Marcel Dekker, New York 2001) p. 329
45.21 S.O. Kasap, J.A. Rowlands: Proc. IEEE **90**, 591 (2002)
45.22 F. Greuter, G. Blatter: Semicond. Sci. Technol. **5**, 111 (1990)
45.23 M.H. Chowdhury, M.Z. Kabir: J. Phys. D **44**, 015102 (2011)
45.24 A. Zuck, M. Schieber, O. Khakhan, Z. Burshtein: IEEE Trans. Nucl. Sci. **50**, 991 (2003)
45.25 S. Tokuda, H. Kishihara, S. Adachi, T. Sato: Proc. SPIE **5030**, 861 (2003)
45.26 R.A. Street, S.E. Ready, K. Van Schuylenbergh, J. Ho, J.B. Boyec, P. Nylen, K. Shah, L. Melekhov, H. Hermon: J. Appl. Phys. **91**, 3345 (2002)
45.27 S. Tokuda, H. Kishihara, S. Adachi, T. Sato: J. Mater. Sci. Mater. Electron. **15**, 1 (2004)
45.28 M. Simon, R.A. Ford, A.R. Franklin, S.P. Grabowski, B. Mensor, G. Much, A. Nascetti, M. Overdick, M.J. Powell, D.U. Wiechert: IEEE Trans. Nucl. Sci. **52**,

188 (2005)
45.29 S.O. Kasap, K.V. Koughia, B. Fogal, G. Belev, R.E. Johanson: Semicond **14**, 816 (2003)
45.30 I.M. Blevis, D.C. Hunt, J.A. Rowlands: J. Appl. Phys. **85**, 7958 (1999)
45.31 M.F. Stone, W. Zhao, B.V. Jacak, P. O'Conner, B. Yu, P. Rehak: Med. Phys. **29**, 319 (2002)
45.32 S.O. Kasap: J. Phys. D **33**, 2853 (2000)
45.33 W. Que, J.A. Rowlands: Phys. Rev. B **51**, 10500 (1995)
45.34 E. Fourkal, M. Lachaine, B.G. Fallone: Phys. Rev. B **63**, 195204 (2001)
45.35 D. Mah, J.A. Rowlands, J.A. Rawlinson: Med. Phys. **25**, 444 (1998)
45.36 J.B. Frey, G. Belev, O. Tousignant, H. Mani, L. Laperriere, S.O. Kasap: J. Appl. Phys. **112**, 014502 (2012)
45.37 S.A. Mahmood, M.Z. Kabir, O. Tousignant, H. Mani, J. Greenspan, P. Botka: Appl. Phys. Lett. **92**, 223506 (2008)
45.38 B. Polischuk, Z. Shukri, A. Legros, H. Rougeot: Proc. SPIE **3336**, 494 (1998)
45.39 S.O. Kasap, J.B. Frey, G. Belev, O. Tousignant, H. Mani, L. Laperriere, A. Reznik, J.A. Rowlands: Phys. Status Solidi (b) **246**, 1794 (2009)
45.40 M. Choquette, Y. Demers, Z. Shukri, O. Tousignant, K. Aoki, M. Honda, A. Takahashi, A. Tsukamoto: Proc. SPIE **4320**, 501 (2001)
45.41 A. Reznik, S.D. Baranovskii, O. Rubel, G. Juska, S.O. Kasap, Y. Ohkawa, K. Tanioka, J.A. Rowlands: J. Appl. Phys. **102**, 053711 (2007)
45.42 K. Kikuchi, Y. Ohkawa, K. Miyakawa, T. Matsubara, K. Tanioka, M. Kubota, N. Egami: Phys. Status Solidi (c) **8**, 2800 (2011)
45.43 S.O. Kasap, J.B. Frey, G. Belev, O. Tousignant, H. Mani, J. Greenspan, L. Laperriere, O. Bubon, A. Reznik, G. DeCrescenzo, K.S. Karim, J.A. Rowlands: Sensors **11**, 5112 (2011)
45.44 A. Reznik, S.D. Baranovskii, O. Rubel, K. Jandieri, J.A. Rowlands: Phys. Status Solidi (c) **8**, 790 (2011)
45.45 S. Abbaszadeh, A. Tari, W.S. Wong, K.S. Karim: IEEE Trans. Electron. Devices **61**, 3355 (2014)
45.46 S. Abbaszadeh, N. Allec, S. Ghanbarzadeh, U. Shafique, K.S. Karim: IEEE Trans. Electron. Devices **59**, 2403 (2012)
45.47 O. Bubon, G. DeCrescenzo, J.A. Raowlands, A. Reznik: J. Non-Cryst. Solids **358**, 2431 (2012)
45.48 M.Z. Kabir: J. Mater. Sci. Mater. Electron. **26**, 4659 (2015)
45.49 G. Juska, K. Arlauskas: Phys. Status Solidi (a) **59**, 389 (1980)
45.50 N. Hijazi, M.Z. Kabir: Can. J. Phys. **93**, 1407 (2015)
45.51 Q. Zhao, L.E. Antonuk, Y. El-Mohri, H. Du, A. Sawant, Z. Su, J. Yamamoto: Med. Phys. **37**, 2738 (2010)
45.52 H. Jiang, Q. Zhao, L.E. Antonuk, Y. El-Mohri, T. Gupta: Phys. Med. Biol. **58**, 703 (2013)
45.53 H. Du, L.E. Antonuk, Y. El-Mohri, Q. Zhao, Z. Su, J. Yamamoto, Y. Wang: Phys. Med. Biol. **53**, 1325 (2008)
45.54 M. Schieber, A. Zuck, M. Braiman, J. Nissenbaum, R. Turchetta, W. Dulinski, D. Husson, J.L. Riester: IEEE Trans. Nucl. Sci. **NS-44**, 2571 (1997)
45.55 Z. Su, L.E. Antonuk, Y. El-Mohri, L. Hu, H. Du, A. Sawant, Y. Li, Y. Wang, J. Yamamoto, Q. Zhao: Phys. Med. Biol. **50**, 2907 (2005)
45.56 M.Z. Kabir, S.O. Kasap: Appl. Phys. Lett. **80**, 1664 (2002)
45.57 R.C. Whited, L. Van den Berg: IEEE Trans. Nucl. Sci. **NS-24**, 165 (1977)
45.58 S.K. Chaudhuri, K. Nguyen, R.O. Pak, L. Matei, V. Buliga, M. Groza, A. Burger, K.C. Mandal: IEEE Trans. Nucl. Sci. **NS-61**, 793 (2014)
45.59 Y. Eisen, A. Shor: Mater. Res. Soc. Symp. Proc. **487**, 129 (1997)
45.60 T.L. Chu, S.S. Chu, C. Ferekides, C.Q. Wu, J. Britt, C. Wang: J. Appl. Phys. **70**, 7608 (1991)
45.61 A. Ruzin, Y. Nemirovsky: J. Appl. Phys. **82**, 4166 (1997)
45.62 J.G. Mainprize, N.L. Ford, S. Yin, E.E. Gordon, W.J. Hamilton, T.O. Tümer, M.J. Yaffe: Med. Phys. **29**, 2767 (2002)
45.63 M.Z. Kabir, S.O. Kasap: J. Phys. D **36**, 2352 (2003)
45.64 G. Zentai, L. Partain, R. Pavlyuchkova, C. Proano, G. Virshup, L. Melekhov, A. Zuck, B.N. Breen, O. Dagan, A. Vilensky, M. Schieber, H. Gilboa, P. Bennet, K. Shah, Y. Dmitriev, J. Thomas: Proc. SPIE **5030**, 77 (2003)
45.65 X. Zhu, H. Sun, D. Yang, J. Yang, X. Li, X. Gao: J. Mater. Sci. Mater. Electron. **25**, 3337 (2014)
45.66 R.A. Street, S.E. Ready, F. Lemmi, K.S. Shah, P. Bennett, Y. Dmitriyev: J. Appl. Phys. **86**, 2660 (1999)
45.67 A. Brauers, U. Schiebel: X-ray image sensor, US Patent (Application) 5729021 (1998)
45.68 M.Z. Kabir: J. Appl. Phys. **104**, 074506 (2008)
45.69 D.R. Ouimette, S. Nudelman, R. Aikens: Proc. SPIE **3336**, 470 (1998)
45.70 P.R. Bennett, K.S. Shah, L.J. Cirignano, M.B. Klugerman, L.P. Moy, F. Olschner, M.R. Squillante: IEEE Trans. Nucl. Sci. **46**, 266 (1999)
45.71 S. Yakunin, M. Sytnyk, D. Kriegner, S. Shrestha, M. Richter, G. Matt, H. Azimi, C. Brabec, J. Stangl, M. Kovalenko, W. Heiss: Nat. Photonics **9**, 444 (2015)
45.72 T. Baikie, Y. Fang, J.M. Kadro, M. Schreyer, F. Wei, S.G. Mhaisalkar, M. Graetzel, T.J. White: J. Mater. Chem. A **1**, 5628 (2013)
45.73 G. Giorgi, J.I. Fujisawa, H. Segawa, K. Yamashita: J. Phys. Chem. Lett. **4**, 4213 (2013)
45.74 D.C. Hunt, S.S. Kirby, J.A. Rowlands: Med. Phys. **29**, 2464 (2002)
45.75 S.A. Mahmood, M.Z. Kabir: Curr. Appl. Phys. **9**, 1393 (2009)
45.76 Z. He: Nucl. Instr. Meth. A **463**, 250 (2001)
45.77 R.A. Street: Appl. Phys. Lett. **57**, 1334 (1990)
45.78 M.Z. Kabir: Nucl. Instr. Meth. A **736**, 156 (2014)
45.79 S.A. Mahmood, M.Z. Kabir: J. Vac. Sci. Tech. A **29**, 031603 (2011)
45.80 G. Vincent, A. Chantre, D. Bois: J. Appl. Phys. **50**, 5484 (1978)
45.81 X. Zhu, H. Sun, D. Yang, J. Yang, X. Li, X. Gao: J. Mater Sci. Mater. Electron. **25**, 3337 (2014)
45.82 M.Z. Kabir, S.O. Kasap: J. Vac. Sci. Tech. A **20**, 1082 (2002)
45.83 E. Rosencher, V. Mosser, G. Vincent: Phys. Rev. B **29**, 1135 (1984)

45.84 J.M. Boone: X-ray production, interaction, and detection in diagnostic imaging. In: *Handbook of Medical Imaging*, Vol. 1, ed. by J. Beutel, H.L. Kundel, R.L. Van Metter (SPIE Press, Washington 2000) p. 1
45.85 M.Z. Kabir, S.O. Kasap: J. Vac. Sci. Tech. A **22**, 975 (2004)
45.86 M.Z. Kabir, S.O. Kasap: J. Phys. D **35**, 2735 (2002)
45.87 M. Rabbani, R. Shaw, R.L. Van Metter: J. Opt. Soc. Am. A **4**, 895 (1987)
45.88 J.G. Mainprize, D.C. Hunt, M.J. Yaffe: Med. Phys. **29**, 976 (2002)
45.89 M.Z. Kabir, S.O. Kasap, W. Zhao, J.A. Rowlands: IEE Proc.-CDS **150**, 258 (2003)
45.90 W. Que, J.A. Rowlands: Med. Phys. **22**, 365 (1995)
45.91 D.C. Hunt, O. Tousignant, Y. Demers, L. Laperrière, J.A. Rowlands: Proc. SPIE **5030**, 226 (2003)
45.92 G. Pang, W. Zhao, J.A. Rowlands: Med. Phys. **25**, 1636 (1998)
45.93 S. Tokuda, H. Kishihara, S. Adachi, T. Sato: J. Mater. Sci. Mater. Electron. **15**, 1 (2003)
45.94 S. Siddiquee, M.Z. Kabir: J. Vac. Sci. Tech. A **33**, 041514 (2015)
45.95 S.A. Mahmood, M.Z. Kabir, O. Tousignant, J. Greenspan: IEEE Trans. Nucl. Sci. **59**, 597 (2012)
45.96 V. Loustauneau, M. Bissonnette, S. Cadieux, M. Hansroul, E. Masson, S. Savard, B. Polischuk: Proc. SPIE **5368**, 162 (2004)
45.97 M.Z. Kabir, M. Yunus, S.O. Kasap: Proc. SPIE **5368**, 170 (2004)
45.98 M.Z. Kabir, L. Chowdhury, G. DeCrescenzo, O. Tousignant, S.O. Kasap, J.A. Rowlands: Med. Phys. **37**, 1339 (2010)

46. Phase-Change Memory Materials

Alexander V. Kolobov, Junji Tominaga, Paul Fons

Phase-change materials are Te-containing alloys, typically lying along the $GeTe$-Sb_2Te_3 quasibinary tie line. Their ability to switch, reversibly and extremely quickly, between the crystalline and amorphous phases, combined with the high stability of both phases, makes them ideally suitable for memory applications. They have been long used in optical data storage in the form of DVD and Blu-Ray disks and recently have also emerged as a leading candidate for electronic nonvolatile memory devices. In this chapter, a detailed description of these materials is provided starting with the global and local structures of the two phases, which were extensively studied both experimentally and using ab initio computer simulations, and followed by the discussion of possible atomistic mechanisms of the phase-change process, with special accent on the role of electronic excitation. The chapter is concluded by a brief description of the present and emerging applications of this class of chalcogenide materials.

46.1 **Structure of Ge-Sb-Te Phase-Change Alloys** 1150
46.1.1 Crystalline Phase 1150
46.1.2 Amorphous Phase 1152
46.2 **Mechanism of the Phase-Change Process** 1156
46.2.1 Structural Studies 1156
46.2.2 Photo-Assisted Amorphization of $Ge_2Sb_2Te_5$ 1156
46.3 **Present Applications and Future Trends** 1158
References 1160

Nonvolatile memory devices are key elements of a wide variety of electronic and portable systems such as digital cameras, solid-state disks, smartphones, computers, e-books, tablets, etc., and their market has been increasing exponentially over the last decade. Even though flash memory represents today the leading technology, to allow its scalability down to the 16 nm technology node and beyond, new architectures are necessary. Therefore, new emerging nonvolatile memory concepts are under investigation and one of the leading candidates is phase-change memory (PCM). PCM has been successfully used in optical memory devices such as DVD-RAM since the 1990s and recently commercial production of electronic nonvolatile phase-change random access memory (PRAM) has been launched by two of the world's leading memory makers Samsung and Micron. PCM is considered to be a storage class memory in which unlike flash its elements are individually addressable, have a demonstrated cyclability of over 10^{11} cycles, and have access times on the tens to hundreds of nanoseconds scale.

The basic idea of PCM of utilizing the property contrast between the amorphous and crystalline phases (the SET and RESET states) of some materials belongs to *Ovshinsky* and dates back to the 1960s [46.1] and is based on the following. When a liquid is cooled, it can solidify either discontinuously to form a crystal or in a continuous way to form a glass. The process is schematically illustrated in Fig. 46.1.

A very simplified picture of the phase-change process is based on the idea that as the temperature of the melt decreases, the viscosity becomes larger, and at a certain point the structure can no longer follow the change in temperature. As a result, if the melt is cooled slowly, the equilibrium crystalline phase is formed. If, on the other hand, the melt is cooled fast enough so that the structure cannot follow the change in temperature, the disordered phase is frozen; one obtains a supercooled liquid and subsequently a glass. When the obtained glass is annealed at temperatures between the glass-transition temperature and the melting point, it gradually transforms into the crystalline

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_46

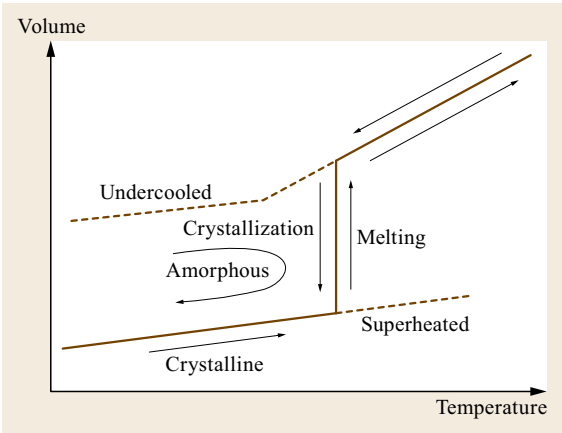


Fig. 46.1 A volume–temperature phase diagram demonstrating the formation of either a crystalline or amorphous phase from the melt

state. One can thus cycle a material between the ordered crystalline and disordered amorphous phases. In the device-relevant phase-change process, a short intense laser or current pulse melts the material that is subsequently quenched into the amorphous phase. A longer pulse of lower intensity reverts the material to the crystalline phase.

The research and development of phase-change optical storage has a long history starting with the discovery of the switching effect in some chalcogenide alloys [46.1–3]. In particular, in amorphous $Si_{12}Te_{48}As_{30}Ge_{10}$ it was found that when a voltage that exceeded a critical value, called the threshold voltage, was applied, the chalcogenide alloy switched from the low-conductivity state to the high-conductivity state [46.1]. As the current in the high-conductivity state was reduced below the characteristic value termed the holding current, the system switched back to the original low-conductivity state. The observed switching process was reversible and could be repeated many times. The observed increase in the conductivity was attributed to the formation of a current filament, growing in diameter with increasing current flow. In the same paper [46.1], it was reported that by tuning the composition a memory effect was observed that was attributed to a change in the structure of the material, namely:

> *[. . .] after switching from a highly resistive state, structural changes result in the preservation of a conductive state even when the current is totally removed. The material can be reversibly switched back to the highly resistive state by application of a current pulse of either polarity exceeding a threshold value.*

This finding marked the beginning of intensive – and extensive – research on phase-change materials. In order for a material to become commercially relevant, it has to satisfy simultaneously several requirements such as sufficiently large property contrast between the two states and high stability of both phases at operating temperatures alongside high switching speed in both directions. Less obvious but equally important are the need for low thermal conductivity of the crystalline phase (to ensure low-energy switching), the relative softness of the material needed to withstand stresses generated at the amorphous-crystalline boundaries allowing for high cyclability, and good scalability. While there are many materials that satisfy some of these requirements, very few satisfy them all.

Years of research have singled out GeTe-based PCM alloys. In practical applications, GeTe is usually alloyed with Sb or other additives such as C or N to tune desirable properties, such as thermal stability, switching speed, or optical contrast (e.g., $Ge_2Sb_2Te_5$ is used in DVDs but $Ge_8Sb_2Te_{11}$ is used in Blu-ray disks). What makes the GeTe alloys special? Since the properties of a material are determined by its structure, we start by discussing the structure of GeTe-based alloys.

46.1 Structure of Ge-Sb-Te Phase-Change Alloys

46.1.1 Crystalline Phase

Binary GeTe

It is natural to start the description with the binary compound GeTe, which is the simplest material in the Ge-Sb-Te (GST) system. In the low-temperature ferroelectric phase, GeTe possesses a rhombohedral structure with the space group $R3m$. This structure can be viewed as a rock-salt structure slightly distorted along the ⟨111⟩ direction with a subsequent shear relaxation. The driving force for the formation of the rhombohedral phase has been a subject of several studies in the past [46.4]. In this phase, Ge and Te atoms are sixfold coordinated to each other with subsets of three shorter (2.83 Å) and three longer (3.15 Å) bonds often described as a Peierls distortion [46.5] due to the reduced coupling between the orthogonal p-type orbitals that constitute the basis for bonding in GeTe.

Based on diffraction studies, mainly neutron diffraction [46.6], it was concluded from a Bragg peaks analysis that GeTe undergoes a displacive ferroelectric-to-paraelectric transition at the Curie temperature, T_c,

around 705 K whereupon the structure changes to the rock-salt structure (space group ($Fm\bar{3}m$) with a concomitant disappearance of the Peierls distortion). Subsequent use of local probes such as extended x-ray absorption fine structure (EXAFS) or pair-distribution function analysis of total scattering unambiguously demonstrated that the structure *locally* remains distorted above T_c in essentially the same manner as it is distorted at lower temperatures but the distribution of shorter and longer bonds becomes random, which manifests itself as a rock-salt average structure from the perspective on long-range averaging probes such as Bragg diffraction [46.7, 8].

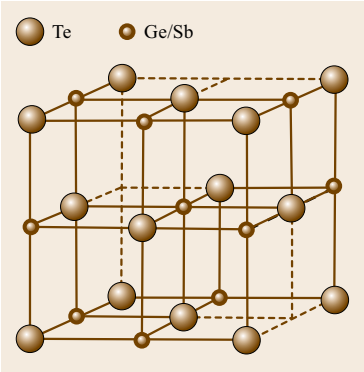


Fig. 46.2 Metastable cubic structure of $Ge_2Sb_2Te_5$

$Ge_2Sb_2Te_5$ and Related Materials

The first report on the structure of Ge-Sb-Te dates back to the late 1960s when *Petrov* et al. [46.9] investigated the structures of $Ge_2Sb_2Te_5$ and $GeSb_4Te_7$. Analysis of electron-diffraction results led the authors to conclude that the studied crystals possessed a hexagonal cell with space group $P\bar{3}m1$ and the lattice constants $a = 4.20 \pm 0.02$ and $c = 16.96 \pm 0.06$ Å and $a = 4.21 \pm 0.02$ and $c = 23.65 \pm 0.08$ Å for $Ge_2Sb_2Te_5$ and $GeSb_4Te_7$, respectively. The structure of $Ge_2Sb_2Te_5$ was described as a layered structure with the stacking sequence along the *c*-axis of Te-Sb-Te-Ge-Te-Te-Ge-Te-Sb-Te. This sequence is sometimes referred to as the Petrov sequence. The stable phase of $Ge_2Sb_2Te_5$ was subsequently reexamined by *Kooi* et al. [46.10]. Very similar lattice parameters were obtained ($a = 4.2$ Å and $c = 17.2$ Å) but the stacking sequence was found to be different, namely, Te-Ge-Te-Sb-Te-Te-Sb-Te-Ge-Te, that is, with the Ge and Sb positions exchanged. A later study using synchrotron radiation found that the Ge/Sb layers were randomly occupied by these two elements [46.11].

Yamada [46.12] was the first to study the structure of a thin layer, crystallized from the amorphous phase. It was found that it possesses a structure that was different from that of the stable trigonal phase. This issue was further explored using x-ray diffraction (XRD) by *Nonaka* et al. [46.13] and by *Yamada* and *Matsunaga* [46.14].

Based on these results, the crystal structure of $Ge_2Sb_2Te_5$ was identified as the rock-salt (NaCl) structure ($Fm\bar{3}m$). From Rietveld refinement, it was concluded that the anion face-centered cubic (fcc) sublattice was fully occupied by Te atoms with Ge and Sb atoms randomly located on the cation sublattice. The stoichiometry of the structure requires that there are vacancies on the Ge/Sb sites and their presence has been confirmed by structure refinement. The proposed structure is shown in Fig. 46.2. It was subsequently shown that the rock-salt structure is characteristic of a wide range of GST alloys [46.15–18].

EXAFS studies performed by various groups [46.19, 20] demonstrated that the Ge-Te bond length in metastable $Ge_2Sb_2Te_5$ is 2.83 ± 0.01 Å, that is, significantly shorter than might be expected based on the rock-salt structure as determined by XRD and the experimental lattice parameter of slightly over 6.00 Å. Based on these results, it was argued [46.20] that the structure of metastable $Ge_2Sb_2Te_5$ does not possess rock-salt symmetry but is locally distorted. The local structure around the Ge species is very similar to that of the binary GeTe, which possesses a rhombohedral structure with subsets of three shorter and three longer bonds. The distortion of the Ge atom location was subsequently confirmed experimentally using scattering measurements [46.21, 22]. The Sb-Te bond length was also found to be shorter than half the lattice parameter (2.94 Å [46.20]). The distortion present in GST alloys is usually referred to as Peierls distortion [46.5].

While the obtained bond lengths are significantly shorter than those expected from the experimental lattice parameters and the rock-salt symmetry, they are at the same time significantly longer that the sum of the corresponding covalent radii, which suggests that the bonds are not purely covalent.

The rock-salt like arrangement of atoms in the crystalline phase requires the formation of six bonds by each participating atom while at the same time, the number of valence electrons located on Ge and Sb atoms is lower. A way to ensure six bonds is by virtue of sharing the valence electrons among several bonds with less than two electrons per bond on average through resonance bonding, a concept that is described as follows. *Lucovsky* and *White* [46.23] were the first to discuss this possibility and its consequences for IV-VI crystals such as GeTe.

In particular, in the cited work [46.23] resonance bonding in IV-VI and chalcogen (Se and Te) crystals was considered and it was suggested that *we might expect materials that exhibit resonance bonding in the*

crystalline phase to have different properties in the amorphous phase. It was also stated that long-range order is crucial for resonance bonding to exist. If long-range order is lost, the system will be unable to achieve resonant bonding [46.23]. This approach has been subsequently elaborated in [46.24–26] to describe the unusually large property contrast in GST alloys and the fast switching rate.

How does the bond length asymmetry effect the electron density distribution? A plot of the charge density difference (CDD) between a simulated relaxed GeTe model at 0 K and isolated pseudo-atoms shown in Fig. 46.3 demonstrates that the electron density is only significant along the short bonds. In contrast, the electron density pile-up along the long bonds is *significantly* lower demonstrating that there is a pronounced bonding energy hierarchy between short and long bonds [46.27]. The existing bonding energy hierarchy should manifest itself in the different response of the short and long bonds to thermal and/or electronic excitations. The fact that long-range order is associated with weak secondary bonds suggests that the crystalline phase is intrinsically fragile.

Vacancies present in the structure deserve a dedicated paragraph. *Kolobov* et al. [46.20] suggested that vacancies serve to stabilize the structure by ensuring the appropriate charge balance. The role of vacancies was a subject of several consequent studies, where it was also concluded that vacancies, whose concentration varies as $x/(1+2x)$ for the $GeTe_{1-x}$-$(Sb_2Te_3)_x$ system [46.15], are not electronic defects but are an intrinsic feature of the GST rock-salt structure [46.15, 24, 28]. Vacancies were also argued to account for the p-type conductivity of the crystalline phase of GeTe [46.29]. *Wuttig* et al. [46.28] demonstrated that introduction of vacancies into Ge/Sb sublattice decreased the total energy of the system.

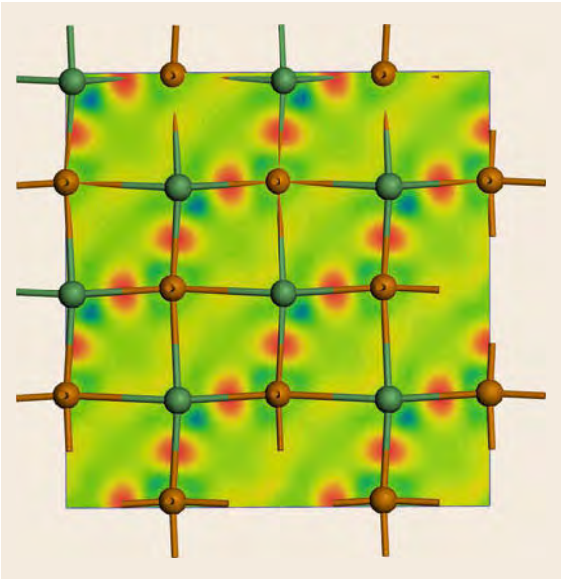


Fig. 46.3 Difference in electron charge density for GeTe at 0 K and isolated pseudo-atoms. A significant bonding charge can be seen midway along the shorter bonds which corresponds to covalent bonding. At the same time, the bonding localized along the longer bonds is *significantly* lower clear demonstrating a pronounced bonding energy difference between the short and long bonds. (Reprinted from [46.27] by permission from Macmillan Publishers Ltd., copyright (2011))

46.1.2 Amorphous Phase

Experimental Studies

Of special interest may be the structure of amorphous binary GeTe (a-GeTe) that is the simplest phase-change material and an end point of the quasibinary GeTe-Sb_2Te_3 tie-line. Despite years of studies by various groups, there is still controversy regarding the structure of amorphous GeTe. X-ray diffraction structural studies made by *Betts* et al. [46.30] back in the 1970s revealed that the interatomic distance and the coordination number of a-GeTe are in poor agreement with those of c-GeTe. From a detailed analysis of the diffraction data, these authors concluded that a random covalent model with a 4(Ge):2(Te) local coordination was the most appropriate as a local coordination model for a-GeTe. Subsequent electron diffraction studies [46.31, 32] obtained similar radial distribution functions. The 4(Ge):2(Te)-coordinated structure was also suggested on the basis of EXAFS measurements around both Ge and Te *K*-edges [46.33–35]. The Ge-Te bond length (2.59 Å) was also found to be consistent with the 4(Ge):2(Te) bonding geometry [46.35]. Raman scattering and far-infrared absorption spectra for a-GeTe [46.36] were equally interpreted in terms of the presence of $GeTe_4$ tetrahedra. These results strongly supported a random covalent network model of the 4(Ge):2(Te)-coordinated atoms, and excluded the possibility of a c-GeTe microcrystalline structure. Combined photoemission and inverse photoemission studies [46.37] also favored the 4(Ge):2(Te) coordination.

On the other hand, neutron scattering [46.38] and Mossbauer spectrometry of ^{125}Te nuclei [46.39] studies suggested a 3(Ge):3(Te)-coordinated local structure, demonstrating that the local structure of a GeTe is still far from being resolved.

The first report on the local structure of the amorphous phase of GST alloys belongs to the LETI group in France [46.19]. These authors performed EXAFS studies on as-deposited and thermally crystallized layers of

GeTe and GST and obtained the bond lengths characteristic of the two states. In particular, they found the Ge-Te bond to be 2.62 Å and the Sb-Te bond to be 2.84 Å (as compared to 2.83 and 2.94 Å, respectively in the crystalline phase), that is, the bonds are *shorter* in the amorphous phase. This work, however, only reported the obtained numerical values and stayed short of drawing further conclusions.

Kolobov et al. [46.20] were the first to go beyond simple measurements and proposed structural models. Their results are described as follows in more detail. It should be noted that in this work laser-crystallized and laser-amorphized states of GST were studied, that is, exactly the structures that are used in optical memory devices (the measurements were performed on real-device structures).

For the amorphous phase, it was found that Ge-Te and Sb-Te bonds become shorter and the structure possessed more local order than the crystalline phase as evidenced by the more intense and narrower peaks in the spectra corresponding to the amorphous phase (Fig. 46.4). It is informative to note that the measured Ge-Te and Sb-Te bond lengths are very close to the sum of the corresponding covalent radii for the elements ($r_{Ge} = 1.21$ Å, $r_{Sb} = 1.40$ Å, $r_{Te} = 1.36$ Å [46.40]). (We note here that the covalent radius for Ge is given assuming sp^3-hybridization).

The observed bond shortening and increased local order are highly unusual for covalent solids when, due to anharmonicity of the interatomic potential, disorder usually results in longer and weaker bonds and suggests that the local structures in the two cases are significantly different. It should also be noted that despite the bond shortening, the density of the amorphous phase is about 5% lower than that of the crystalline phase.

The conclusion that a significant change occurs in the local structure finds more support in the experimentally measured x-ray absorption near-edge structure (XANES) spectra [46.20], where the overall observed changes in XANES upon the phase transition could be reproduced by first-principles multiple-scattering (FEFF) simulations [46.41] when the Ge atoms were placed into tetrahedral symmetry sites [46.20].

Amorphous $Ge_2Sb_2Te_5$ has subsequently been studied using EXAFS by other groups [46.42–44] and similar results were obtained although in addition the presence of Ge-Ge bonds was also reported. It is interesting to note that the obtained values of the fraction of Ge-Ge bonds were similar to the concentration of tetrahedral Ge sites, suggesting a correlation between the two.

Bond shortening in the as-deposited amorphous phase of $Ge_2Sb_2Te_5$ has also been observed by high-energy x-ray scattering experiments [46.45].

A gradual upwards drift in the electrical resistance of Ge-Sb-Te alloys has been observed after switching to the amorphous state that potentially can cause problems for multilevel storage applications in PRAM. This phenomena, often referred to as *drift* of the amorphous state, is usually attributed to the gradual [46.46, 47] relaxation of bonds toward a lower energy state. In particular, the loss of Ge-Ge bonds in the amorphous state leads to a widening of the bandgap and a gradual increase in resistivity.

Computer Simulations

Welnic et al. [46.48] were the first to apply ab-initio simulations based on density functional theory for phase-change materials, using the spinel structure (in which Ge atoms are tetrahedrally coordinated) to model

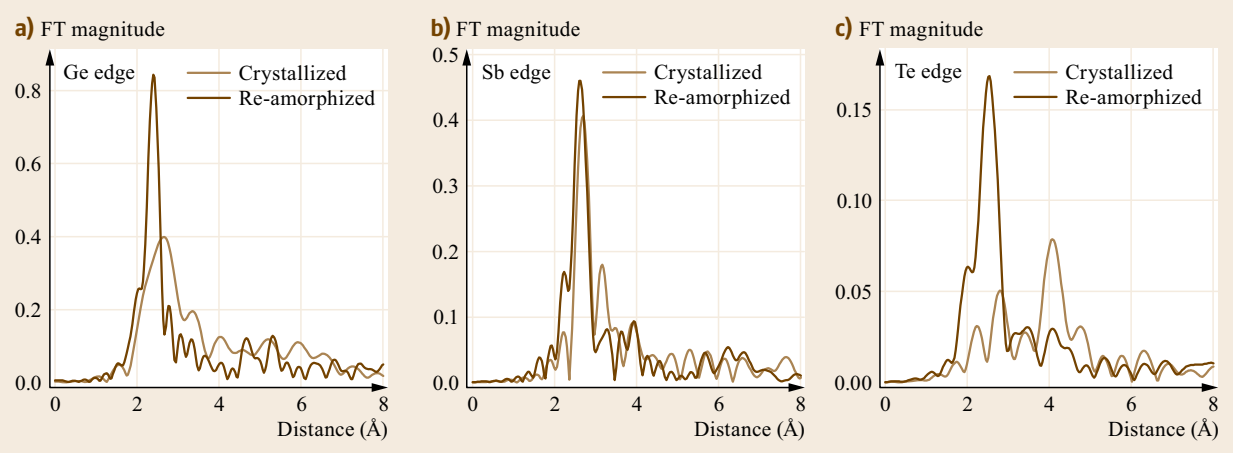


Fig. 46.4a–c Fourier transformed EXAFS spectra for Ge **(a)**, Sb **(b)**, and Te **(c)** K-edges for laser crystallized and laser amorphized $Ge_2Sb_2Te_5$. (Reprinted from [46.20] by permission from Macmillan Publishers Ltd, copyright (2004))

the amorphous phase. Calculations of the band structure showed a band opening at the Γ point upon transformation from the rock-salt to the spinel phase in agreement with experiment.

A significant step forward was the in-silico generation of the melt-quenched amorphous phase. The first simulation of the melt-quenched amorphous $Ge_2Sb_2Te_5$ was reported by *Caravati* et al. [46.49]. The liquid structure (270 atom) generated at 2300 K was equilibrated for 6 ps and then quenched in 16 ps and further equilibrated for 18 ps at 990 K. In order to generate a model of a-$Ge_2Sb_2Te_5$, the liquid was then brought to 300 K in 18 ps. Subsequently these studies were extended [46.50]. The obtained results can be summarized as follows. Ge and Sb atoms are mostly fourfold coordinated and form bonds preferentially with Te atoms. A large fraction of Ge atoms ($\approx 33\%$) were found on tetrahedral sites. It was additionally found that the presence of bonds with Ge or Sb favors a tetrahedral geometry.

Akola and *Jones* [46.51, 52] performed similar simulations using significantly larger cells (460 atom as opposed to the 270 atom of [46.49]) and carried out molecular dynamics simulations over hundreds of picoseconds, times that are comparable to the experimentally observed (ca. 1 ns) amorphization times.

The Ge-Sb (2.78 Å) and Sb-Te (2.93 Å) bond lengths of a GST were found to be shorter than in the crystalline form, but longer than those obtained from experimental studies. The most prominent topologies reported were Ge-Te_4, Sb-Te_3, and Te-$Ge(Sb)_3$. The total coordination numbers using a distance cut-off of 3.2 Å were 4.2 (Ge), 3.7 (Sb), and 2.9 (Te). About 60% of Ge atoms were found to be fourfold coordinated but only 34% were tetrahedrally bonded.

The study found square fragments dominated the structure of a GST. Denoting Te atoms as A and Ge/Sb atoms as B, the authors introduced ABAB squares as the building blocks of the a GST structure (Fig. 46.5). The rapid amorphous-to-crystalline transition was viewed by the authors as a vacancy-supported reorientation of ABAB squares. The same authors subsequently studied the $Ge_8Sb_2Te_{11}$ composition used in Blu-ray disks [46.53].

Hegedüs and *Elliott* [46.54] performed a comprehensive study of phase-change atomistics in $Ge_2Sb_2Te_5$ throughout the phase-change cycle. Starting with the liquid phase, they found the existence of four-membered rings (4-rings), analogous to the ABAB blocks of [46.51], even at temperatures as high as 1073 K. They found that the concentration of 4-rings increased as the temperature decreased and this increase was found to be correlated with the size of the maximal cluster of *connected square rings*.

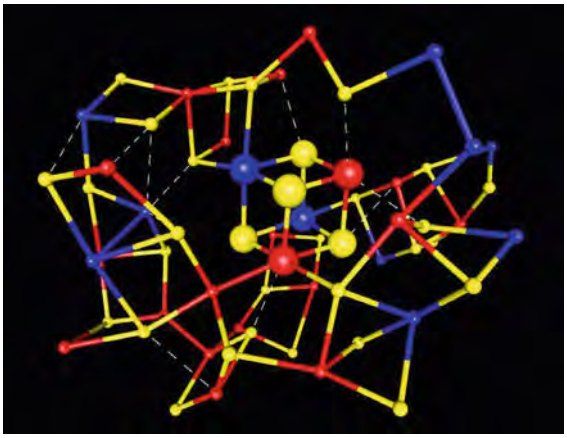


Fig. 46.5 A fragment of the amorphous structure showing an ABAB cube [46.51]. (Copyright 2007 by the American Physical Society. Reprinted with permission)

The results obtained using the DFT simulations described earlier clearly demonstrated that not all Ge atoms acquire tetrahedral symmetry positions in the amorphous phase, although it was argued that the concentration of the tetrahedral Ge sites depends on the simulation details and can reach a value of 50% [46.55].

Krbal et al. [46.56] performed a detailed Ge *K*-edge XANES analysis on the MQ amorphous model of *Hegedüs* and *Elliott* [46.54] that contained 90 atom in the simulation cell, including 20 Ge atoms. The analysis of this amorphous model of GST [46.54] revealed that the local structure around Ge atoms can be grouped into three characteristic bonding configurations with respect to the first-nearest neighbor geometry:

1. Purely tetrahedral (T_d) configurations
2. Pyramids (P_y), with a Ge atom at the apex and with Te-Ge-Te angles very close to 90°
3. Highly distorted octahedral (O_h) structures.

What makes the group-IV element Ge form P_y configurations in the MQ amorphous phase? The answer to this question derives from the definition of a covalent bond, a concept that is characterized by the sharing of a pair of electrons between two bonded atoms [46.57]. Germanium has a $s^2p_x^1p_y^1p_z^0$ electronic configuration in its outer shell. In most materials, these orbitals hybridize and the electronic configuration changes to sp^3, allowing for the formation of four equivalent covalent bonds.

A rather special situation is created when Ge bonds to a chalcogen. The electronic configuration of Te is $(s^2)p_x^1p_y^1p_z^2$, that is, there are two unpaired p electrons and a nonbonding lone-pair in the outer shell. The Te lone-pair p electrons have the option of interact-

ing with the empty p-orbital of an unhybridized Ge atom, forming a two-electron dative bond, which, once formed, is indistinguishable from usual covalent bonds. A similar approach was used to explain the existence of Ge(3):Te(3) configurations in [46.58], where the authors also considered the formation of Ge-Te bonds that involved the lone-pair electrons of Te.

How does the incorporation of Sb atoms affect the structure and properties of this type of material? An Sb atom has one unpaired valence electron on each of its three p-orbitals and can thus form three conventional covalent bonds. On the other hand, Te atoms can form two conventional covalent bonds with Sb atoms; its lone-pair electrons cannot be used. As a result, two Sb atoms (shown in magenta in Fig. 46.6) are needed to replace three Ge atoms to ensure that all interatomic bonds are saturated, which determines the stability of the GeTe-Sb_2Te_3 tie-line. This substitution generates a vacancy on a Ge site. The three Te atoms (a Te triad) located next to the vacancy are twofold coordinated and possess unused lone-pair electrons. Because the Te atoms are low (twofold) coordinated, they can rather easily move in space in the directions show by the arrows in the figure. With 20% of Ge sites being vacant in a typical GST alloy $Ge_2Sb_2Te_5$, the concentration of lone pairs is on the order of 10^{21} cm^{-3}.

The presence of lone-pair orbitals nearly aligned with covalent Ge-Te bonds leads to the formation of three-center four-electron (3c–4e) Te-Ge-Te bonds [46.59] as illustrated in Fig. 46.7. The creation of such

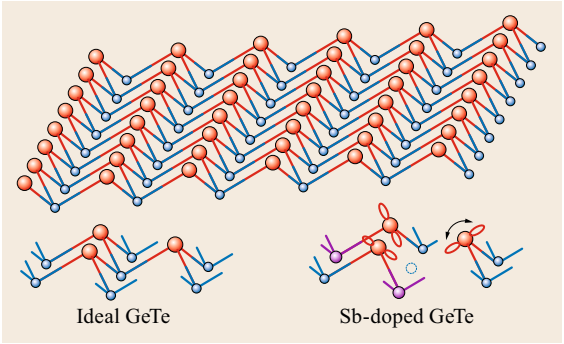


Fig. 46.6 Schematic of the formation of a vacancy and twofold coordinated Te atoms as a result of Sb substitution in GeTe [46.59]. Ge atoms – *blue*, Sb atoms – *magenta*, Te atoms – *orange*. (Copyright 2013 by the American Physical Society. Reprinted with permission)

bonds with equal Te-Ge and Ge-Te distances on both sides of the central Ge atom has several important consequences. First, it leads to the strengthening of interlayer interactions reducing the bonding energy hierarchy characteristic of the ideal binary rhombohedral GeTe phase and makes the structure locally more *cubic*. Second, the similarity in the bond-lengths and energies lead to a more conducive environment for resonance bonding in the structure with 3c–4e bonds, making quasibinary GeTe-Sb_2Te_3 different from the ideal GeTe case that is characterized by a strong bond energy hierarchy.

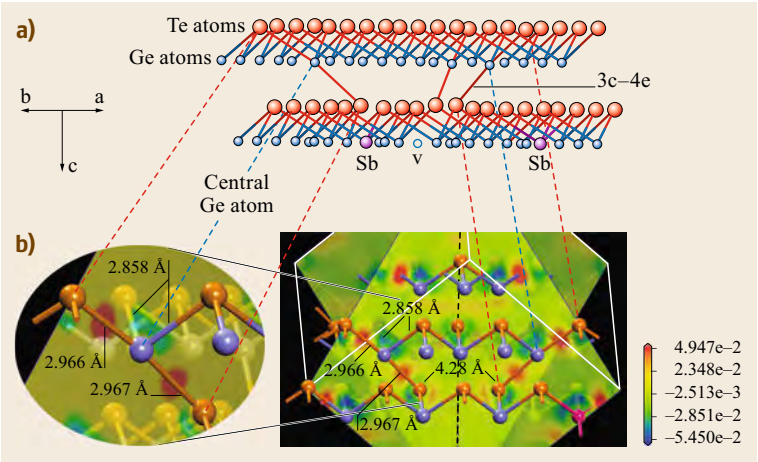


Fig. 46.7a,b Schematic of the formation of three-center four-electron Te-Ge-Te bonds **(a)** utilizing Te lone-pair electrons of the twofold coordinated Te atoms located around the Ge vacancy. Te atoms are shown in *orange*, Ge atoms in *blue*, Sb atoms in *magenta*, and the Ge-site vacancy is shown as an *empty circle*. **(b)** Shows the result of DFT simulation of the three-center Te-Ge-Te bond complexes. Two CDD slice – separated by a *black dashed line* – one for each of the two 3c–4e bonds, are shown on the *right*. The *red spots* of similar size and color midway between the Ge and Te atoms in the CDD map indicate covalent (like) interaction of similar strength on *both* sides of the central Ge atom along the Te-Ge-Te bond direction [46.59]. (Copyright 2013 by the American Physical Society. Reprinted with permission)

46.2 Mechanism of the Phase-Change Process

46.2.1 Structural Studies

In early work, it was tacitly assumed that upon exposure of the crystalline phase to an intense laser or current pulse, phase-change materials melt and are subsequently quenched into a (completely) disordered amorphous state. Recent years have witnessed significant progress in the understanding of the atomistic mechanism of phase change.

The first model that attempted to provide an atomistic description of the phase-change process was the umbrella-flip model [46.20]. Subsequent ab initio studies performed by different groups have not only confirmed the existence of Ge atoms on tetrahedral symmetry sites (about 30% of the total number of Ge atoms) but also found the presence of regular ABAB building blocks (A = Ge, Sb, B = Te) alternatively referred to as four-membered rings, square rings, or even-membered rings [46.45, 51, 52, 54] in amorphous GST. The crystallization was explained by the ordering of pre-existing four-membered rings.

The proposal that the bonding in the crystalline state of GST is resonant while in the amorphous state it is purely covalent [46.23–25] implicitly suggested that the underlying mechanism of the phase transition consisted of the establishment – or destruction – of resonance bonding between the covalently bonded fragments. In the crystalline phase, resonant bonding leads to the ordering and alignment of p orbitals on adjacent molecular units. This alignment is lost in the amorphous phase and this has a drastic effect on materials's properties such as optical contrast and resistivity [46.26].

The phase-change (amorphization) process can be viewed as the destruction of three-centre bonds with the transformation of symmetrical three-center Te-Ge-Te bonds in the crystalline phase with equal Ge-Te distances on both sides of the central Ge atom into asymmetric structures with different Ge-Te distances in the amorphous phase [46.59] as illustrated in Fig. 46.8. This process requires very small atomic displacements and may alternatively be viewed as the formation/destruction of connected four-membered rings [46.54] or ABAB blocks [46.51], provided the bond cut-off distances are properly chosen to account for the formation and/or destruction of the 3c–4e bonds.

Subsequently, it was argued that Ge-Ge bonds in the amorphous phase are formed as a result of the formation of dynamic three-center bonds [46.60], which becomes possible due to the presence of lone-pair electrons subtended at partially hybridized threefold coordinated Ge atoms [46.61]. The Ge-Ge bonds serve to stabilize the amorphous phase. The process is illustrated in Fig. 46.9.

Finally, it is important that the *destabilization of the subsystem of* the weaker bonds with the preserved covalent backbone not only allows for the low-temperature amorphization of the material, however, the preserved covalent backbone also ensures the memory of the initial crystalline structure making the (reverse) crystallization process fast and ensuring high cyclability of the crystallization–amorphization process.

46.2.2 Photo-Assisted Amorphization of $Ge_2Sb_2Te_5$

While it is generally believed that the role of light in the phase-change process is to heat the material above the melting point – which is likely to be correct for longer pulses – exposure to shorter pulses is different. The photo-induced amorphization of $Ge_2Sb_2Te_5$ upon exposure to 600 ps pulses was first studied using in-situ x-ray absorption spectroscopy [46.62].

It was found that following exposure to a short laser pulse the white-line intensity (the spectra maximum after the absorption edge) first monotonically decreases and reaches a minimum value in about 1 ns. This fast initial decrease in the white-line intensity is followed by its partial recovery and within ≈ 2 ns a new saturation value is reached. Most interestingly, the minimum value of the white-line intensity during the amorphization

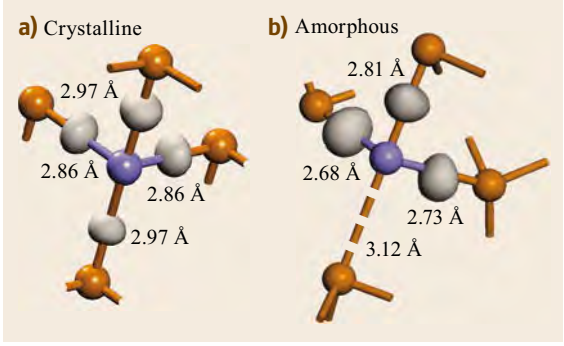


Fig. 46.8a,b Fragments of *defective octahedral Ge sites* in the crystalline **(a)** and amorphous **(b)** phases. Despite very similar *atomic* structures, indicating minute atomic motion, the character of interatomic interaction changes drastically as illustrated by the shown CDD isosurfaces (*gray*) obtained through DFT simulations. While in the crystalline phase the central Ge atom has covalent like interactions with four Te neighbors, in the amorphous phase, it is covalently bonded to only three Te atoms [46.59]. (Copyright 2013 by the American Physical Society. Reprinted with permission)

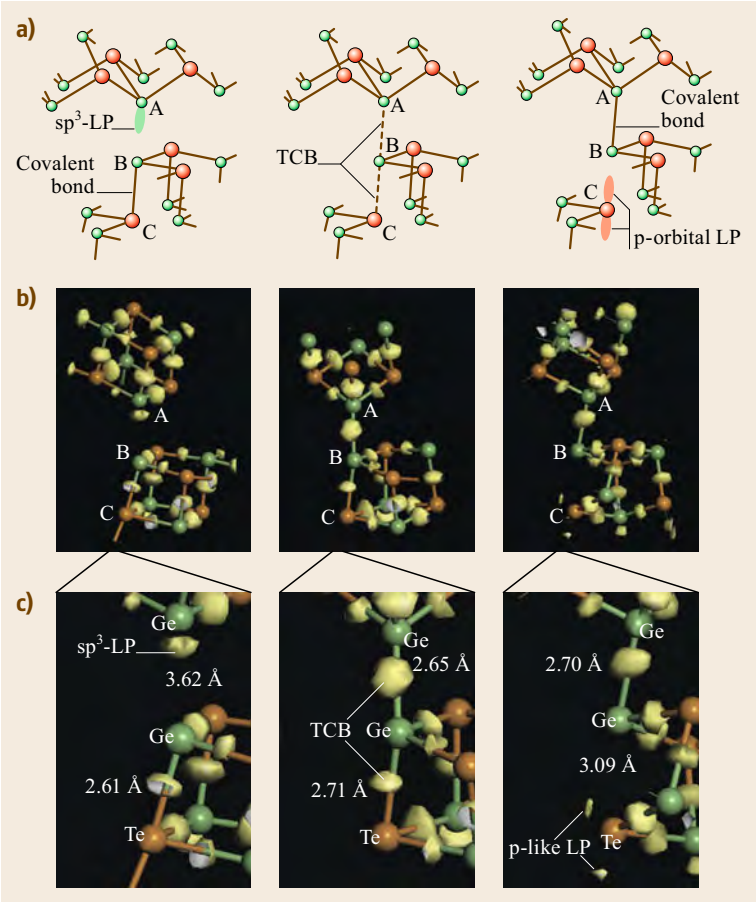


Fig. 46.9 **(a)** Schematic of the formation of a tetrahedral Ge configuration. Ge atoms are shown in *green* and Te atoms are shown in *orange*. When a Ge atom with an LP-orbital (*marked* A) comes close to another Ge atom (B) and is aligned with the neighboring Ge-Te bond (between atoms B–C) (*left panel*), a three-center A–B–C bond is established (*middle*), whose subsequent rupture at the opposite arm results in the formation of a GeTd–GePy configuration (between atoms A–B), leaving behind a twofold coordinated Te atom (C) (*right*). **(b)** Evolution of CDD clouds during the in-silico amorphization process using DFT simulations substantiating the schematic shown in the *upper panel*. CDD clouds corresponding to the LP-electrons of an sp^3 hybridized Ge orbital and a Te lone-pair p-orbital can be seen in the *left* and *right panels*, respectively, in addition to increased CDD midway between Ge and Te atoms that are signatures of covalent bonds. The presence of CDD clouds on both sides of the Ge atom (*marked* B in the figure) in the central panel is evidence of the formation of a transient three-center Ge-Ge-Te bond. **(c)** Zooms into vicinities of the atoms that participate in the formation of three-center bonds. (Reproduced from [46.60])

process is *significantly higher* than that corresponding to the static liquid state [46.63] suggesting that $Ge_2Sb_2Te_5$ does not melt in the conventional sense upon its transformation from the crystalline to the amorphous phase [46.62].

The role of electronic excitation in the phase-change process was also investigated theoretically [46.64]. The electronic excitation was modeled by removing electrons from the high lying valence band states according to the strength of the excitation, where the authors used a jellium background charge to compensate for the loss of charged carriers. From these studies it was concluded that electronic excitation considerably lowers the critical amorphization temperature and reduces the atomic diffusion coefficient with respect to that of the corresponding liquid phase by at least one order of magnitude.

Experimental evidence for the important role of nonthermal processes is gradually growing. Thus, rapid nonthermal control of resonant bonding was demonstrated in [46.65] using a combination of single-shot femtosecond electron diffraction and optical spec-

troscopy. In line with this, in [46.66], ultrafast time-resolved electron diffraction and single-shot optical pump-probe measurements were used to study lattice dynamics in polycrystalline $Ge_2Sb_2Te_5$ initiated by a femtosecond near-ultraviolet pulse and it was concluded that a nonthermal crystal-to-amorphous phase transition was initiated by the displacement of Ge atoms.

46.3 Present Applications and Future Trends

Phase-change memory alloys have been used since the 1990s in rewritable optical disks such as compact disks (CD), and later also in digital versatile disks (DVD) and BluRay disks (BD). The structure of these disks is essentially the same and is shown in Fig. 46.10. What varies is the thickness of individual layers, which is determined by the optics used (the length of the numerical aperture and laser wavelength), and the exact composition of the PCM material. Thus, in DVDs $Ge_2Sb_2Te_5$ is used but in order to optimize the optical contrast at shorter wavelengths $Ge_8Sb_2Te_{11}$ is used in BDs. Figure 46.11 shows the different phase change alloys used for rewritable optical disks on a Ge-Sb-Te ternary phase diagram.

Electrical cell designs for the large-scale integration of phase-change devices have reached commercial development. While lateral device designs such as that proposed by NXP Semiconductors (Fig. 46.12) showed early initial promise [46.68], industry has focused on variations of the mushroom cell design (Fig. 46.13) with a layer of phase-change material (typically GST) sandwiched between two conducting electrodes due to the greater areal density achievable as well as the possibility for three-dimensional stacking of cells. Energy requirements for mushroom like cells have been found to scale roughly linearly with contact area down to effective contact diameters of 3 nm [46.69, 70]. In 2010, Samsung reported on the commercial production of 512 Mbit PRAM chips for use its cellular phones. In 2012, a 20 nm sized mushroom-type device was also announced in a 8 Gb chip by Samsung [46.71]. In 2012, Micron also reported full-scale commercial production of 45 nm node, 1 Gbit PRAM cells. Micron and Intel in July 2015 jointly announced a new generation of what is believed to be a new stackable form of PRAM called *3-D cross-point memory* with the initial chips produced reportedly having a capacity of 128 Gbit. These chips were slated to reach the market in 2017. Considering the widespread use of memory devices, for example, in computers or mobile phones, it may be fair to say that most households in the world have phase-change memory devices. Few other electronic materials can compete with this.

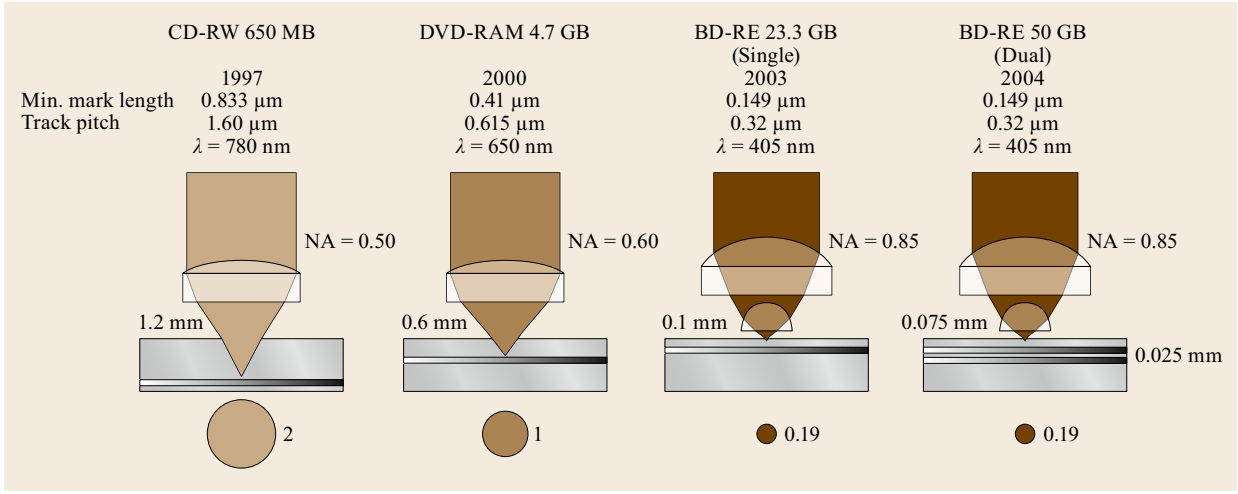


Fig. 46.10 Comparison of optics, recording densities, recording capacities and disk structures used for CD, DVD, and BD. Changes in the lens system are shown in the upper portion; the disk structure is depicted beneath. The change in beam cross-section is displayed at the bottom. From CD to DVD and DVD to BD, the effective laser beam cross section has been reduced by a factor of 2 and 5, respectively. The recording density increased by a factor of 7 going from CD to DVD. Going from DVD to the dual-layer BD another increase by a factor of 11 has been accomplished. (Reprinted from [46.67] by permission from Macmillan Publishers Ltd., copyright (2007))

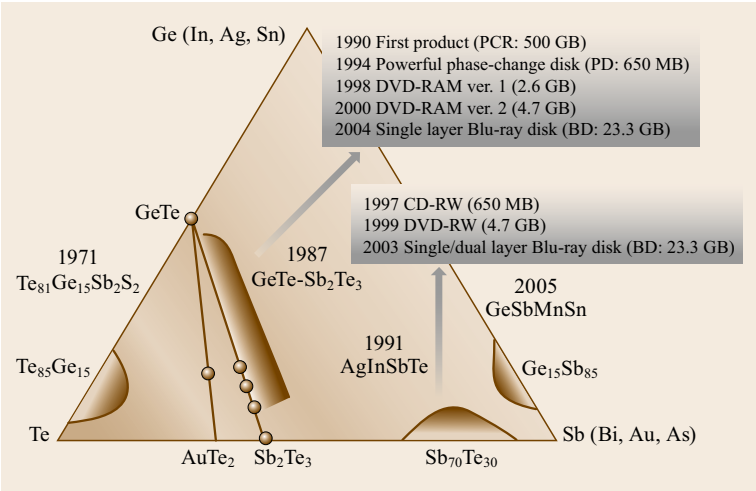


Fig. 46.11 Ternary phase diagram depicting different phase-change alloys, their year of discovery as a phase-change alloy and their use in different optical storage products. (Reprinted from [46.67] by permission from Macmillan Publishers Ltd., copyright (2007))

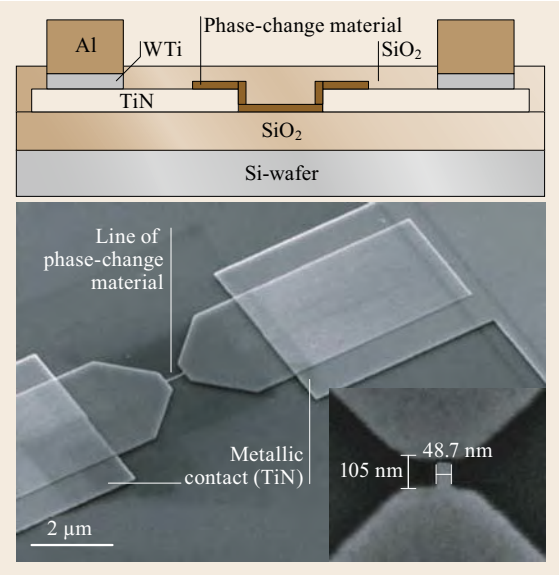


Fig. 46.12 A schematic of a lateral type line cell along with a SEM image of an actual cell. (Reprinted from [46.68] by permission from Macmillan Publishers Ltd., copyright (2005))

A major development in energy efficiency was realized by the spatial separation of GeTe and Sb_2Te_3 into atomically thin layers. Devices based upon this structure were recently shown to result in a significant reduction (by 90%) in energy consumption, which was interpreted as being due to significantly decreased entropic losses [46.72] arising from the one-dimensional motion of Ge atoms at interfaces [46.72, 73]. As a consequence, this kind of memory materials was named interfacial phase-change memory (iPCM).

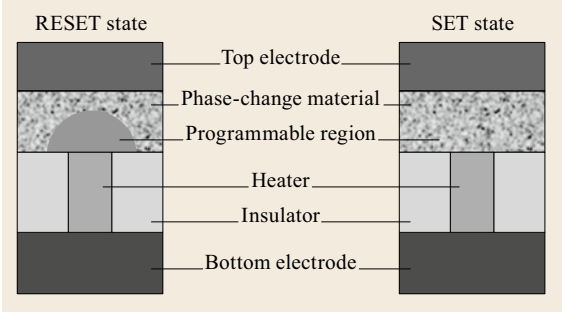


Fig. 46.13 A schematic of a mushroom-type cell in both RESET and SET states

Subsequent studies demonstrated that iPCM, as well as hexagonal $Ge_2Sb_2Te_5$ with certain stacking order, can be a topological insulator or a Dirac semimetal [46.74, 75]. In addition, interesting magnetic responses from iPCM were observed [46.76, 77] that were not present in composite material of the same average composition, making iPCM-based structures promising for spintronic and topotronic applications.

In another development, researchers have grown PRAM on flexible (Kapton) substrates with bottom electrode dimensions of 150 nm. Problems with conventional submicron lithography and flexible substrates were avoided by using a Si-based block co-polymer for patterning in lieu of conventional optical lithography [46.78]. Repeated cycling even after flexing of the substrate demonstrated the potential application of such cell designs to future wearable electronics.

Although the ideas behind neuron-based computing date back to the 1940s with the advent the theory of threshold computing [46.79], the development of electrical devices capable of carrying out such operations in a physical system came much later. In the 1980s

with the pioneering work of *Hopfield* et al. [46.80], the concept of neural networks came into vogue. *Ovshinksky* was the first person to discuss the cognitive computing and neural networks in the context of phase change materials [46.81]. An example of $Ge_2Sb_2Te_5$ based synapses exhibiting the spike-timing-dependent plasticity required for neuromorphic computing was reported in 2012 [46.82]. More recently, a large-scale 165 000 synapse neural network based upon phase-change materials was fabricated and in a prototypical use applied to the recognition of handwriting [46.83].

With the commercial realization of phase-change memory, the field of phase-change materials is primed for further progress. With parallel developments occurring in diverse areas such as neuromorphic computing, iPCM structures, as well as manipulation of the topological properties of chalcogenide-based superlattices, the next 10 year of research will bring a rich array of different devices to the forefront of technology.

References

46.1 S. Ovshinsky: Phys. Rev. Lett. **21**, 1450 (1968)
46.2 A.D. Pearson, W. Northover, J.F. Dewald, W. Peck: *Chemical, physical, and electrical properties of some unusual inorganic glasses in Glass Tech* (Plenum, New York 1962) p. 357
46.3 B.T. Kolomiets, E.A. Lebedev: Radio Eng. Electron. **8**, 1941 (1963)
46.4 K. Rabe, J. Joannopoulos: Phys. Rev. B **36**, 6631 (1987)
46.5 J.-P. Gaspard, A. Pellegatti, F. Marinelli, C. Bichara: Philos. Mag. B **77**, 727 (1998)
46.6 T. Chattopadhyay, J. Boucherle, H. Von Schnering: J. Phys. C **20**, 1431 (1987)
46.7 P. Fons, A.V. Kolobov, M. Krbal, J. Tominaga, K. Andrikopoulos, S. Yannopoulos, G. Voyiatzis, T. Uruga: Phys. Rev. B **82**, 155209 (2010)
46.8 T. Matsunaga, P. Fons, A.V. Kolobov, J. Tominaga, N. Yamada: Appl. Phys. Lett. **99**, 231907 (2011)
46.9 I. Petrov, R. Imamov, Z. Pinsker: Sov. Phys. Cryst. **13**, 339 (1968)
46.10 B. Kooi, J.T.M.D. Hosson: J. Appl. Phys. **92**, 3584 (2002)
46.11 T. Matsunaga, N. Yamada, Y. Kubota: Acta Cryst. B **60**, 685 (2004)
46.12 N. Yamada: MRS Bull. **21**, 48 (1996)
46.13 T. Nonaka, G. Ohbayashi, Y. Toriumi, Y. Mori, H. Hashimoto: Thin Solid Films **370**, 258 (2000)
46.14 N. Yamada, T. Matsunaga: J. Appl. Phys. **88**, 7020 (2000)
46.15 T. Matsunaga, R. Kojima, N. Yamada, K. Kifune, Y. Kubota, Y. Tabata, M. Takata: Inorg. Chem. **45**, 2235 (2006)
46.16 T. Matsunaga, N. Yamada: Phys. Rev. B **69**, 104111 (2004)
46.17 T. Matsunaga, R. Kojima, N. Yamada, K. Kifune, Y. Kubota, M. Takata: Appl. Phys. Lett. **90**, 161919 (2007)
46.18 T. Matsunaga, H. Morita, R. Kojima, N. Yamada, K. Kifune, Y. Kubota, Y. Tabata, J.-J. Kim, M. Kobata, E. Ikenaga: J. Appl. Phys. **103**, 093511 (2008)
46.19 B. Hyot, X. Biquard, L. Poupinet: Local structure of amorphous and crystalline GeTe and GeSbTe, Proc. Eur. Phase Change Symp. (2001)
46.20 A. Kolobov, P. Fons, A. Frenkel, A. Ankudinov, J. Tominaga, T. Uruga: Nat. Mater. **3**, 703 (2004)
46.21 S. Shamoto, N. Yamada, T. Matsunaga, T. Proffen, J.W. Richardson Jr., J.-H. Chung, T. Egami: Appl. Phys. Lett. **86**, 081904 (2005)
46.22 S.I. Shamoto, K. Kodama, S. Iikubo, T.U. Taguchi, N. Yamada, T. Proffen: Jpn. J. Appl. Phys. **45**, 8789 (2006)
46.23 G. Lucovsky, R. White: Phys. Rev. B **8**, 660 (1973)
46.24 J. Robertson, K. Xiong, P.W. Peacock: Thin Solid Films **515**, 7538 (2007)
46.25 K. Shportko, S. Kremers, M. Woda, D. Lencer, J. Robertson, M. Wuttig: Nat. Mater. **7**, 653 (2008)
46.26 B. Huang, J. Robertson: Phys. Rev. B **81**, 081204R (2010)
46.27 A.V. Kolobov, M. Krbal, P. Fons, J. Tominaga, T. Uruga: Nat. Chem. **3**, 311 (2011)
46.28 M. Wuttig, D. Lüsebrink, D. Wamwangi, W. Welnic, M. Gillessen, R. Dronskowski: Nat. Mater. **6**, 122 (2007)
46.29 A.H. Edwards, A.C. Pineda, P.A. Schultz, M.G. Martin, A.P. Thompson, H.P. Hjalmarson, C.J. Umrigar: Phys. Rev. B **73**, 045210 (2006)
46.30 F. Betts, A. Bienenstock, S.R. Ovshinsky: J. Non-Cryst. Solids **4**, 554 (1970)
46.31 D. Dove, M. Heritage, K. Chopra, S. Bahl: Appl. Phys. Lett. **16**, 138 (1970)
46.32 O. Uemura, Y. Sagara, M. Tsushima, T. Kamikawa, T. Satoh: J. Non-Cryst. Solids **33**, 71 (1979)
46.33 M. Wakagi, Y. Maeda: Phys. Rev. B **50**, 14090 (1994)
46.34 A. Kolobov, P. Fons, J. Tominaga, A. Ankudinov, S. Yannopoulos, K. Andrikopoulos: J. Phys. Cond. Mat. **16**, S5103 (2004)
46.35 K. Hirota, K. Nagino, G. Ohbayashi: J. Appl. Phys. **82**, 65 (1997)
46.36 G. Fisher, J. Tauc, Y. Verhelle: Vibrational modes and bonding in the amorphous Ge_xTe_{1-x} system. In: *Amorphous and Liquid Semiconductors*, ed. by J. Stuke (Taylor and Francis, London 1974) p. 1259
46.37 S. Hosokawa, Y. Hari, T. Kouchi, I. Ono, H. Sato, M. Taniguchi, A. Hiraya, Y. Takata, N. Kosugi, M. Watanabe: J. Phys.-Cond. Mat. **10**, 1931 (1998)
46.38 S.J. Pickart, Y.P. Sharma, J.P. de NeufvilleL: J. Non-Cryst. Solids **34**, 183 (1979)
46.39 P. Boolchand, B.B. Triplett, S.S. Hanna, J.P. de Neufville: *Mossbauer Effect Methodology* (Plenum, New York 1974)

46.40 P. Pykkö, M. Atsumi: Chem. Eur. J. **15**, 186 (2009)
46.41 A. Ankudinov, B. Ravel, J. Rehr, S. Conradson: Phys. Rev. B **58**, 7565 (1998)
46.42 P. Jovari, I. Kaban, J. Steiner, B. Beuneu, A. Schops, M.A. Webb: Phys. Rev. B **77**, 035202 (2008)
46.43 D. Baker, M. Paesler, G. Lucovsky, P. Taylor: J. Non-Cryst. Solids **352**, 1621 (2006)
46.44 M. Paesler, D. Baker, G. Lucovsky, A. Edwards, P. Taylor: J. Phys. Chem. Solids **68**, 873 (2007)
46.45 S. Kohara, K. Kato, S. Kimura, H. Tanaka, T. Usuki, K. Suzuya, H. Tanaka, Y. Moritomo, T. Matsunaga, N. Yamada: Appl. Phys. Lett. **89**, 201910 (2006)
46.46 K.V. Mitrofanov, A.V. Kolobov, P. Fons, X. Wang, J. Tominaga, Y. Tamenori, T. Uruga, N. Ciocchini, D. Ielmini: J. Appl. Phys. **115**, 173501 (2014)
46.47 S. Gabardi, S. Caravati, G.C. Sosso, J. Behler, M. Bernasconi: Phys. Rev. B **92**, 054201 (2015)
46.48 W. Welnic, A. Pamungkas, R. Detemple, C. Steimer, S. Blugel, M. Wuttig: Nat. Mater. **5**, 56 (2005)
46.49 S. Caravati, M. Bernasconi, T. Kühne, M. Krack, M. Parrinello: Appl. Phys. Lett. **91**, 171906 (2007)
46.50 S. Caravati, M. Bernasconi, T.D. Kuehne, M. Krack, M. Parrinello: J. Phys. Cond. Mat. **21**, 499803 (2009)
46.51 J. Akola, R. Jones: Phys. Rev. B **76**, 235201 (2007)
46.52 J. Akola, R. Jones: J. Phys. Cond. Mat. **20**, 465103 (2008)
46.53 J. Akola, R. Jones: Phys. Rev. B **79**, 134118 (2009)
46.54 J. Hegedüs, S. Elliott: Nat. Mater. **7**, 399 (2008)
46.55 M. Micoulaut, K. Gunasekera, S. Ravindren, P. Boolchand: Phys. Rev. B **90**, 094207 (2014)
46.56 M. Krbal, A.V. Kolobov, P. Fons, J. Tominaga, S.R. Elliott, J. Hegedus, T. Uruga: Phys. Rev. B **83**, 054203 (2011)
46.57 L. Pauling: *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* (Cornell Univ. Press, Ithaca 1960)
46.58 M. Xu, Y. Cheng, H. Sheng, E. Ma: Phys. Rev. Lett. **103**, 195502 (2009)
46.59 A.V. Kolobov, P. Fons, J. Tominaga, S.R. Ovshinsky: Phys. Rev. B **87**, 165206 (2013)
46.60 A.V. Kolobov, P. Fons, J. Tominaga: Sci. Rep. (2015), doi:10.1038/srep13698
46.61 A.V. Kolobov, P. Fons, J. Tominaga: Phys. Rev. B **87**, 155204 (2013)
46.62 P. Fons, H. Osawa, A.V. Kolobov, T. Fukaya, M. Suzuki, T. Uruga, N. Kawamura, H. Tanida, J. Tominaga: Phys. Rev. B **82**, 041203 (2010)
46.63 A. Kolobov, P. Fons, M. Krbal, R. Simpson, S. Hosokawa, T. Uruga, H. Tanida, J. Tominaga: Appl. Phys. Lett. **95**, 241902 (2009)
46.64 X.-B. Li, X.Q. Liu, X. Liu, D. Han, Z. Zhang, X.D. Han, H.-B. Sun, S.B. Zhang: Phys. Rev. Lett. **107**, 015501 (2011)
46.65 K.V. Mitrofanov, P. Fons, K. Makino, R. Terashima, T. Shimada, A.V. Kolobov, J. Tominaga, V. Bragaglia, A. Giussani: Sci. Rep. **6**(20), 633 (2016)
46.66 M. Hada, W. Oba, M. Kuwahara, I. Katayama, T. Saiki, J. Takeda, K.G. Nakamura: Sci. Rep. **5**, 13530 (2015)
46.67 M. Wuttig, N. Yamada: Nat. Mater. **6**, 824 (2007)
46.68 M. Lankhorst, B. Ketelaars, R. Wolters: Nat. Mater. **4**, 347 (2005)
46.69 H.-S. Wong, S. Raoux, S. Kim, J. Liang, J.P. Reifenberg, B. Rajendran, M. Asheghi, K.E. Goodson: Proc. IEEE **98**, 2201 (2010)
46.70 F. Xiong, A.D. Liao, D. Estrada, E. Pop: Science **332**, 568 (2011)
46.71 Y. Choi, I. Song, M.-H. Park, H. Chung, S. Chang, B. Cho, J. Kim, Y. Oh, D. Kwon, J. Sunwoo, J. Shin, Y. Rho, C. Lee, M.G. Kang, J. Lee, Y. Kwon, S. Kim, J. Kim, Y.-J. Lee, Q. Wang, S. Cha, S. Ahn, H. Horii, J. Lee, K. Kim, H. Joo, K. Lee, Y.-T. Lee, J. Yoo, G. Jeong: A 20 nm 1.8 V 8 Gb PRAM with 40 MB/s program bandwidth, Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Pap. (ISSCC) (2012) p. 46
46.72 R.E. Simpson, P. Fons, A.V. Kolobov, T. Fukaya, M. Krbal, T. Yagi, J. Tominaga: Nat. Nanotech. **6**, 501 (2011)
46.73 X. Yu, J. Robertson: Sci. Rep. (2015), doi:10.1038/srep12612
46.74 A.A. Burkov, L. Balents: Phys. Rev. Lett. **107**, 127205 (2011)
46.75 J. Tominaga, A.V. Kolobov, P. Fons, T. Nakano, S. Murakami: Adv. Mat. Interf. **1**, 1300027 (2014)
46.76 J. Tominaga, R.E. Simpson, P. Fons, A.V. Kolobov: Appl. Phys. Lett. **99**, 152105 (2011)
46.77 D. Bang, H. Awano, J. Tominaga, A.V. Kolobov, P. Fons, Y. Saito, K. Makino, T. Nakano, M. Hase, Y. Takagaki: Sci. Rep. (2014), doi:10.1038/srep05727
46.78 B.H. Mun, B.K. You, S.R. Yang, H.G. Yoo, J.M. Kim, W.I. Park, Y. Yin, M. Byun, Y.S. Jung, K.J. Lee: ACS Nano **9**, 4120 (2015)
46.79 W. McCulloch, W. Pitts: Bull. Math. Biophys. **5**, 115 (1943)
46.80 J.J. Hopfield: Proc. Nat. Acad. Sci. **79**, 2554 (1982)
46.81 S.R. Ovshinsky, B. Pashmakov: Innovation providing new mulitiple functions in phase-change materials to achieve cognitive computing, Proc. Symp. GG/HH – Adv. Charact. Tech. Data Storage Mater. – Phase Change Nonmagnetic Mater. Data Storage (2003)
46.82 D. Kuzum, R.G.D. Jeyasingh, B. Lee, H.-S.P. Wong: Nano Lett. **12**, 2179 (2012)
46.83 G. Burr, R. Shelby, S. Sidler, C. di Nolfo, J. Jang, I. Boybat, R. Shenoy, P. Narayanan, K. Virwani, E. Giacometti: IEEE T. Electron Dev. **62**, 3498 (2015)

47. Carbon Nanotubes and Bucky Materials

Mark Baxendale

The chapter details the underlying phenomena that underpin electronic applications that have followed from the discoveries of C_{60} and carbon nanotubes. The reduced dimensionality of these self-organised structures, high electron mobility, weak electromigration, and the plethora of quantum electronic effects exhibited by these structures suggest they are serious candidates for molecular electronics. The detail of the surface chemistry and conditions of synthesis assume greater importance than for conventional electronic materials since all atoms are on the exterior of these structures, as is outlined with references to the wider literature. Essential electronic structure information is given with reference to the transport measurements that have contributed greatly to the evolution of the field with emphasis on the Coulomb blockade and ballistic transport phenomena. The major electronic applications are then outlined, giving the state-of-the-art figures of merit for performance and comments on prospects for realisation.

47.1 **Carbon Nanotubes** 1163
47.1.1 General 1163
47.1.2 Geometry 1165
47.1.3 Synthesis and Chemistry 1165
47.1.4 Electronic Structure and Transport 1166
47.1.5 Nanoelectronic Devices 1168
47.1.6 Networks 1168
47.1.7 Other Electronic Applications 1169
47.2 **Bucky Materials** 1170
References 1170

The discovery of C_{60} (Fig. 47.1) in 1986, and subsequently higher fullerenes, followed by that of carbon nanotubes in 1991 accelerated the interest in carbon-based molecules and compounds as electronically or optically active materials that had begun in the early 1970s. The award of the 2010 Physics Nobel Prize to Andre Geim and Konstantin Novoselov for the isolation and groundbreaking experiments on graphene emphasized the importance of carbon-based electronic materials. The emergence of the interdisciplinary field of nanotechnology has beneficially attracted attention and methodologies from other fields of science and engineering that have considerably enhanced the physical electronics research effort [47.1, 2].

47.1 Carbon Nanotubes

47.1.1 General

The electronic and structural description of carbon nanotubes and other sp^2-bonded carbon nanostructures with curved surfaces (bucky materials, the term *bucky* is derived from buckminsterfullerene: the name given to C_{60} soon after discovery) will inevitably contain references to that of the familiar two-dimensional (2-D) planar counterpart, graphite. The strength of the in-plane covalent bond of the hexagonal graphitic network produces a short C–C distance of 0.142 nm and high mechanical stiffness.

Carbon nanotube is a generic term that includes both single- and multi-walled structures. Single-wall carbon nanotubes (SWNTs) are seamless cylindrical graphitic macromolecules of nanoscale diameter and micrometer length, capped by hemispherical ends. The closure of the cylinder is a result of pentagonal inclusions in the hexagonal carbon network of the nanotube walls, Fig. 47.2. For SWNTs, every carbon atom is a surface atom. The smallest SWNT diameter reported to date, 4 Å, corresponds to the predicted lower limit for stable SWNT formation from consideration of the stress energy built into the cylindrical structure [47.3]. SWNTs tend to form closely packed bundles, or *ropes*, of several tens of individual single-wall nanotubes.

Multi-wall nanotubes (MWNTs) comprise several to tens of concentric cylinders of these graphitic shells with a layer spacing of 3.4 Å (the inter-planer spacing of graphite). SWNT diameters are typically ≈

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_47

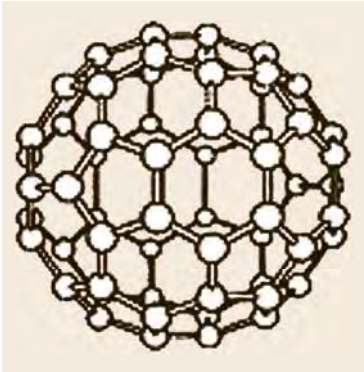


Fig. 47.1 The C_{60} molecule, diameter 0.7 nm

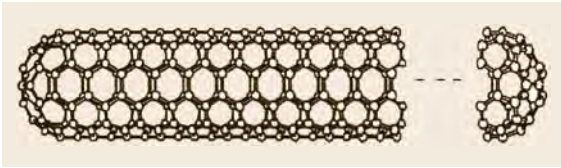


Fig. 47.2 Single-wall carbon nanotube, length 1–10 μm, diameter typically 1.4 to > 10 nm

1 nm and a MWNT diameter can be in the range 2–100 nm with typical values of 10–20 nm. Nanotube lengths are typically 1–10 μm but can be ≈ 1 mm, giving rise to astonishing aspect ratios (length/diameter) of > 1000. MWNTs of similar dimensions can also be formed from a single graphite sheet rolled into a scroll structure [47.4]. Double-wall carbon nanotubes are a recent addition to the set of carbon nanostructures.

A consequence of the high C–C bond strength is that the in-plane Young's modulus of a carbon nanotube is ≅ 1 TPa, making the carbon nanotube one of the stiffest known materials, while the tensile strength is ≅ 150 Gpa (nanotubes have 600 times the strength/weight ratio of steel). Nanotubes deform elastically by buckling, have very low defect density over hundreds to thousands of interatomic spacings, and the strain energy built into the cylindrical structure tends to promote self-repair. The stiff nanotube structure means that the optical-phonon population is minimal even at room temperature; the result can be ballistic electron transport over micrometer length scales (several orders of magnitude greater than that in conventional semiconductor nanostructures).

Another consequence of the short C–C distance is spatially extensive overlapping hybridised atomic orbitals – a π-electron system – with highly mobile delocalised electrons on the interior and exterior of the structure. One carbon atom contributes one unpaired π-electron to the nanotube. The promise of very high carrier mobility in the π-electron system is one motivation for carbon-based electronics; mobilities of circa $10^5\,\mathrm{cm^2V^{-1}s^{-1}}$ have been measured in SWNTs at 50 K and suspended graphene flakes at 240 K [47.5, 6].

Another is the range of electronic attributes that are displayed by the carbon allotropes, e.g., room-temperature resistivity can range from $10^{-8}\,\Omega$m for SWNTs to $> 10^{18}\,\Omega$m for crystalline C_{60}.

The π-electron system and stiff structure produce electrical conductivity, very high current-carrying capacity with weak electromigration. The best figures of merit for nanotube systems that have been measured since discovery are: electrical conductivity, $10^8\,\Omega^{-1}\mathrm{m}^{-1}$ (comparable with that of copper), carrier mobility: $10^4\,\mathrm{cm^2/Vs}$ (comparable with that of high-quality GaAs), thermal conductivity $10^4\,\mathrm{W/mK}$ (comparable with that of diamond), and sustainable current density of $10^{13}\,\mathrm{A/m^2}$ (> 1000 times greater than that of a metal nanowire).

Nanotube circumferential crystal momentum vectors are quantised due to the periodic boundary conditions imposed by the finite nanoscale diameter whereas those directed along the micrometer-scale axis show continuous variation; thus, one-dimensional (1-D) electronic transport is intrinsic for the SWNT. Electronic systems of reduced dimensionality – dots, wires and sheets of charge – created by top-down silicon processing have assumed importance in conventional microelectronics. Carbon nanostructures offer the possibility of self-organised molecular-level low-dimensional electronic systems, e.g.: zero-dimensional (0-D) (C_{60} and short SWNTs), 1-D (SWNTs), 2-D (graphene, large-diameter MWNT outer shell), and three-dimensional (3-D) (coupled-layer MWNTs).

Electronic modification of carbon nanotubes by insertion of C_{60} and other fullerenes into the central capillary to form a mixed-dimensionality hybrid structure called the *peapod* is a science promises a rich future for research and applications [47.7].

Carbon nanotubes are chemically stable in ambient conditions and structurally stable in vacuum for temperatures far greater than 1000 °C. Viewed in total, the above properties potentially make carbon nanotubes excellent electronic materials. These considerations place carbon nanotube electronics among a number of competing technologies poised to complement or replace silicon-based complementary metal–oxide–semiconductor (CMOS) technology in the < 100 nm-feature-size domain. Individual SWNT-based field effect transistors, unipolar and bipolar, show potential for integrated optoelectronic applications. Much recently attention has been given to the production of large-area flexible transparent conducting SWNT networks with indium-doped tin oxide-matching electrical and optical performance [47.8].

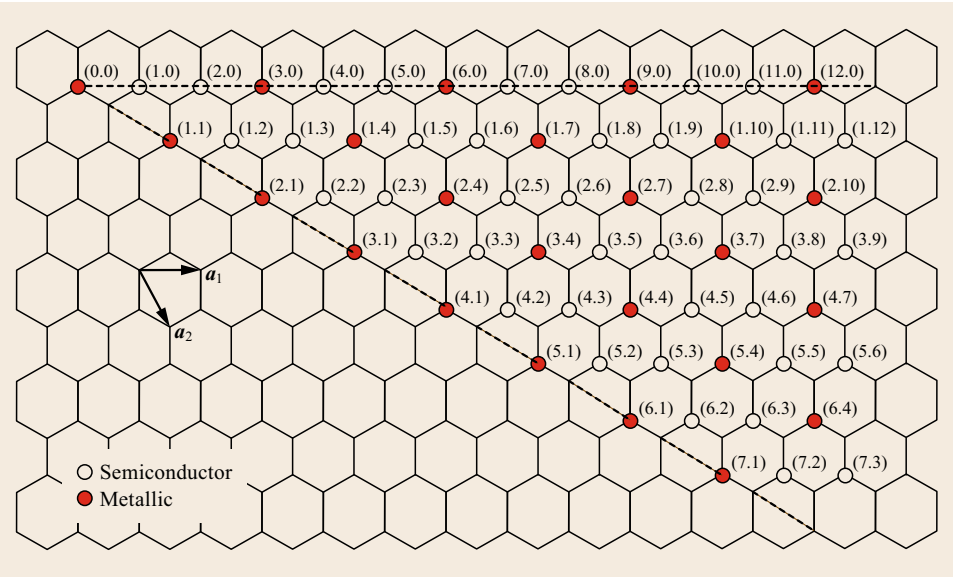


Fig. 47.3 A map of conductivity type for an individual SWNT indicating the n, and m indices of the chiral vector and the unit vectors $\mathbf{a}_1$ and $\mathbf{a}_2$

47.1.2 Geometry

The unique feature of the electronic structure of carbon nanotubes is that it can range from the metallic to the semiconductor depending on the details of the microstructure. Essentially, the way the hexagonal network connects to itself to form a cylinder determines the electronic structure: the chiral vector usually used to describe the wrapping of the network is

$$\boldsymbol{C} = n\boldsymbol{a}_1 + m\boldsymbol{a}_2 \equiv (n, m)\,,$$

where $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are the unit vectors of the hexagonal network, and n and m are integers. The resulting nanotube can then be described in the form (n, m), Fig. 47.3.

Electronic band-structure calculations show that an (n, m) nanotube is metallic at room temperature if $2n + m$ is a multiple of 3, otherwise it is a semiconductor with a band gap of $E_g = 0.9\,d\,\mathrm{eV}$, where d is the nanotube diameter expressed in nanometers, i. e., typically $E_g \approx 0.5\,\mathrm{eV}$. This remarkable property signals the possibility of band-gap engineering by control of the microstructure. The wrapping angle, or chiral angle, is given by the angle between $\boldsymbol{a}_1$ and $\boldsymbol{C}$. From geometric and symmetry considerations,

$$0 < |m| < n\,, \quad 0° = \theta = 30°\,,$$
$$\cos(\theta) = 2n + \frac{m}{2\sqrt{(n^2 + m^2 + nm)}}\,,$$

and nanotube diameter $d = 0.078\sqrt{(n^2 + m^2 + nm)}$ nanometers. The general case (n, m) is referred to as the chiral nanotube; there are two special cases: (a) the zigzag nanotube ($m = 0$, $\theta = 0°$), and (b) the armchair nanotube ($n = m$, $\theta = 30°$, all metallic) [47.2]. Assuming a random distribution of nanotube diameters in the reaction products, the relative populations of metallic to semiconducting electronic structure is 1 : 3. Single-symmetry (n, m) synthesis or post-synthesis sorting according to electronic structure are key fields for research effort. The task is proving rather difficult but there appear to be no fundamental barriers to progress.

A stable defect in the hexagonal network is the pentagon–heptagon pair (Stone–Wales or 5–7 defect). This defect will cause a sharp bend in an otherwise well-graphitised SWNT. Controlled introduction of defects allows the possibility of constructing Y- and T-junctions and other complex geometries from nanotubes, including rings and coils. Moreover, a 5–7 defect can connect a metallic to a semiconductor nanotube giving an Angstrom-scale heterojunction and hence a device density 10^4 times greater than present-day microelectronics [47.9].

47.1.3 Synthesis and Chemistry

The quality of carbon nanotubes in terms of crystallinity and impurity content depends on the conditions of synthesis by arc discharge using graphitic electrodes, laser vaporisation of graphite, or chemical vapour deposition (CVD) using hydrocarbon gas and metal nanoparticle catalysts. Each method produces different nanotube samples and sample-to-sample variation from the same source. An extensive review of synthesis methods was published by *Rakov* [47.10]. The key factor for crystallinity, and therefore electronic quality, is the forma-

tion temperature: > 2000 °C is required for complete graphitisation. Little is know about the preferences the various growth processes have for certain nanotube symmetries over others.

Reaction products can contain significant quantities of residual metal catalyst particles and non-tubular forms of carbon (typically 30 wt% of the raw material). Purification processes ranging from the simple to complex are commonly used: oxidation and acid washing [47.10], solvent treatment followed by ultrafiltration [47.11], or flocculation using aqueous surfactants [47.12]. Controlled growth of MWNTs perpendicular to a substrate has been widely explored due to its technological importance [47.13]

Surface functionalisation of SWNTs is required for several applications as a means of modifying charge exchange with the ambient. Oxidation of nanotubes with the acid mixture H_2SO_4–HNO_3 leads to high concentrations of carboxylic, carbonyl, and hydroxyl groups on the surface and removal of the tips to expose the interior surface. Carboxyl (−COOH) groups are then readily derivatised by a variety of reactions. Covalent functionalisation, however, necessarily disrupts the rigid structure and π-electron system with consequent degradation of mechanical and electrical properties but provides the best stability and accessibility [47.14]. Non-covalent routes to nanotube functionalisation offer ease of synthesis and minimum disruption of the tubular structure [47.15].

Exposure to ambient humidity, oxygen and other gases profoundly affects the measured electronic properties of nanotubes through charge exchange and the quality of metallic electrical contacts made to nanotubes. The electronic properties of carbon nanotube systems are extremely sensitive to the presence of molecular oxygen due to the formation of the charge-transfer complex $C_p^{+\delta}$–$O_2^{-\delta}$, i. e., oxygen-exposed nanotubes are p-type [47.16]. Nanotubes exhibit ultrahigh sensitivity at room temperature to O_2, N_2O, NH_3 but not H_2 [47.17]. Nanotubes must be elevated to temperatures > 700 °C in ultrahigh vacuum to completely remove absorbents.

47.1.4 Electronic Structure and Transport

The electronic structure of SWNTs is predicted by tight-binding calculations extended from the 2-D approach for graphite to the cylindrical 1-D nanotube system [47.2], Fig. 47.4. The tight-binding modelling follows from the original band structure calculations of graphite by *Wallace* in 1947 [47.18].

In the case of nanotubes circumferential crystal momentum vectors are quantised due to the periodic boundary conditions imposed by a finite nanoscale diameter. This simple approach has been proven, mainly by low-temperature scanning tunnelling spectroscopy, to be a good description of the electronic structure of carbon nanotubes [47.19]. The curvature of the nanotubes is ignored in this model but can introduce some important modifications [47.20]. However, the original model is, nevertheless, adequate for the classification of nanotubes into metallic and semiconducting types according the simple rule outlined in Sect. 47.1.1. In common with other 1-D conductors with half-filled bands the expectation was that carbon nanotubes will undergo a Peierls distortion, i. e., and a gap will open in the band structure at the Fermi energy due to atomic displacement modifying the unit cell. However, it has been shown theoretically that the effect very quickly diminishes as a function of nanotube radius [47.2].

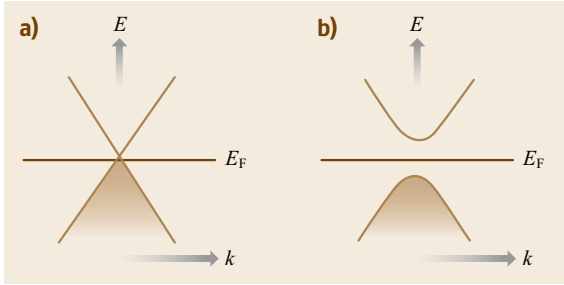


Fig. 47.4 **(a)** Electron energy E versus wavevector k showing the intersection of the dispersion relations of two bands at the Fermi energy E_F; electron occupancy is indicated by shading. The essentially zero-gap semiconductor structure is common to those sp^2-bonded carbon allotropes for which k can vary in two dimensions resulting in a conical shape of the E-k plots (graphene, carbon nanotubes). **(b)** The case of semiconducting SWNT; a band-gap appears owing to the combination of the conical E-k shape and quantisation of the circumferential crystal momentum vector

Electrical transport measurements on individual SWNTs, SWNT ropes, and MWNTs have produced diverse results in the decade since discovery and are still the focus of intense research effort. The picture is complicated by several factors:

1. The current pathway through MWNTs is greatly complicated by interlayer coupling.
2. Two-terminal resistances in the range 100 Ω–100 MΩ were first reported [47.21].
3. The processing, contacting, and manipulation methods used undoubtedly introduced non-intrinsic resistances and potential barriers.

Nevertheless, the best transport parameters that have emerged since discovery, quoted in Sect. 47.1, are highly desirable from many device points of view. The

quality of lithographically defined contacts is critical for electrical transport measurements. There are three possible outcomes of the contacting process:

1. The contacts are of good quality; then the total resistance will be the diffusive or quantum resistance for that structure (6.5 kΩ for a ballistic SWNT) [47.22].
2. No contact is made and the nanotube is isolated by tunnel barriers; in this case the nanotube will act as a Coulomb island and blockade phenomena will be observed at low temperature [47.23].
3. The contacts are of an intermediate type that allows the nanotube to act as a cavity and Fabry–Perot interference phenomena are observed [47.24].

The exact physics of contacting to carbon nanotubes is poorly understood; the exact atomic arrangement at the interface is likely to play a role as are the ambient gases.

Successful transport measurements have yielded a plethora of quantum-electronic and low-dimensional transport phenomena. *Tans* et al. [47.23] demonstrated true quantum wire behaviour; the remarkable feature of this work was to identification of an electron phase-coherence length on the micrometer scale at 4 K for nanotubes: several orders of magnitude greater than that in conventional semiconductor nanostructures. With an extraordinarily skilful experiment *Bachtold* et al. [47.25] demonstrated the Aharonov–Bohm effect by circulating circumferential current in a contacted MWNT with the axis parallel to a magnetic field. The periodicity of the magneto-conductance variation with applied field corresponded to an interferometer diameter equal to that of the outer MWNT shell. This finding gave weight to an earlier conclusion from quantum conductance measurements of *Frank* et al. [47.26], namely that current is only carried in the outer shell for MWNTs. This was posed as a general statement but was contradicted by the controlled layer-by-layer conductivity analysis of *Collins* et al. [47.27] in which the current-carrying contribution of each MWNT layer was determined by vaporisation of successive MWNT layers. This work concluded that at least eight layers contribute to the current-carrying capacity of a MWNT.

The landmark paper of *Frank* et al. identified ballistic conductance in MWNT systems at room temperature by controlled lowering of a MWNT bundle into a liquid mercury contact [47.26]. This was a remarkable result since the implication is that the electron mean free path is on the micrometer scale at room temperature. A ballistic conductor has a conductance given by the Landauer formula

$$G = 2\left(\frac{e^2}{h}\right)\sum_{i}^{N} T_i \,,$$

where N is the number of conducting channels, T is the transmission factor (experimental result: $T = 1$ for most nanotube systems), and $G_0 = 2e^2/h$ is the quantum conductance (the factor of 2 is due to spin degeneracy). Frank et al. measured a conductance of $1G_0 >$ per MWNT, implying that current is carried in the outer shell only and there is a *missing* $1G_0$ since a metallic carbon nanotube has two crossing 1-D energy bands at the Fermi level, which yields a conductance of $2G_0$ for a single nanotube shell. However, subsequent results have shown the expected $2G_0$ result [47.28]. Therefore, the controversy over the experiment by Frank et al. remains unsolved.

SWNTs were expected to exhibit the predicted electron–electron interaction effects in 1-D systems, the so called Luttinger liquid (LL) state. In a strictly 1-D system the independent electron approximation breaks down and the normal 3-D Fermi gas picture is replaced by the LL. In the LL description electron–electron interactions dominate and the electrons are in a highly correlated state. The LL state requires a high degree of order since disorder destroys the correlated ground state. The main features of the LL are suppression of the density of states at the Fermi energy according to a power law and separation of spin and charge excitations. In transport measurements, the LL state is manifested as a tunnelling conductance from a normal electrode according to $G(T) \approx T^{\alpha}$, for $eV \ll k_{\mathrm{B}}T$, and at large voltages $G(V) = \mathrm{d}I/\mathrm{d}V \approx V^{\alpha}$ for $eV \gg k_{\mathrm{B}}T$, where the parameter α is determined by the strength of the interaction between the electrons. Such power laws have been observed in contacted SWNTs, although stronger evidence, such as a clear demonstration of spin–charge separation, in necessary to identify the LL state unambiguously [47.29].

Coulomb blockade phenomena have been observed in low-temperature transport measurements on carbon nanotube systems often by unintentional incorporation of tunnel junctions between the metal electrode and the nanotube. The effect occurs when a mesoscopic *island* with capacitance C (self-capacitance of the nanotube plus the barrier capacitance at the nanotube–electrode junction) is isolated by tunnel barriers from two electrodes. If the charging energy for a single electron on the island, e^2/C, is larger than the thermal energy $k_{\mathrm{B}}T$, then the system will not conduct due to electron–electron repulsion. Coulomb blockade can be modulated by a gate electrode from full blocking to a conducting state limited only by the tunnel barriers. Such a system forms a single-electron transistor (SET). The capacitance of a nanotube is $\approx 10^{-17}F$, thus SET operation conditions are fulfilled at temperatures below 10 K when the entire nanotube defines the Coulomb island. To minimise the capacitance and therefore el-

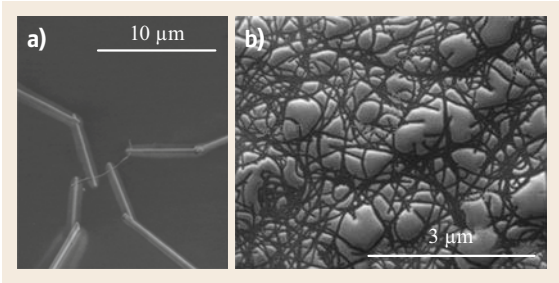


Fig. 47.5 **(a)** Example of a contacted individual MWNT on the insulating SiO_2 layer of a conducting silicon substrate. **(b)** Example of a network comprising SWNT ropes depostied on a quartz substrate

evate the operating temperature inter-tube islands can be defined by crossing nanotubes or intra-tube islands bending with a scanning probe microscope (kinks and bends in SWNTs can create tunnel barriers). The latter approach has been used to create a SET that operates close to room temperature [47.30]. Graphene is likely to be the carbon allotrope to produce usable SET devices owing to the ease with which flakes can be lithographically processed [47.31].

Superconductivity in nanotube systems was first observed as the proximity effect with SWNTs or SWNT ropes connecting two superconducting electrodes [47.32]. The first report signalling intrinsic superconductivity in SWNT ropes was by *Kociak* et al. in 2001 [47.33]. Subsequently, 1-D superconductivity has been observed in arrays of small-diameter SWNTs at 15 K and in indvidual MWNTs up to 12 K [47.34, 35].

Two broad categories of exploitation of the above electronic attributes have arisen; nanoelectronic devices, with an individual SWNT or MWNT as the acive channel in a source-drain configuration deposited on the top SiO_2 layer of a heavily doped silicon substrate (which serves as a backgate), and applications which exploit networks of *ropes* or bundled SWNTs, Fig. 47.5. Although, the intermediate approach of using highly aligned SWNT arrays as the active channel has had some success [47.36].

47.1.5 Nanoelectronic Devices

Semiconducting nanotubes are especially important for nanoelectronic device applications. Semiconducting behaviour has been observed in individual SWNTs; in SWNT ropes and MWNTs mixed metallic and semiconducting current pathways greatly complicate the I–V characteristic. Field-effect transistors (FETs) with individual SWNT channels have been demonstrated and found to have higher transconductance than can be achieved with state-of-the-art metal–oxide–semiconductor field-effect transistors (MOSFETs) [47.37]. The conducting state is attained with negative gate voltages, implying that the carbon nanotube forms a normally off p-type conduction channel (probably due to unintentional doping by exposure to oxygen). Logic gates made from nanotube FETs have been demonstrated [47.38]. The nature of the metal electrode–SWNT Schottky barrier is still not fully understood.

The prospects for highly integrated circuits consisting of nanotube elements must be considered remote at this time for several reasons (some previously mentioned):

1. Current synthesis techniques produce mixtures of metallic or semiconductor nanotubes and these tend to form ropes or bundles.
2. The precise nature of the metal electrode–nanotube contact is poorly understood.
3. Scanning probe manipulation techniques used to fabricate prototype individual devices are not scalable.

Nevertheless, carbon nanotube-based FETs, both unipolar and bipolar, show some promise for intergrated optoelectronic applications [47.39].

47.1.6 Networks

This work is largely driven by the necessity to find an alternative to the conventional transparent conducting electrode material, indium-doped tin oxide (ITO), owing to the terrestrial rarity of indium and the desirability of flexibility for some emerging technologies. Individual SWNTs constitute ideal elements for flexible transparent conducting networks with ITO-matching electrical and optical performance, namely, that is, sheet resistance much less than $< 100\,\Omega/\mathrm{sq}$ optical transmission of 90% at 550 nm wavelength. The basic network element is unavoidably a bundle rather than an individual SWNT. Electrical transport within the bundle and the inter-bundle contact resistance has greatest influence on the electrical transport within the network; contact resistances can be minimised by careful removal of residual surfactant introduced by solution processing [47.5, 6, 11]. SWNT networks have been modelled as metallic filamentary pathways interrupted by thin tunnel barriers, backscattering by zone-boundary phonons, fluctuation-assisted tunnelling, and variable-range hopping (VRH). The contribution of each component is a function of network thickness; VRH dominates in the thinnest networks and metallic conduc-

tivity increases with thickness [47.16, 17]. Approaches to transparent conducting electrode production include chemical doping and electrical modification by various methods, including acid treatment [47.11, 12], decoration by metal nanoparticles [47.3, 13, 40], conducting polymer composites, graphene-SWNT hybrids [47.14, 15], metallic enrichment, and purification through ultracentrifugation [47.7].

Many approaches for the production of such hybrid networks have been reported, including, physical evaporation of metal nanoparticles onto the SWNT network, nanoparticle attachment by chemical reaction with functionalised SWNTs, and electroless nanoparticle deposition methods [47.24, 29–33].

The room temperature sheet resistance of the resultant hybrid networks has been reported to be both increased and decreased relative to the un-decorated SWNT network [47.29, 34, 35]. A 1/1500 reduction of room temperature sheet resistance of SWNT network by Au-nanoparticle decoration with an optical transmittance of 90% at 550 nm wavelength has been reported [47.40].

Metal nanoparticle-decorated SWNT networks show promise for selective gas sensing based on differential analyte-nanoparticle reactivity and charge exchanges with the network. There are also useful plasmonic and thermoelectric benefits from such hybrid systems [47.20–22]. The metal nanoparticle-network interaction is not well understood. Decoration of individual SWNTs with gold nanoparticles results in electron transfer from the nanotube to the nanoparticle and the formation of a potential barrier at the SWNT-metal interface; the barrier characteristics depend on the work function of the metal [47.13, 23, 24, 26–28, 36].

47.1.7 Other Electronic Applications

Although nanoelectronics has been the driving force for carbon nanotube research, many other electronic applications are being explored; this section outlines some of them.

Interconnects

Ballistic transport on the micrometer scale at room temperature and high current-carrying capacity suggests carbon nanotubes are good candidates for high-bandwidth dissipation-less interconnect for nanoscale circuit elements. Since chemical processing tends to degrade nanotube electrical properties, methods of directed growth are being explored as a means of *wiring* integrated circuits, e.g., guiding through via holes [47.41], by electric field [47.42], or surface modification [47.43].

Field-Emission Displays

The nanoscale diameter, aspect ratio of ≈ 1000, and high conductivity make carbon nanotubes ideal candidates for field emitters [47.44]. Field-emission currents from single nanotubes and aligned or randomly oriented nanotube thin films have been extensively studied. The promise is of low-threshold-field electron emission with a current density sufficient to drive a phosphor screen for display purposes. SWNTs and MWNTs have proved to be remarkably good field emitters with threshold emission fields in the range $1-10\,\mathrm{V}/\mu\mathrm{m}$ and capable of carrying very high current density. The exact emission mechanism, the role played by surface absorbents, and the cause of emission current saturation are under debate. Nevertheless, major manufacturers have recently produced prototype field-emission displays with carbon nanotube sources; with further development, this is likely to be the first major application of carbon nanotubes to reach the marketplace [47.45].

Electron-Beam Lithography

The present minimum feature size for silicon microelectronics is 130 nm, which is achieved using extreme-UV optical lithography. Alternative technologies are being explored for the future production of < 100 nm devices. Conventional electron-beam lithography can achieve a 5-nm line width but is limited by writing time because a single beam is used to write the entire pattern. Thus arrays of electron guns operating in parallel are being considered as a route to reduction of the writing time while maintaining the high resolution of electron-beam lithography. Carbon nanotubes are currently the most promising candidates for use as the emission source [47.46].

Electro-Optic Materials

Polymer–nanotube composite materials have been studied from the point of view of applications in electro-optics. Loading the layers of organic light-emitting diodes (LEDs) with low concentrations of nanotubes effectively increased the lifetime of the devices by preventing the build up of local hot spots thorough the high thermal conductivity that can be achieved in a nanotube percolation network [47.47]. Nanotube-induced local ordering of the matrix polymer suggests efficiency improvements may also be possible [47.48]. Using nanotube mats as an electrode in solar-cell applications apparently gave no improvement in device performance; however, these early measurement were performed with defective MWNTs so there may be a case for further investigation using the high-quality SWNTs available today.

Nanofabricators

Carbon-nanotube atomic force microscopy tips for fabricating oxide nanostructures in Si and Ti by anodisation is a rapidly expanding field aimed at Tbit/cm^2 data storage [47.49]. The technique utilises ambient moisture in the oxidation process with the tip biased negative relative to the surface. Line widths of 5–10 nm suitable for antidotes and tunnel junctions can be achieved.

Actuators

Electromechanical actuators based on SWNT sheets have been shown to generate higher stresses than natural muscle when operating in physiological conditions and higher strains than high-modulus ferroelectrics [47.50]. The actuation mechanism is a geometrical expansion of the carbon–carbon covalent bond caused by electrochemical double-layer charging. Work densities per cycle substantially higher than any previously known technology are predicted for SWNT sheets with mechanical properties close to those of individual SWNTs. This can be achieved by inter-tube binding and alignment optimisation.

Sensors

The conductivity of nanotube systems is highly sensitive to gaseous ambients, which affect the sign and amount of injected charge. The dimensions of a nanotube sensing element is such that very low quantities of analyte species will produce a measurable response. Nanotube gas sensors certainly have prospects to challenge conventional gas sensors for certain uses [47.19]. In addition to transduction, bio-sensors require a bioreceptor (e.g., enzyme or cell) immobilisation matrix. Carbon-nanotube-based bio-sensors meet both requirements and have been found to promote homogeneous electron-transfer reactions.

47.2 Bucky Materials

Over the last decade the field of bucky materials has been dominated by carbon nanotubes, as the number of publications and patent submissions testify. After intense research activity since their discovery, the most noteworthy of C_{60}-based electronic materials are superconducting heterofullerides of composition K_2MC_{60}, where M = Fe, Ni, Cu, Ag, Co ..., with $T_c \approx 15$ K. However the future for encapsulation of species in the C_{60}, and higher fullerene, cage appears promising, in particular as the basis of devices for quantum information processing. Other non-carbon-nanotube structures are now under intense investigation; these include BN, W, MoS_2, heterogeneous nanotubes and the ternary compound $Mo_6C_{9-x}H_x$ (C = chalcogen, H = halogen, $3 < x < 6$). The science of these structures is still in its infancy but, given the rich experience of over a decade of carbon nanotube research, there is reason to be optimistic of exciting new science and technology.

References

47.1 M. S. Dresselhaus, G. Dresselhaus, P. Avouris (eds): *Carbon Nanotubes: Synthesis, Structure, Properties, and Applications*, Topics Appl. Phys. Ser., Vol. 80 (Springer, Berlin Heidelberg New York 2000)

47.2 R. Saito, G. Dresselhaus, M.S. Dresselhaus: *Physical Properties of Carbon Nanotubes* (Imperial College Press, London 1998)

47.3 N. Wang, Z.K. Tang, G.D. Li, J.S. Chen: Nature **408**, 50 (2000)

47.4 L.M. Viculus, J.J. Mack, R.B. Kaner: Science **299**, 1361 (2003)

47.5 S. Ilani, P.L. McEuen: Annu. Rev. Condens. Matter Phys. **1**, 1 (2010)

47.6 K.I. Bolotin, K.J. Sikes, J. Hone, H.L. Stormer, P. Kim: Phys. Rev. Lett. **101**, 096802 (2008)

47.7 X. Liu, T. Pichler, M. Knupfer, M.S. Golden, J. Fink, H. Kataura, Y. Achiba, K. Hirahara, S. Iijima: Phys. Rev. B **65**, 45419 (2002)

47.8 G. Grüner: J. Mater. Chem. **16**, 3533 (2006)

47.9 L. Chico, V.H. Crespi, L.X. Benedict, S.G. Louie, M.L. Cohen: Phys. Rev. Lett. **76**, 971 (1996)

47.10 E.G. Rakov: Russ. Chem. Rev. **69**, 25 (2000)

47.11 A.C. Dillon, T. Genett, K.M. Jones, J.L. Alleman, P.A. Parilla, M.J. Heben: Adv. Mater. **11**, 1354 (1999)

47.12 K. Tohji, H. Takahashi, Y. Shinoda, N. Shimizu, B. Jeyadevan, I. Matuoka, Y. Sato, A. Kasuya, S. Ito, Y. Nishina: J. Phys. Chem. B **101**, 1974 (1997)

47.13 J.-M. Bonard, T. Stora, J.-P. Salvetat, F. Maier, T. Stoeckli, C. Duschul, L. Forro, W.A. de Heer, A. Chatelain: Adv. Mater. **9**, 827 (1997)

47.14 A. Huczko: Appl. Phys. A **74**, 617 (2002)

47.15 S.E. Baker, W. Cai, T.L. Lasseter, K.P. Weidkamp, R.J. Hamers: Nano. Lett. **2**, 1413 (2002)

47.16 R.J. Chen, Y. Zhang, D. Wang, H. Dai: J. Am. Chem. Soc. **123**, 3838 (2001)

47.17 G.U. Sumanasekera, C.K.W. Adu, S. Fang, P.C. Eklund: Phys. Rev. Lett. **85**, 1096 (2000)

47.18 P.R. Wallace: Phys. Rev. B **71**, 622 (1947)

47.19 J. Kong, N.R. Franklin, C. Zhou, M.G. Chapline, S. Peng, K. Cho, H. Dai: Science **287**, 5453 (2000)
47.20 J.W.G. Wildöer, L.C. Venema, A.G. Rinzler, R.E. Smalley, C. Dekker: Nature **391**, 59 (1998)
47.21 M. Ouyang, J.L. Huang, C.L. Cheung, C.M. Lieber: Science **292**, 702 (2001)
47.22 T.W. Ebbesen, H.J. Lezec, H. Hiura, J.W. Bennett, H.F. Ghaemi, T. Thio: Nature **382**, 54 (1996)
47.23 S.J. Tans, M.H. Devoret, H. Dai, A. Hess, R.E. Smalley, L.G. Geerlings, C. Dekker: Nature **386**, 474 (1997)
47.24 M. Bockrath, D.H. Cobden, P.L. McEuen, N.G. Chopra, A. Zettl, A. Thess, R.E. Smalley: Science **275**, 1922 (1997)
47.25 A. Bachtold, C. Strunk, J.-P. Salvetat, J.-M. Bonard, L. Farrro, T. Nussbaumer, C. Schonenberger: Nature **397**, 673 (1999)
47.26 S. Frank, P. Poncharal, Z.L. Wang, W.A. de Heer: Science **280**, 1744 (1998)
47.27 P.C. Collins, M.S. Arnold, P. Avouris: Science **292**, 1331 (2001)
47.28 W. Liang, M. Bockrath, D. Bozovic, J.H. Hafner, M. Tinkham, H. Park: Nature **411**, 665 (2001)
47.29 M. Bockrath, D.H. Cobden, L. Jia, A.L. Rinzler, R.E. Smalley, L. Balents, P.L. McEuen: Nature **397**, 598 (1999)
47.30 H.W. Postma, T. Teepen, Z. Yao, M. Grifoni, C. Dekker: Science **293**, 76 (2001)
47.31 T. Ihn, J. Güttinger, F. Molitor, S. Schnez, E. Schurtenberger, A. Jacobsen, S. Hellmüller, T. Frey, S. Dröscher, C. Stampfer, K. Ensslin: Mater. Today **13**, 44 (2010)
47.32 A.F. Morpurgo, J. Kong, C.M. Marcus, H. Dai: Science **286**, 263 (1999)
47.33 M. Kociak, A.Y. Kasumov, S. Gueron, B. Reulet, L. Vaccarrini, I.I. Khodos, Y.B. Gorbatov, V.T. Volkov, H. Bouchiat: Phys. Rev. Lett. **86**, 2416 (2001)
47.34 Z. Wang, W. Shi, H. Sie, T. Zhang, N. Wang, Z. Tang, X. Zhang, R. Lortz, P. Sheng, I. Sheikin: A. Demuer, Phys. Rev. B, **81**, 174530 (2010)
47.35 I. Takesue, J. Haruyama, N. Kobayashi, S. Chiashi, S. Maruyama, T. Sugai, H. Shinohara: Phys. Rev. Lett. **96**, 057001 (2006)
47.36 S.J. Kang, C. Kocabas, T. Ozel, M. Shim, N. Pimparkar, M.A. Alam, S.V. Rotkin, J.A. Rogers: Nat. Nanotechnol. **2**, 230 (2007)
47.37 S.J. Wind, J. Appenzaller, R. Martel, V. Derycke, P. Avouris: Appl. Phys. Lett. **80**, 3817 (2002)
47.38 A. Bachtold, P. Hadley, T. Nakanishi, C. Dekker: Science **294**, 1317 (2001)
47.39 P. Avouris, Z. Chen, V. Perebeinos: Nat. Nanotechnol. **2**, 605 (2007)
47.40 C.F. McAndrew, M. Baxendale: Nanotechnology **24**, 305202 (2013)
47.41 F. Kreupl, A.P. Graham, G.S. Duesberg, W. Steinhögl, M. Liebau, E. Unger, W. Hönlein: Microelectron. Eng. **64**, 399 (2002)
47.42 Y. Zhang, A. Chang, J. Cao, Q. Wang, W. Kim, Y. Li, N. Morris, E. Yenilmez, J. Kong, H. Dai: Appl. Phys. Lett. **79**, 3155 (2001)
47.43 J. Liu, M.J. Cavasant, M. Cox, D.A. Walters, P. Boul, L. Wei, A.J. Rimberg, K.A. Smith, D.T. Colbert, R.E. Smalley: Chem. Phys. Lett. **303**, 125 (1999)
47.44 J.-M. Bonard, M. Croci, C. Klinke, R. Kurt, O. Noury, N. Weiss: Carbon **40**, 1715 (2002)
47.45 N.S. Lee, D.S. Chung, I.T. Han, J.H. Kang, Y.S. Choi, H.Y. Kim, S.H. Park, Y.W. Jin, W.K. Yi, M.J. Yun: Diamond Rel. Mater. **10**, 265 (2001)
47.46 W.I. Milne, K.B.K. Teo, M. Chhowalla, G.A.J. Amaratunga, J. Yuan, J. Robertson, P. Legagneux, K. Pirio, K. Bouzehouane, D. Pribat, W. Bruenger, C. Trautmann: Curr. Appl. Phys. **1**, 317 (2001)
47.47 B. McCarthy, J.N. Coleman, R. Czerw, A.B. Dalton, M. in het Panhius, A. Maiti, A. Drury, P. Bernier, J.B. Nagy: J. Phys. Chem. B **106**, 2210 (2002)
47.48 H. Ago, K. Petritsch, M.S.P. Shaffer, A.H. Windle, R.H. Friend: Adv. Mater. **11**, 1281 (1999)
47.49 H. Dai, N. Franklin, J. Han: Appl. Phys. Lett. **73**, 1508 (1998)
47.50 R.H. Baughman, C. Cui, A.A. Zakhidov, Z. Iqbal, J.N. Barisci, G.M. Spinks, G.G. Wallace, A. Mazzoldi, D. De Rossi, A.G. Rinzler, O. Jaschinski, S. Roth, M. Kertesz: Science **284**, 1340 (1999)

48. Graphene

Henry H. Radamson

More than a decade has passed since the graphene was synthesized for the first time but still there are many applications that have not been explored. The first intended application for graphene was its high mobility property for high-speed electronics. This task became more difficult when the problem with zero bandgap of graphene made an obstacle to close the current flow of transistor. Different remedies, for example, downscaling the dimensions to nano-ribbons and bilayer structures were proposed to create a reasonable bandgap in graphene for electronic applications. The main idea was to improve the figure of merit I_{on}/I_{off} for RF devices. Although many important achievements were made, but later graphene was considered as a better choice for photonic applications. For example, graphene was used as light absorber in the lateral direction for modulators, or sensitive to light in photodetectors when it is vertically illuminated. Nowadays the graphene has inaugurated a new platform in flexible and transparent electronics. As example, electronics giant producers have already integrated graphene for flexible screens.

During recent years, many research efforts have been spent on other two-dimensional (2-D) crystals (transition metal dichalcogenides, silicene, and germanene) as competitors to graphene but it is still a long way to integrate such materials for industrial applications.

48.1 **Graphene Synthesis** 1174
48.2 **Band Structure and Electronic Applications** 1176
48.3 **Characterization of Graphene Material** 1178
48.3.1 AFM 1178
48.3.2 Transmission Electron Microscopy 1178
48.3.3 Raman Spectroscopy 1179
48.3.4 XRD Analysis 1180
48.4 **Photonic Applications** 1180
References 1182

Carbon is the most fascinating material in the universe which may be constructed into different forms and shapes. The electrical, optical, and mechanical properties of carbon depend on how the atoms are packed in relation to their adjacent atoms. As an example, diamond is a single crystalline form of carbon atoms with sp^3 hybridzation while graphene is a single-sheet arrangement of carbon atoms in a honeycomb shape with sp^2 hybridzation (Fig. 48.1a). These configurations place diamond in the group of wide bandgap materials while graphene is almost a zero bandgap material.

In nature, stacks of graphene planes can be found in graphite material and they can be ordered in the hexagonal or rhombohedral sequence. Graphene as a two-dimensional crystal material was experimentally detached from graphite by Geim and Novoselov in 2004. The discovery led to Nobel prize in 2010 with motivation of the synthesis of a 2-D carbon sheet with sp^2 bonds in comparison with sp^3, which is the most common form and is quantum mechanically favorable for bonding of carbon atoms [48.1].

In graphene, sp^2 hybrids establish σ bonds with the neighboring atoms (Fig. 48.1b). These bonds have a length of 0.142 nm and are the main reason behind the tremendous Young's modulus close to 1 TPa, which makes the strong mechanical property of graphene. This amazing Young modulus value is roughly 100 times larger than for steel with similar film thickness [48.1]. The remaining 2p orbital is half filled and it creates an extended π-bond where the electrons play the main role in electrical conduction and determine the photonic and electrical characteristics of graphene [48.2].

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_48

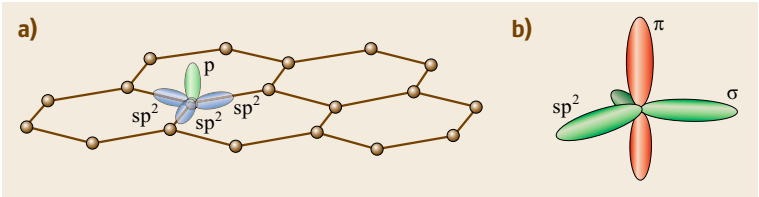


Fig. 48.1 **(a)** Schematic image of graphene with sp^2 hybrids of carbon atoms and **(b)** the formed π and σ bonds. (After [48.3])

Graphene is mostly known for its high electron mobility, which is $2.5\times10^5\,\mathrm{cm^{-2}V^{-1}s^{-1}}$ at room temperature. This high mobility value is a result of decreased electron–phonon interaction resulting in significantly lower carrier scattering. The electron mobility in graphene is almost 200 times higher than Si [48.4] and 4 times larger than III–V semiconductors [48.5]. This would make graphene a very attractive material for high-speed transistors.

Graphene has demonstrated excellent electron and thermal conduction. The electron saturation velocities up to $7\times10^7\,\mathrm{cm\,s^{-1}}$ have been observed, which is remarkably higher than high field saturation velocities for III–V and Si materials [48.6]. Similarly, current density close to $108\,\mathrm{A/cm^2}$ have been measured, which is almost five orders of magnitude larger than copper. The optical properties of graphene have also been widely studied. Graphene has optical opacity of 2.3% within a broad spectrum, which is wavelength independent in the interval of far-infrared and blue light [48.7].

Graphene absorbs more photons per unit thickness and surface. This means that graphene has larger efficiency per volume and bigger chances to saturate with high-intensity light pulses. Finally, graphene is a material with many outstanding properties, which is not seen altogether in one material alone. However, due to some undesired properties, there are still obstacles for graphene to replace materials for electronic technology in the present time. For example, graphene will not probably replace Si or III–V materials in high-frequency electronics in very near future. Meanwhile, graphene could be integrated in transparent electronics instead of today's most used material, indium-tin oxide (ITO) which is expensive and difficult to synthesize for mass production. Graphene has attracted attention for this purpose for electronics giant manufacturers such as Samsung and Sony [48.8].

More work is also required to decrease the graphene's parasitic contact in order to integrate for high frequency and flexible electronic applications. Recently, graphene has inaugurated totally a new domain within THz electronics, millimeter waves, and room-temperature ballistic electronics [48.9].

48.1 Graphene Synthesis

There are many ways to synthesize graphene (Fig. 48.2). The first article about isolating graphene sheet from graphite was published in 2004 by *Novoselov* and *Geim* [48.1]. Their method is the so-called exfoliation and uses a scotch-tape to attach flakes of graphene monolayers. The detached graphene can be transferred directly on Si (or BN) wafers or dissolved in acetone. The size of flakes is usually some micrometers and in the best case a few millimeters.

This method is simple, easy accessible, and flakes have relative low defect density. The high-quality graphene is the key for the high electron mobility and exfoliated graphene layers have demonstrated typical values of 2×10^4 and $2.5\times10^5\,\mathrm{cm^2\,V^{-1}s^{-1}}$ on SiO_2 and hexagonal-BN (h-BN) substrates at room temperature, respectively [48.10]. Similar mobility values have been demonstrated for suspended graphene [48.11] when the substrate interaction in this case is eliminated.

In many cases, the exfoliated graphene can be suspended in nonaqueous solvents or water-based surfactants [48.12]. The benefit of this method is using the surface tension of these solvents to make graphite to split in thinner platelets. Another approach is to oxidize graphite to form graphene oxide, which is easily soluble in water. Using ultrasound sonication may cause thining of the graphene stack to a monolayer. However, a total removal of all oxide is hard to achieve [48.13]. There are many reports about patterning reduced graphene oxide (RGO) and using laser scribing [48.14]. These materials can be used for supercapacitors, printed, and flexible electronics. However the complexity of such manufacturing techniques makes them limited to lab research and not suitable for any industrial application.

Another method to synthesize graphene is thermal decomposition of SiC. This method was demonstrated for the first time by *Berger* et al. [48.15] and it is based on the thermal annealing of SiC where sublimation of

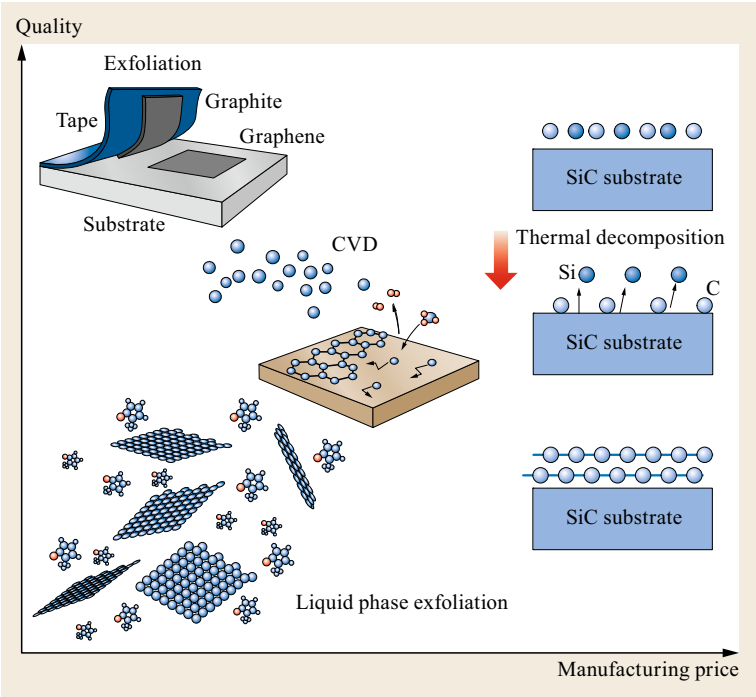


Fig. 48.2 A drawing of graphene quality versus manufacturing cost for different synthesis methods

Si occurs and the C atoms are segregated on Si- and the C-face of the wafer and form graphene layers. The process temperatures are around 1600 °C in 100 mbar for Si-face and 1450 °C and 0.1 μbar for C-face in Ar atmosphere [48.16]. As a result of Si desorption, the surface of SiC holds narrow terraces of graphene and they are connected by steps.

This method has demonstrated electron mobility values of around $3.0 \times 10^4\,\mathrm{cm^2V^{-1}s^{-1}}$ [48.17] depending on the alignment of SiC substrate terraces with the direction of carrier transport through graphene stack.

This method provides high quality graphene layers but has the drawbacks of the high cost of SiC substrates wafers and formation of small-sized graphene of a few micrometers which is too small for industrial applications to process a complete transistor.

The formed graphene stack on SiC demonstrates a bandgap of ≈ 260 meV. This value may vary depending on the stack thickness and it reaches almost zero for 4 graphene monolayers [48.18] due to substrate interaction or induced strain by SiC [48.19].

During recent years, chemical vapor deposition (CVD) of graphene has become very popular method to synthesize graphene. The method is based on the decomposition of a carbon precursor, for example, methane (CH_4) or ethanol (C_2H_6O) around 1000 °C on a foil of a transition metal, for example, Cu, Ni, Fe as catalyst material [48.20]. Later the formed graphene is separated or peeled off from surface of catalyst metal by using a transfer polymer (typically polymethylmethacrylate (PMMA) or polydimethylsiloxane (PDMS)) where the catalyst is etched away [48.21]. However, there is a high risk to damage the graphene sheet by the mechanical stress during the removal action and the quality of the graphene also depends on the surface roughness of catalyst metal. Different electronic component producers have stepped in using CVD method [48.22]. Nowadays very large area graphene is produced and developed for transparent electronics for displayer application. Using this method, carrier mobility of $0.5\,\mathrm{cm^2V^{-1}s^{-1}}$ has been demonstrated on SiO_2 surface.

48.2 Band Structure and Electronic Applications

Six carbon atoms are bonded in graphene honeycomb configuration with hexagonal lattice. Therefore, graphene can be considered as a triangular lattice consisting of a basis of two atoms per unit cell present at the two equivalent lattice sites A and B. A simple calculation gives the 2 lattice vectors, where $a = 0.142\,\text{nm}$ is the nearest-neighbor C-C spacing (Fig. 48.3a).

In principle, a honeycomb configuration has six points at the corners of the first Brillouin zone (FBZ) where they are distinguished in two groups of three equivalent points. These points are denoted by K and K′ (Fig. 48.3b). These corners are called Dirac points and the electrons and holes in the vicinity of those points are massless and follow linear relations (not the quadratic form) of energy momentum dispersion.

Graphene acts as a massless Dirac fermion and its electronic properties are described by the Dirac equation instead of the Schrodinger equation [48.23].

The graphene band structure has two notable properties: at first, at the Dirac point positions the valence and conduction band edges are very close (or intersecting), resulting in almost no energy gap (Fig. 48.3c). This places graphene in the category of zero bandgap semiconductors or as a semimetal. Secondly, as mentioned above the energy momentum dispersion has linear behavior and this leads to a constant carrier group velocity over energy (where is the Fermi velocity and estimated to $8\times10^5\,\text{ms}^{-1}$). As a result, the carriers have effective masses, which are directly related to momentum and have zero value at zero energy. This is an exceptional characteristic of graphene which is distinguished from common semiconductors, where the dispersion curves have parabolic shape and carrier velocity is dependent on the second derivative of the dispersion curves (Fig. 48.4).

Graphene has a linear density of states (DOS) as well. Each point q appears with twofold spin degenerate ($g_s = 2$) and since there are two Dirac points K and K′ hence another twofold valley degenerate ($g_v = 2$). Therefore DOS is written as [48.25]

$$\rho(E) = \frac{g_s g_v}{2\pi(\hbar v_F)^2}|E|\,, \tag{48.1}$$

where E stands for energy and v_F for Fermi velocity. This makes the DOS distribution of graphene very different from metals and semiconductors (Fig. 48.5). The integral of DOS equation times the Fermi–Dirac distribution at a given temperature results in a noticeable electron-sheet density for graphene. Furthermore, graphene sheet is not perfectly planar and has some ripples and this makes charge inhomogeneities in graphene.

Graphene possesses an ambipolar characteristic against an applied electric field. It means in contrary to metals, in the semimetal graphene, electrons may convert to holes in response to an applied electric field [48.26]. This means that n-type or p-type graphene is not predefined.

Graphene has extremely high carrier mobility. There are three macroscopic forces, which can affect the carrier mobility in graphene. These forces are diffusion, drift, and drag force. Diffusion force is an outcome of carriers' random motion, drift force is generated by electrostatic field and drag is the force due to the carrier scattering within the lattice [48.22]. In graphene, four scattering mechanisms have influence on the mobility. These mechanisms are addressed to the charged impurities, acoustic or optical phonons, and surface roughness. Experimentally, it has been demonstrated that the charged impurities in graphene have the crucial

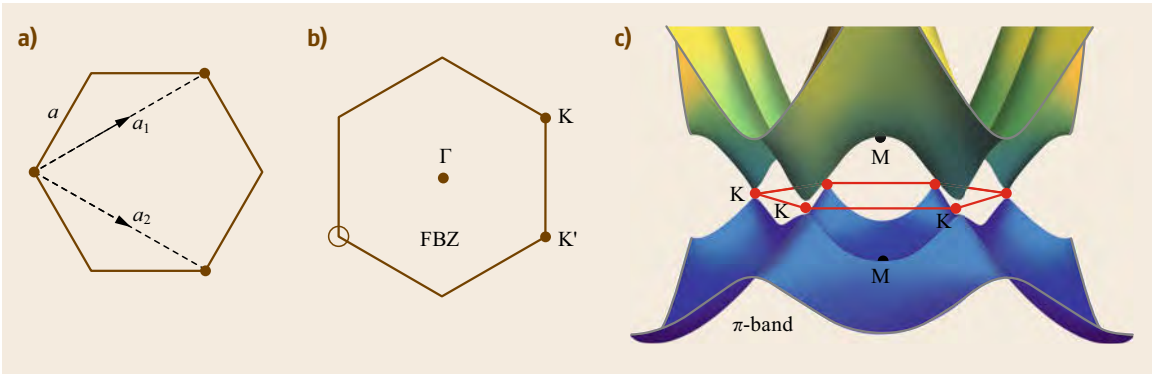


Fig. 48.3a–c A schematic view of graphene (a) honeycomb shape, (b) the first Brillouin zone, and (c) the band structure of graphene. (After [48.24])

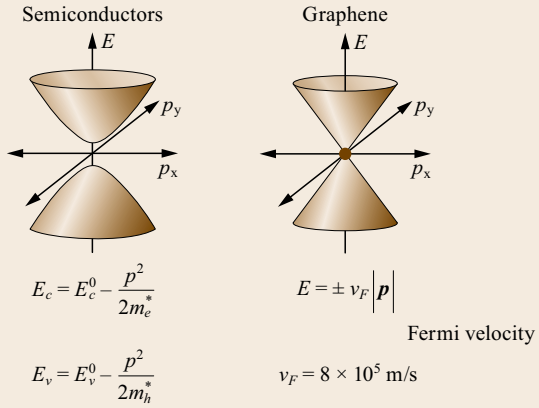


Fig. 48.4 The energy band diagram in semiconductors and graphene

effect on the carrier mobility. The charged impurities can be reduced in graphene by annealing or by using a high-k dielectric [48.27].

Application of graphene in high speed electronics has been shadowed for years since the lack of a bandgap makes it impossible to close the current flow. This blocks the application of graphene for logics since a bandgap around 400–500 meV is sought for this application. Integration of graphene as a channel material in field effect transistors (FETs) offers a high transconductance; however, the absence of a well-defined saturation region diminishes the benefit of a graphene power amplifier. A high I_{on}/I_{off} ratio is the key criteria for RF application.

One way to solve this problem is to decrease the lateral size of graphene to few a nanometers (so-called graphene nanoribbon (GNR)) where the electronic wavefunction could be localized and a quasi 1-D structure could be established (Fig. 48.6). In this way, a ballistic quantum transport is built up in graphene which is compatible with High Electron Mobility Transistor (HEMT) devices.

Another way to create a small bandgap in graphene is to form a bilayer graphene (BLG) [48.29] (or so-called Bernal stacked graphene), where the two graphene monolayers are placed on each other with a misalignment (Fig. 48.6).

Unfortunately Bernal BLG can mainly be synthesized by exfoliation method. Using BLG, a bandgap of 130 meV and the I_{on}/I_{off} value of ≈ 100 have been obtained [48.29].

Many research studies have proposed another alternative: vertical graphene transistor. This type of transistor functions by current tunneling from an electrode through a thin dielectric layer to graphene. In a more practical way, another graphene layer can be deposited to act as a second electrode. This extra gate allows triode modulation of the tunnel current [48.30]. This design has demonstrated an I_{on}/I_{off} ratio around 50 but no RF action has been shown so far.

An alternative to the above configuration is integration of graphene in a graphene hot-electron transistor,

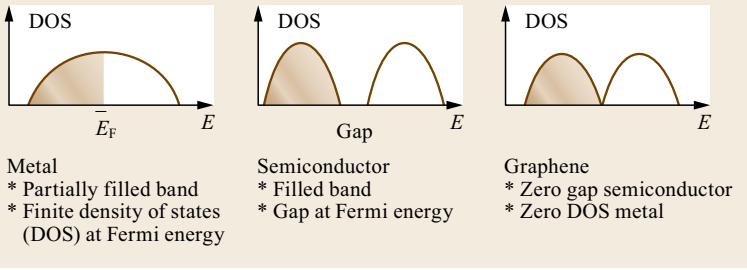


Fig. 48.5 Electronic structure and DOS for metals, semiconductors and graphene

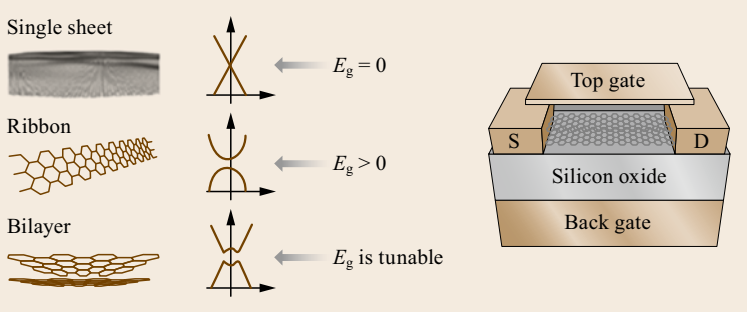


Fig. 48.6 Schematic of band diagram for different graphene material and GNR transistor. (After [48.28])

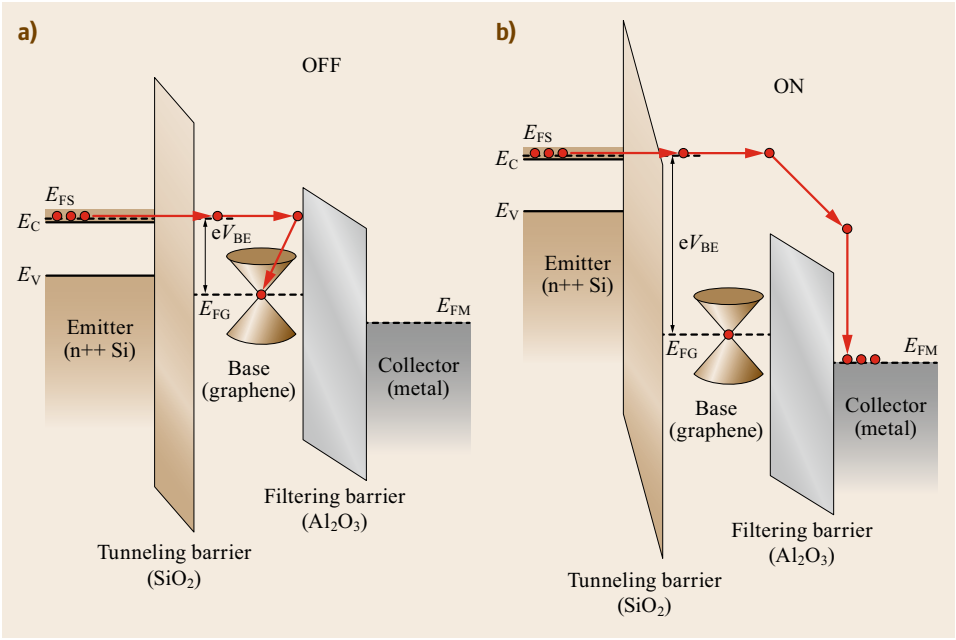


Fig. 48.7a,b Vertical graphene transistor **(a)** in off and **(b)** on state. (After [48.31])

which is the application of hot electrons in a metal-insulator-metal-insulator-metal (MI-M-I-M) transistor, an approach similar to Bipolar Junction Transistor (BJT). In this type of transistor, graphene is applied as low-resistivity base electrode and consists of an emitter base tunnel junction and a base collector filtering dielectric (Fig. 48.7a,b). Although these transistors have shown I_{on}/I_{off} ratio around 10^5 they have a very low current gain, which leads to collector currents 10 orders of magnitude less than the other types of transistors [48.32].

48.3 Characterization of Graphene Material

Graphene material is usually in a few monolayers on a specific substrate and it is difficult to be distinguished by optical microscopes. The characterization techniques, for example, atomic force microscopy (AFM), transmission electron microscopy (TEM), x-ray diffraction (XRD), and Raman spectroscopy are mainly applied to study graphene.

48.3.1 AFM

AFM is a common technique to estimate the thickness of graphene material. In general, 2-D crystallites such as graphene are a few nanometers thick elevated on the surface. Tapping mode is mostly used for the thickness measurements [48.33]. However, AFM technique has low throughput and therefore the measured data are not without error.

As an example, *Novoselov* et al. [48.33] have shown a monolayer graphene has 0.4 nm height, and Gupta et al. showed a height of 0.7 nm for single monolayer of graphene. One explanation for these different outcomes in measurements could be that these groups used the different amplitudes of the cantilever for their experiments.

This shows that extra care has to be taken to analyze the AFM results in order to have a comparable and reliable method to measure the thickness of graphene [48.33, 34].

48.3.2 Transmission Electron Microscopy

Transmission electron microscopy (TEM) is one of the most accurate methods to study graphene and provide atomic resolution (Fig. 48.8a–c). Graphene appears both in single- and multilayers in a flat sheet or folded form. The electron beam is parallel to the folded graphene sheets and they appear as one dark line for single sheets (Fig. 48.8a) and double dark lines for bilayer sheets (Fig. 48.8b). In the case of multiple folded sheets the number of dark lines could appear in any shape [48.35]. This can make it difficult to determine the number of graphene sheets.

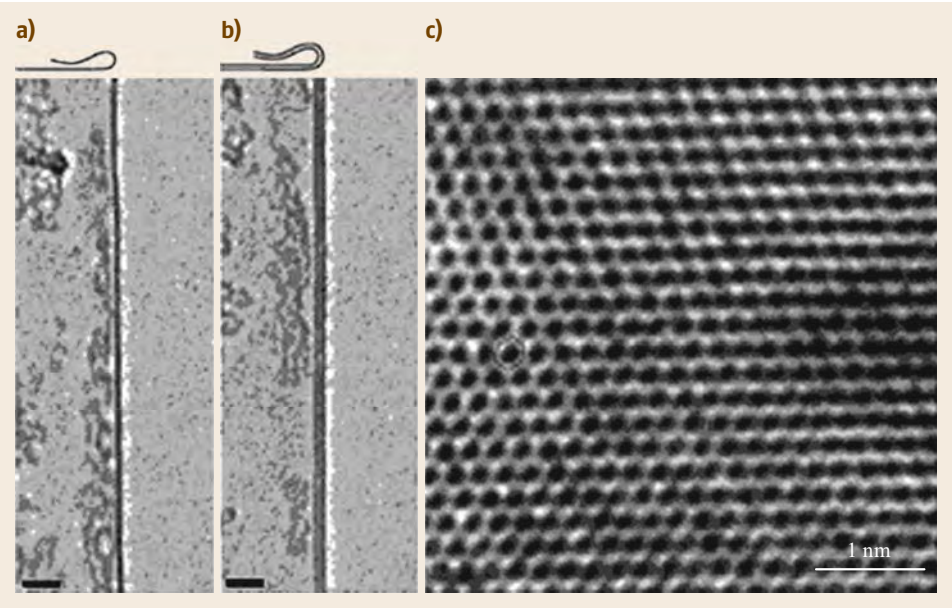


Fig. 48.8a–c TEM images: **(a)** cross-section of single-folded graphene, **(b)** cross-section of double-folded graphene and **(c)** top view of graphene material. (Reprinted from [48.36], with permission from Elsevier)

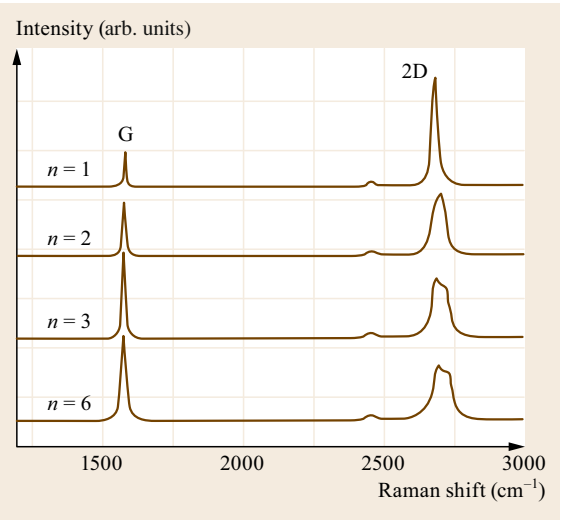


Fig. 48.9 Raman curves of graphene with n layers. (After [48.37])

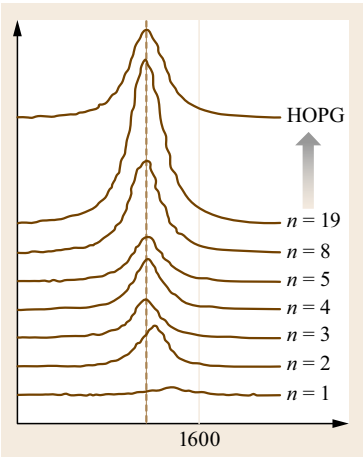


Fig. 48.10 G-band of graphene for a series of graphene layers n. (After [48.38])

48.3.3 Raman Spectroscopy

Raman spectroscopy is a technique that is based on inelastic interactions of phonons in a layer. For graphene, Raman can provide information about the number of monolayers and the quality of graphene material. There are three main bands in Raman analysis:

- The D band
- The 2-D band
- The G band.

The D Band

This band is an indicator of the defects in the graphene material and the presence of sp^3 bonds of carbon. The intensity of the D band is low for graphene (or graphite) with high quality. This peak normally appears at the edges of graphene flakes [48.39].

The 2-D Band

This band indicates the number of graphene sheets and also reveals if the graphene lies on a specific substrate. If the 2-D band is a single peak with narrow full-width-half-maximum (FWHM) then the graphene is a single sheet. However, if the graphene consists of multilayers then the shape of the peak is broad with small peaks. Figure 48.9 shows different cases for graphene

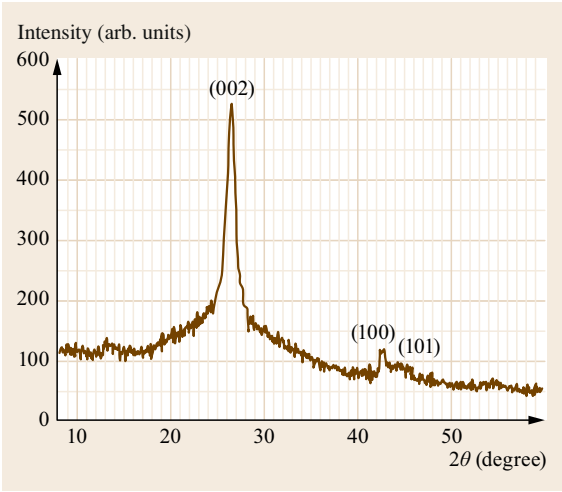


Fig. 48.11 X-ray curve for graphene. (After [48.40])

material. As an example, Raman results from a bilayer graphene shows peaks at wavelengths of 514 and 633 nm [48.41].

The G Band

The position of this band shows the number of graphene sheets. A peak is seen around $1587\,\mathrm{cm}^{-1}$ if the graphene consists of a single sheet, otherwise the G band moves downwards. Figure 48.10 shows the change of G band for graphene layers with different numbers of sheets [48.38].

48.3.4 XRD Analysis

X-ray diffraction is a technique that is usually used to study thin films. In the case of a very thin layer of graphene, a high-intensity x-ray beam is required. The XRD technique provides information about the crystalline quality and layer thickness of graphene. The latter parameter could be relatively determined by the intensity of the peaks. The XRD data of the graphene material gives three main peaks at reflections:

- (002) at 26.3°
- (101) 44.6°
- (100) 43.2° (Fig. 48.11) [48.40, 42].

48.4 Photonic Applications

Graphene has very promising optical properties for modulators [48.43, 44], photodetectors [48.45, 46], and plasmonic structures. Despite its thin thickness, it absorbs light in the vertical or lateral direction due to distinctive electronic structure [48.47]. The material absorbs a large variety of wavelengths within interband, intraband, and collective plasmon excitations. The extraordinary benefit of graphene is tuning the absorption by regulating the charge density.

The optical conductivity is written by a universal value of $\sigma_{\mathrm{uni}} = \pi e^2/2h$. For photons with higher energies, the optical conductivity is increased gradually towards peaks, which lie in UV range with energy of 4.62 eV (Fig. 48.12). This is related to nonlinear part of dispersion relation, which is labeled by interband transitions from the bonding to the antibonding π states in the graphene Brillouin zone [48.48].

Graphene absorbs ideally 2.3% of the normal incident light through intra- and interband transitions. This absorption may deviate from the ideal universal value of 2.3% depending on the underneath layer (Fig. 48.13a). This deviation is related to phonons (where interband absorption is allowed in the presence of phonons) [48.49].

Graphene can be placed on a waveguide to absorb light along the lateral direction (Fig. 48.13b). The length of graphene sheet may determine the amount of the absorption and for a long enough sheet a total absorption can be obtained. The absorption of light in the lateral direction occurs when a graphene sheet is placed on a waveguide. The length of graphene (or waveguide) sheet determines the amount of absorption and a complete absorption can be obtained if the length of graphene is long enough.

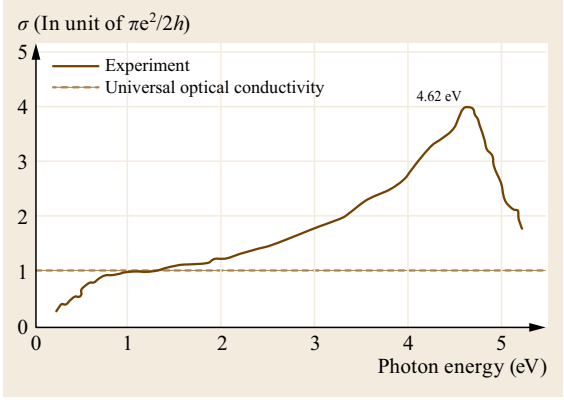


Fig. 48.12 Experimental optical conductivity (*solid line*) and the universal optical conductivity (*dashed line*) of a single graphene layer. The discrepancy of the optical data from the universal values at low energies is due to spontaneous doping. (After [48.48])

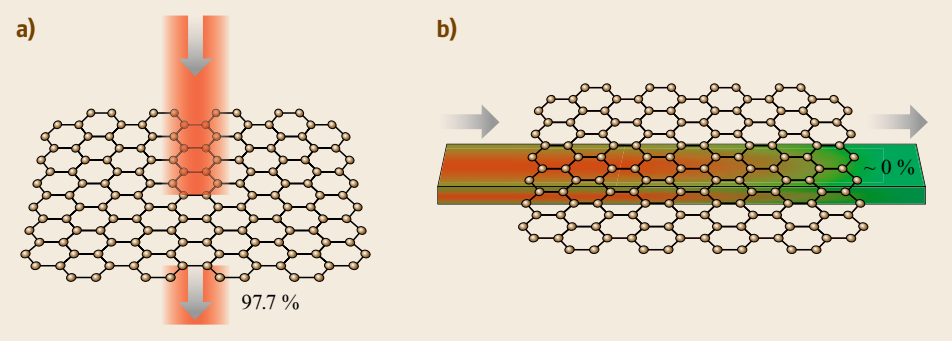


Fig. 48.13 **(a)** Normal incident light shines on a graphene layer. **(b)** Graphene on a waveguide and the light interacts with graphene along the waveguide sheet

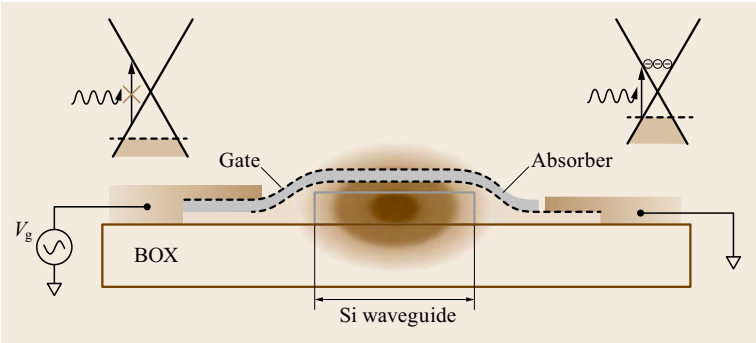


Fig. 48.14 A bilayer graphene modulator integrated on a waveguide. By applying a voltage to the gate, charge-sheet density in the first graphene layer is affected. (After [48.44])

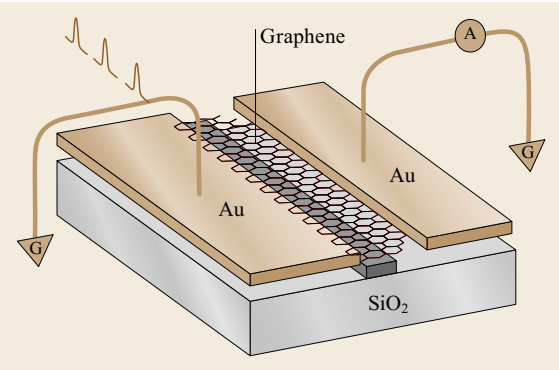


Fig. 48.15 A photodetector with graphene active layer. (After [48.46])

An innovative method to deal with graphene low DOS is to use two graphene sheets isolated by an oxide layer (Fig. 48.14). In this design, the first graphene sheet layer is for absorption, however, the second sheet operates as a gate.

Si is a material with a bandgap of 1.12 eV, and is transparent for $\lambda > 1.1\,\mu m$. Therefore, Si can be used as a waveguide to conduct telecom wavelengths.

The modulator in Fig. 48.14 may function when the first graphene sheet is grounded and the second one is biased with a gate voltage, V_g [48.44].

Graphene has also been used as active photonic material in photodetectors. An electrical current was generated when the contact electrode of graphene transistor was illuminated [48.50, 51] (Fig. 48.15). In order to increase the absorption, the detector can be formed on a waveguide processed on an SOI substrate with a high-k material beneath the graphene sheet.

Although graphene can absorb a broad wavelength a graphene detector has a low responsivity of 0.1 mA/W, which is remarkably lower than III-V detectors.

During recent years, many research efforts have been spent on other 2-D crystals (transition metal dichalcogenides, silicene, and germanene) as competitors to graphene [48.52].

The transition metal dichalcogenides are a group of materials which are known as MX_2, where M is a transition metal element from group IV (e.g., Ti, Zr or Hf), group V (e.g., V, Nb, or Ta), or group VI (e.g., Mo or W), and X is a chalcogen material (e.g., S, Se, or Te).

Even though, there are many reports about the synthesis and properties of these materials there is a long way to go before they can be mass-produced for industrial applications in near future [48.52].

References

48.1 K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, Y. Zhang, S.V. Dubonos, I.V. Grigorieva, A.A. Firsov: Electric field effect in atomically thin carbon films, Science **306**(5696), 666 (2004)

48.2 S. Kim, J. Ihm, H.J. Choi, Y.W. Son: Origins of anomalous electronic structures of epitaxial graphene on silicon carbide, Phys. Rev. Lett. **100**(17), 176802 (2008)

48.3 M.C. Lemme: Current status of graphene transistors, Solid State Phenom. **156–158**, 499 (2009)

48.4 Y.-C. Lin, C. Jin, J.-C. Lee, S.-F. Jen, K. Suenaga, P.-W. Chiu: Clean transfer of graphene for isolation and suspension, ACS Nano **5**, 2362 (2011)

48.5 M.E. Mills, P. Townsend, D. Castillo, S. Martin, A. Achen: Benzocyclobutene (DVSBCB) polymer as an interlayer dielectric (ILD) material, Microelectron. Eng. **33**, 327 (1997)

48.6 Y. Lin, D. Farmer, K. Jenkins: Enhanced performance in epitaxial graphene FETs with optimized channel morphology, Electron Device **32**, 1343 (2011)

48.7 R.R. Nair, P. Blake, A.N. Grigorenko, K.S. Novoselov, T.J. Booth, T. Stauber, N.M.R. Peres, A.K. Geim: Fine structure constant de_nes visual transparency of graphene, Science **320**, 1308 (2008)

48.8 T. Kobayashi, M. Bando, N. Kimura, K. Shimizu, K. Kadono, N. Umezu, K. Miyahara, S. Hayazaki, S. Nagai, Y. Mizuguchi, Y. Murakami, D. Hobara: Production of a 100-m-long high-quality graphene transparent conductive film by roll-to-roll chemical vapor deposition and transfer process, Appl. Phys. Lett. **102**, 023 (2013)

48.9 G. Moddel, Z. Zhu, S. Grover, S. Joshi: Ultrahigh speed graphene diode with reversible polarity, Solid State Commun. **152**, 1842 (2012)

48.10 A.S. Mayorov, R.V. Gorbachev, S.V. Morozov, L. Britnell, R. Jalil, L.A. Ponomarenko, P. Blake, K.S. Novoselov, K. Watanabe, T. Taniguchi, A.K. Geim: Micrometerscale ballistic transport in encapsulated graphene at room temperature, Nano Lett. **11**, 2396 (2011)

48.11 K. Bolotin, K. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim, H. Stormer: Ultrahigh electron mobility in suspended graphene, Solid State Commun. **146**, 351 (2008)

48.12 A.A. Green, M.C. Hersam: Solution phase production of graphene with controlled thickness via density differentiation, Nano Lett. **9**, 4031 (2009)

48.13 G. Eda, G. Fanchini, M. Chhowalla: Large-area ultrathin films of reduced graphene oxide as a transparent and exible electronic material, Nat. Nanotechnol. **3**, 270 (2008)

48.14 M.F. El-Kady, V. Strong, S. Dubin, R.B. Kaner: Laser scribing of high-performance and flexible graphene-based electrochemical capacitors, Science **335**, 1326 (2012)

48.15 C. Berger, Z. Song, X. Li, X. Wu, N. Brown, C. Naud, D. Mayou, T. Li, J. Hass, A.N. Marchenkov, E.H. Conrad, P.N. First, W.A. de Heer: Electronic confinement and coherence in patterned epitaxial graphene, Science **312**, 1191 (2006)

48.16 J.A. Robinson, M. Wetherington, J.L. Tedesco, P.M. Campbell, X. Weng, J. Stitt, M.A. Fanton, E. Frantz, D. Snyder, B.L. VanMil, G.G. Jernigan, R.L. Myers-Ward, C.R. Eddy, D.K. Gaskill: Correlating Raman spectral signatures with carrier mobility in epitaxial graphene: A guide to achieving high mobility on the wafer scale, Nano Lett. **9**, 2873 (2009)

48.17 Z. Guo, R. Dong, P.S. Chakraborty, N. Lourenco, J. Palmer, Y. Hu, M. Ruan, J. Hankinson, J. Kunc, J.D. Cressler, C. Berger, W.A. de Heer: Record maximum oscillation frequency in C-face epitaxial graphene transistors, Nano Lett. **13**, 942 (2013)

48.18 S.Y. Zhou, G.-H. Gweon, A.V. Fedorov, P.N. First, W.A. de Heer, D.-H. Lee, F. Guinea, A.H. Castro Neto, A. Lanzara: Substrate-induced bandgap opening in epitaxial graphene, Nat. Mater. **6**, 770 (2007)

48.19 N. Ferralis, R. Maboudian, C. Carraro: Evidence of structural strain in epitaxial graphene layers on 6H-SiC(0001), Phys. Rev. Lett. **101**, 156801 (2008)

48.20 X. Li, W. Cai, J. An, S. Kim, J. Nah, D. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc, S.K. Banerjee, L. Colombo, R.S. Ruo: Large-area synthesis of high-quality and uniform graphene films on copper foils, Science **324**, 1312 (2009)

48.21 P.J. Ko, H. Takahashi, S. Koide, H. Sakai, T.V. Thu, H. Okada, A. Sandhu: Simple method to transfer graphene from metallic catalytic substrates to flexible surfaces without chemical etching, J. Phys. Conf. Ser. **433**(12), 002 (2013)

48.22 S. Bae, H. Kim, Y. Lee, X. Xu, J.-S. Park, Y. Zheng, J. Balakrishnan, T. Lei, H.R. Kim, Y.I. Song, Y.-J. Kim, K.S. Kim, B. Ozyilmaz, J.-H. Ahn, B.H. Hong, S. Iijima: Roll-to-roll production of 30-inch graphene films for transparent electrodes, Nat. Nanotechnol. **5**, 574 (2010)

48.23 A. Fasolino, J.H. Los, M.I. Katsnelson: Intrinsic ripples in graphene, Nat. Mater. **6**, 858 (2007)

48.24 M. Wilson: Electron band structure in graphene. Near K and K', the dispersion relation is relativistic with an effective mass being zero, Phys. Today **59**(1), 21 (2006)

48.25 T. Fang, A. Konar, H. Xing, D. Jena: Carrier statistics and quantum capacitance of graphene sheets and ribbons, Appl. Phys. Lett. **91**, 092 (2007)

48.26 J.A. Robinson, M. Wetherington, J.L. Tedesco, P.M. Campbell, X. Weng, J. Stitt, M.A. Fanton, E. Frantz, D. Snyder, B.L. VanMil, G.G. Jernigan, R.L. Myers-Ward, C.R. Eddy, D.K. Gaskill: Correlating Raman spectral signatures with carrier mobility in epitaxial graphene: A guide to achieving high mobility on the wafer scale, Nano Lett. **9**, 2873 (2009)

48.27 K.D. Holland, N. Paydavosi, N. Neophytou, D. Kienle, M. Vaidyanathan: RF performance limits and operating physics arising from the lack of a bandgap in graphene transistors, IEEE Trans. Nanotechnol. **12**, 566 (2013)

48.28 F. Xia, D.B. Farmer, Y.-M. Lin, P. Avouris: Graphene field-effect transistors with high on/off current ratio and large transport band gap at room temperature, Nano Lett. **10**, 715 (2010)

48.29 K. Yan, H. Peng, Y. Zhou, H. Li, Z. Liu: Formation of bilayer bernal graphene: Layer by-layer epitaxy via chemical vapor deposition, Nano Lett. **11**, 1106 (2011)
48.30 L. Britnell, R.V. Gorbachev, R. Jalil, B.D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M.I. Katsnelson, L. Eaves, S.V. Morozov, N.M.R. Peres, J. Leist, A.K. Geim, K.S. Novoselov, L.A. Ponomarenko: Field effect tunneling transistor based on vertical graphene heterostructures, Science **335**, 947 (2012)
48.31 C. Zeng, E.B. Song, M. Wang, S. Lee, C.M. Torres Jr., J. Tang, B.H. Weiller, K.L. Wang: Vertical graphene-base hot-electron transistor, Nano Lett. **13**, 2370 (2013)
48.32 W. Mehr, J. Dabrowski, J.C. Scheytt, G. Lippert, Y.-H. Xie, M.C. Lemme, M. Ostling, G. Lupina: Vertical graphene base transistor, IEEE Electron Dev. Lett. **33**, 691 (2012)
48.33 K.S. Novoselov, D. Jiang, F. Schedin, T.J. Booth, V.V. Khotkevich, S.V. Morozov, A.K. Geim: Two-dimensional atomic crystals, Proc. Natl. Acad. Sci. **102**(30), 10451 (2005)
48.34 P. Nemes-Incze, Z. Osva'th, K. Kamara's: *L.P. Biro': Anomalies in thickness measurements of graphene and few layer graphite crystals by tapping mode atomic force microscopy* (Carbon 46(11), 1435 2008)
48.35 J.C. Meyer, A.K. Geim, M.I. Katsnelson, K.S. Novoselov, T.J. Booth, S. Roth: The structure of suspended graphene sheets, Nature **446**(7131), 60 (2007)
48.36 J.L. Qi, W.T. Zheng, X.H. Zheng, X. Wang, H.W. Tian: Relatively low temperature synthesis of graphene by radio frequency plasma enhanced chemical vapor deposition, Appl. Surf. Sci. **257**(15), 6531 (2011)
48.37 C. Casiraghi, A. Hartschuh, E. Lidorikis, H. Qian, H. Harutyunyan, T. Gokus, K.S. Novoselov, A.C. Ferrari, H. Harutyuyan: Rayleigh imaging of graphene and graphene layers, Nano Lett. **7**(9), 2711 (2007)
48.38 A. Gupta, G. Chen, P. Joshi, S. Tadigadapa, P.C. Eklund: Raman scattering from high-frequency phonons in supported n-graphene layer films, Nano Lett. **6**(12), 2667 (2006)
48.39 M. Wall: Raman spectroscopy optimizes graphene characterization, Adv. Mater. Process **170**, 35 (2012)
48.40 G. Wang, J. Yang, J. Park, X. Gou, B. Wang, H. Liu, J. Yao: Facile synthesis and characterization of graphene nanosheets, J. Phys. Chem. C **112**(22), 8192 (2008)
48.41 A.C. Ferrari, J.C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K.S. Novoselov, S. Roth, A.K. Geim: Raman spectrum of graphene and graphene layers, Phys. Rev. Lett. **97**(18), 187401 (2006)
48.42 A. Chakrabarti, J. Lu, J.C. Skrabutenas, T. Xu, Z. Xiao, J.A. Maguire, N.S. Hosmane: Conversion of carbon dioxide to fewlayer graphene, J. Mater. Chem. **21**(26), 9491 (2011)
48.43 M. Liu, X. Yin, E. Ulin-Avila, B. Geng, T. Zentgraf, L. Ju, F. Wang, X. Zhang: A graphene-based broadband, optical modulator, Nature **474**, 64 (2011)
48.44 S.J. Koester, M. Li: High-speed waveguide-coupled graphene-on-graphene optical modulators, Appl. Phys. Lett. **100**, 171107 (2012)
48.45 A. Pospischil, M. Humer, M.M. Furchi, D. Bachmann, R. Guider, T. Fromherz, T. Mueller: CMOS-compatible graphene photodetector, Nat. Photonics **7**, 892 (2013)
48.46 X. Gan, R.-J. Shiue, Y. Gao, I. Meric, T.F. Heinz, K. Shepard, J. Hone, S. Assefa, D. Englund: Chip-integrated ultrafast graphene photodetector with high responsivity, Nat. Photonics **7**, 883 (2013)
48.47 R.R. Nair, P. Blake, A.N. Grigorenko, K.S. Novoselov, T.J. Booth, T. Stauber, N.M.R. Peres, A.K. Geim: Fine structure constant defines visual transparency of graphene, Science **320**(5881), 1308 (2008)
48.48 K.F. Mak, J. Shan, T.F. Heinz: Seeing many-body effects in single- and few-layer graphene: Observation of two-dimensional saddle-point excitons, Phys. Rev. Lett. **106**, 046401 (2011)
48.49 B. Scharf, V. Perebeinos, J. Fabian, P. Avouris: Effects of optical and surface polar phonons on the optical conductivity of doped graphene, Phys. Rev. B **87**, 035414 (2013)
48.50 E.J.H. Lee, K. Balasubramanian, R.T. Weitz, M. Burghard, K. Kern: Contact and edge effects in graphene devices, Nat. Nanotech. **3**, 486 (2008)
48.51 F. Xia, T. Mueller, R. Golizadeh-Mojarad, M. Freitag, Y.-M. Lin, J. Tsang, V. Perebeinos, P. Avouris: Photocurrent imaging and efficient photon detection in a graphene transistor, Nano Lett. **9**, 1039 (2009)
48.52 S. Chang, L.T.L. Lee, T. Chen: Plasmon-enhanced excitonic solar cells. In: *High-Efficiency Solar Cells: Physics, Materials and Devices*, ed. by X. Wang, Z.M. Wang (Springer, Cham 2014) p. 515

49. Magnetic Information-Storage Materials

Charbel Tannous, R. Lawrence Comstock†

The purpose of this chapter is to review the current status of magnetic materials used in data storage. The emphasis is on magnetic materials used in disk drives and in the magnetic random-access memory (MRAM) technology. A wide range of magnetic materials is essential for the advance of magnetic recording both for heads and media, including high-magnetization soft-magnetic materials for write heads, antiferromagnetic alloys with high blocking temperatures and low corrosion propensity for pinning films in giant-magnetoresistive (GMR) sensors and ferromagnetic alloys with large values of giant magnetoresistance. For magnetic recording media, the advances are in high-magnetization metal alloys with large values of switching coercivity. A significant limitation to magnetic recording is found to be the superparamagnetic effect and advances have been made in multilayer ferromagnetic films to reduce the impact of the effect, but also to allow high-density recording have been developed. Perpendicular recording as compared to longitudinal recording is reviewed and it is shown that this technology will soon be replaced first by heat-assisted and later by bit-patterned magnetic recording in order to progress steadily toward areal densities well above 10^{12} bit/in^2 (1 Tbit/in^2 or 1000 Gbit/in^2). While an MRAM cell exploits some of the materials used in GMR sensors, its basic component is the magnetic tunneling junction in which magnetic films are coupled by a thin insulating film and conduction occurs by quantum mechanical tunneling. The status of MRAM cell technology and some closely related key problems are reviewed.

49.1 **Magnetic Recording Technology** 1186
49.1.1 Magnetic Thin Films 1187
49.1.2 The Write Head 1189
49.1.3 Spin-Valve Read Head 1192
49.1.4 Longitudinal Recording Media (LMR) ... 1199
49.1.5 Perpendicular Magnetic Recording 1205
49.2 **Magnetic Random-Access Memory** 1215
49.2.1 Tunneling Magnetoresistive Heads 1218
49.3 **Extraordinary Magnetoresistance (EMR)** 1220
49.4 **Summary** 1220
References 1220

The purpose of this chapter is to summarize the status of magnetic materials used in high-capacity disk drives and magnetic-semiconductor memory devices. The technology of disk drives is selected since these devices have experienced the largest increase in data capacity over time and this has made disk drives the preeminent storage system for digital data. To illustrate this point, consider Fig. 49.1, which is a plot of the areal density (number of data tracks per inch times the number of bits per inch recorded on a track) for disk drives over time [49.1].

The increase in areal density is more than 100% per year up to about 2002, when it reduced to about 20–30%, then increased steeply by 600% between 2006 and now evolving from 150–900 Gbit/in^2, the current value at the time of writing. Practically it means that we are about to reach 1 Tbit/in^2, a value that is considered as a formal limit to (perpendicular) recording and requires a paradigm shift in order to keep further increasing areal density as explained below.

The total data capacity of a disk is approximately the areal density times the recording area, which is twice the number of platters since both faces are used and depends on the size of the disk (2.5 and 3.5 in (64 and 90 mm)) diameter being the most common.

For a 3.5 in disk, a simple rule of thumb implies that 900 Gbit/in^2 areal density means about 900 Gbyte storage per platter face.

Many technologies have contributed to this rapid increase in areal density, including advances in the technology of flying heads with reduced spacing to the disk surface, data codes and error detection and cor-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_49

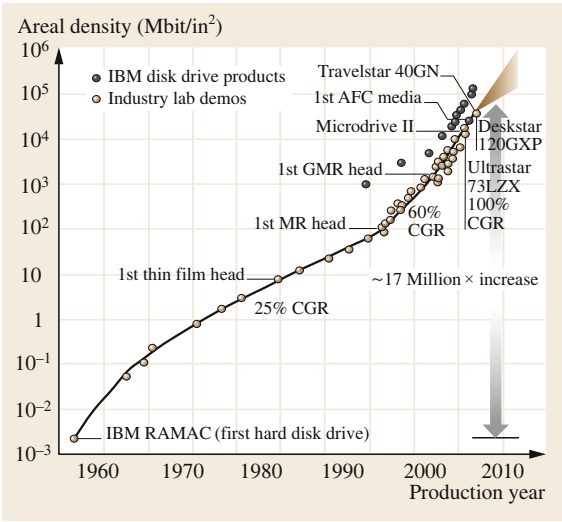


Fig. 49.1 Historical variation of areal density from an IBM perspective. (After [49.1])

rection, advanced servo-control systems for accurate control of magnetic recording heads on data tracks, and improvements in the mechanical structures comprising a disk drive, including advances in motors used to drive the disks. However, this paper discusses only the fundamental technology associated with digital magnetic recording, including the devices used to record and read back the recorded data and the media on which the data is recorded. The discussion is also restricted to the materials and not to any of the mechanical structures associated with the recording heads or disks.

The discussion covers longitudinal magnetic recording (LMR), perpendicular magnetic recording (PMR) and the transition process between them allowing an increase in areal densities in hard disks. In PMR the remanent magnetization is perpendicular to the disk surface reducing the impact of the superparamagnetic limit. Technology underlying MRAM (magnetic random-access memory) is also discussed since it provides nonvolatility to RAM, blurring the borderline separating permanent storage and RAM.

Consequently, the operating system and application programs will always be present in RAM since the computer's very first boot, in sharp contrast with respect to the electronic RAM whose contents totally disappear every time the computer is switched off or when it is jammed and not responding.

MRAM is a possible replacement for the familiar semiconductor memories used in microcomputers – dynamic and static random-access memory (DRAM modules are placed on the computer motherboard close to central processing unit (CPU) whereas static random-access memory (SRAM) is used inside CPU cache memories).

MRAM technology combines a magnetic storage technology together with metal-oxide semiconductor (MOS) devices to result in fast and high-density data memory devices. The technology on which the magnetic part of MRAM is based is an extension of the technology used in magnetic recording devices – the magnetic tunneling junction (MTJ). MRAM technology will also be discussed in this chapter. Parts of this work have been previously published in the Journal of Materials Science: Materials in Electronics [49.2].

49.1 Magnetic Recording Technology

The technology of magnetic recording was one hundred years old [49.3] in 1999. The fundamental concept of magnetic recording is to use a magnetic structure (the write head) driven by current that represents the data to be recorded to generate a magnetic field that can change the state of the magnetization in a closely spaced magnetic recording medium, which in the earliest realization was magnetic wire, and today is either the familiar magnetic tape or a magnetic layer on a rigid disk substrate. The data are recovered by the generation of an output voltage in the read head by sensing the magnetization in the recording medium, e.g., by Faraday's law ($V = N\mathrm{d}\phi/\mathrm{d}t$), where N is the number of turns on the read head and ϕ is the magnetic flux coupled to the read head from the media. The magnetic recording system to be discussed here is that used to store digital data, in which case the current supplied to the write head is in the form of pulses encoded to represent the digital data [49.4, 5] (1 or 0).

In the case of disk drives the write and read heads are separate thin-film structures deposited on the back of a mechanical slider, which uses a hydrodynamic air bearing to fly over the surface of the disk [49.5]. Figure 49.2 is a schematic of a digital LMR system. The recording (write) and read elements are shown together with the magnetic recording surface, which in disk drive technology is a thin metallic film of a cobalt alloy (to be discussed). The digital data are recorded in the magnetic film as transitions between the two possible states of the magnetization (pointing to the left or right) and with the width approximately equal to the width of the write head and the width of a data track. The transi-

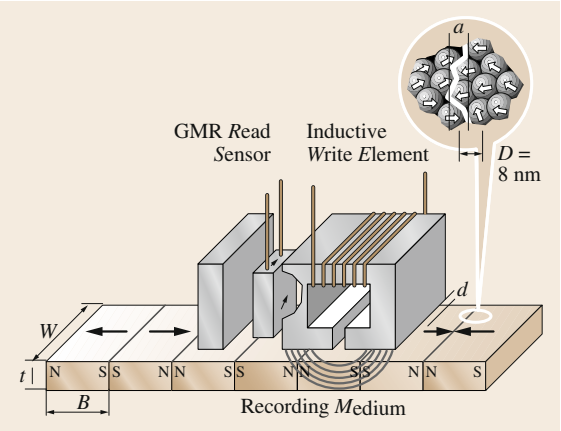


Fig. 49.2 Schematic illustration of a longitudinal magnetic GMR read sensor recording system showing a giant magnetoresistive (GMR) read sensor, write element and the recording medium. The parameters are: t the recording medium thickness (δ is also used), W the width of the recorded track, B the length of the recorded magnetization or bit length and d the head-media separation (HMS) or magnetic spacing (includes overcoats on the disk surface and undercoats on the slider surface). Shown in the *inset*: the transition region between the two directions of the magnetization; D is the media grain size and a is the transition parameter that characterizes the length of the transition region (length $l = \pi a$). (After [49.1])

tion region between the oppositely directed directions of the magnetization is similar to that between magnetic domains and has a length l. We will discuss the different parts of the recording system starting with the write head and including the read head, which in Fig. 49.2 is a giant-magnetoresistive sensor. First, we will discuss some aspects of thin magnetic films that are relevant to magnetic recording components. All magnetic recording components used in disk drives today are fabricated from thin films to allow mass production and for technology reasons, as we will discuss.

49.1.1 Magnetic Thin Films

Induced Uniaxial Anisotropy

In most applications of soft-magnetic films, a uniaxial anisotropy in the plane of the film is required. One way of introducing this uniaxial anisotropy is to induce it by applying a magnetic field in the plane of the film and the effect is referred to as magnetic annealing. The anisotropy can be induced during deposition of the film or induced by a subsequent annealing step, using temperature and a magnetic field to modify the anisotropy. The mechanism of the induced anisotropy is that of directional order, in which atomic pairs in an alloy tend to align with the local magnetization. In nickel-iron films the atomic pairs are the iron atoms and the induced uniaxial anisotropy energy density is typically $K_u \approx 1-3\,\mathrm{kerg/cm^3}$ ($0.1-0.3\,\mathrm{kJ/m^3}$), where the uniaxial energy density is

$$E_k = K_u \sin^2\theta\,, \tag{49.1}$$

where θ is the angle of the magnetization with respect to the direction of the induced anisotropy. To induce the anisotropy it is necessary to saturate the magnetization of the film with a small magnetic field, typically $50-100\,\mathrm{Oe}$ ($4000-8000\,\mathrm{A/m}$), since it is the magnetization not the magnetic field that is responsible for the magnetic annealing. If the anisotropy energy density is positive the energy is minimum is along the direction of the anisotropy, which is referred to as an *easy* axis.

A second method of introducing uniaxial anisotropy in thin films is by stress. If the magnetostriction constant is isotropic with value λ_s for a polycrystalline material then the magnetoelastic energy density is given by

$$E_{me} = \frac{3}{2}\lambda_s\sigma \sin^2\theta\,. \tag{49.2}$$

The angle θ is the angle of the magnetization in the film plane with respect to the direction of the stress (σ). If a sample is unmagnetized it will strain by λ_s as the sample is magnetized to saturation. A typical value of λ_s is $\approx 10^{-6}$. If the magnetostriction constant and the stress are both positive the direction for the minimum energy and hence the easy axis is for $\theta = 90°$. The magnetization, permeability, crystalline anisotropy constant and magnetostriction constant of nickel-iron alloys [49.6] $Ni_xFe_{(1-x)}$ are shown in Fig. 49.3. The maximum permeability is for the Permalloy composition $Ni_{80}Fe_{20}$. Increasing the atomic percentage of iron over that of Permalloy results in increased magnetization since the magnetic moment for iron is $2.2\,\mu_B$ (Bohr magnetons) per atom in the metal, while the value for nickel is $0.6\,\mu_B$ per atom in the metal.

Hysteresis in Soft-Magnetic Films

If we impose an external magnetic field H_e along the direction of the induced anisotropy (easy axis) of the thin film ($\theta = 0°$) the total energy density is given by

$$E_k = K_{eff} \sin^2\theta - MH_e \cos\theta\,. \tag{49.3}$$

K_{eff} is the sum of the uniaxial crystalline anisotropy (K_u) and the stress anisotropy $(3/2)\lambda_s\sigma$. The second term is the magnetostatic energy. The equilibrium angle

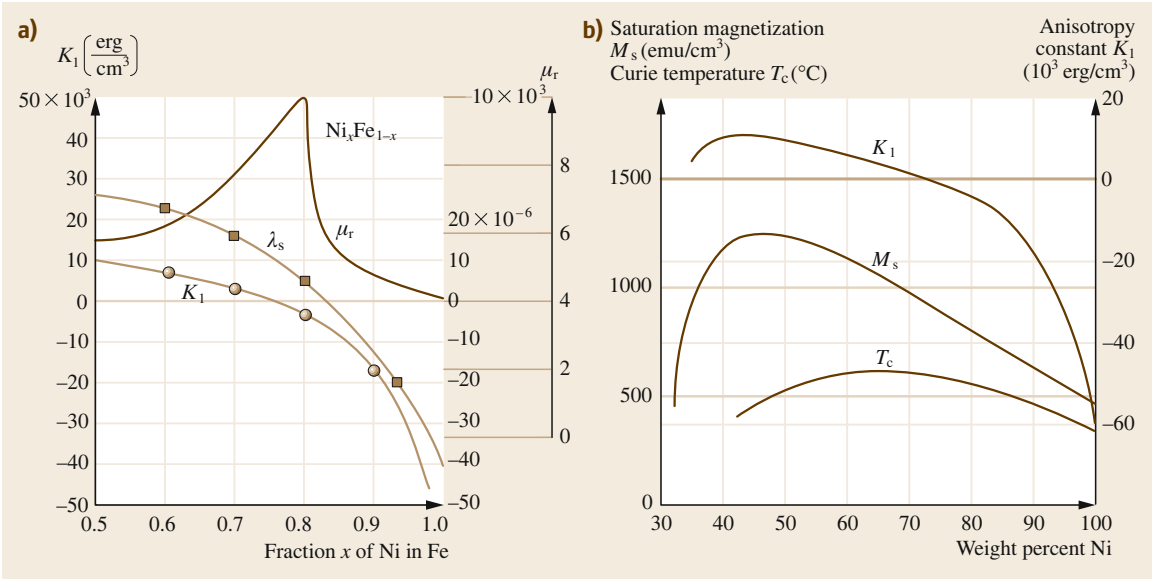


Fig. 49.3 **(a)** Magnetic properties of Ni_xFe_{1-x} alloys, including initial permeability, crystalline anisotropy and magnetostriction. **(b)** Magnetization, Curie temperature and crystalline anisotropy of Ni_xFe_{1-x} alloys. (After [49.7])

for the magnetization is given by minimizing the total energy, $dE/d\theta = 2K_{\text{eff}} \sin\theta \cos\theta + MH_e \sin\theta = 0$. The solution $\sin\theta = 0$ implies that the magnetization lies at either $\theta = 0$ or 180°. However, the energy must be a minimum for stability and the second derivative changes sign at a boundary between stability and instability

$$\frac{d^2E}{d\theta^2} = 2K_{\text{eff}}(\cos^2\theta - \sin^2\theta) + MH_e^* \cos\theta = 0\,.$$

The magnetic field that just satisfies this condition is

$$H_e^* = -H_c = \frac{-2K_{\text{eff}}}{M}\,. \qquad (49.4)$$

The critical magnetic field is the coercive field (H_c). A curve of the magnetization versus applied magnetic field for this case is shown in Fig. 49.4 (curve labeled *easy*), which was measured on a thin film of FeCoNi (to be discussed) [49.8]. The magnetization curve exhibits hysteresis and in the FeCoNi alloy the coercive field is small and the material is magnetically *soft*. Hysteresis is responsible for the use of magnetic materials as digital memory devices since the up or positive state of the magnetization could represent a binary 1 and the down or negative state could represent a binary 0 and these states can be stable even in zero applied field ($H_e = 0$).

The above analysis also applies to magnetically hard films, for which the coercive field is large. We will discuss magnetic recording media and here the coercive field is measured in units of kOe. With the magnetic

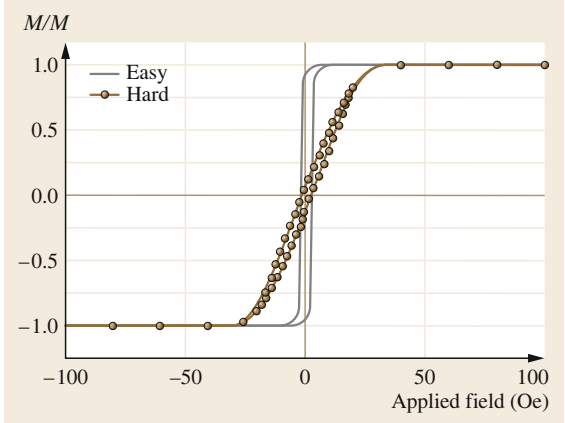


Fig. 49.4 Easy and hard magnetization curves of an electroplated FeCoNi alloy film. (After [49.8])

field applied perpendicular to the easy axis, along the hard axis, the energy density is given by

$$E_k = K_{\text{eff}} \sin^2\theta - MH_e \sin\theta\,.$$

The derivative of the total energy is given by

$$\frac{dE}{d\theta} = 2K_{\text{eff}} \sin\theta \cos\theta - MH_e \cos\theta = 0$$

resulting in the two solutions:

1. $\cos\theta = 0$, or $\theta = 90$ or 270°
2. $\sin\theta = H_e/(2K_{\text{eff}}/M)$.

The solution to this case is also shown in Fig. 49.4 (curve labeled *hard*) and is seen not to exhibit hysteresis. This is the hard-axis magnetization curve and the magnetic field where the magnetization saturates is known as the anisotropy field H_k. The hard-axis magnetization curve is used extensively for magnetic recording heads.

49.1.2 The Write Head

The write head is formed from thin films of soft ferromagnetic alloys patterned in the form of a magnetic yoke. A schematic of an eight-turn thin-film head is shown [49.9] as Fig. 49.5. Heads in disk drives typically have fewer than five turns to reduce inductance. The cross-hatched regions are the soft ferromagnetic material, and the leads that carry the write current are shown on the right-hand side of the top view. Note that the widths of the copper alloy turns are increased near the back region of the head to reduce resistance. The heads are fabricated using electroplating and the substrate (not shown) is a hard ceramic that can be machined to form the slider [49.5]. Electrical current in the form of pulses is coupled to the yoke to generate a magnetic field at the gap. The coil is insulated from the metallic magnetic yoke by layers of baked polymer photoresist. The alloy that has been used most frequently in the past for the magnetic films in the write head is $Ni_{80}Fe_{20}$ (Permalloy), which can be deposited in thin films using electroplating. Electroplating is discussed in *Comstock* [49.5] and in more detail by *Andricacos* and *Romankiw* [49.10].

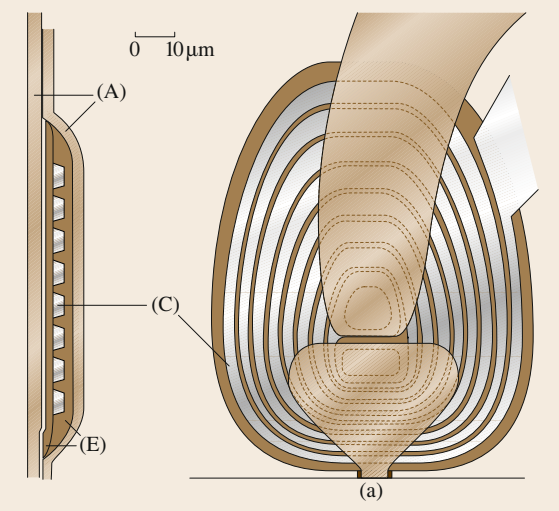


Fig. 49.5 Thin-film head with eight turns showing the poles and yoke structure. A is Permalloy, C is copper and E is baked photoresist. (After [49.9])

The maximum coercivity ($H_{c,max}$) of the recording medium that can be recorded by the write head is given by (49.4) (in Gaussian units) [49.5]

$$H_{c,\,max} = 0.20 \times 4\pi M_s \tan^{-1}\left(\frac{\frac{g}{2}}{d+\delta}\right), \qquad (49.5)$$

where $4\pi M_s$ (Gauss) is the saturation flux density of the material used in the recording or write head, g is the gap length of the write head, d is the HMS (see Fig. 49.2 caption) and δ is the thickness of the recording film (shown as t in Fig. 49.2).

To increase the linear density (transitions along a recorded track), it is necessary to increase the coercive field of the recording media. This is the case since the length of a transition between the two states of the magnetization is given by (49.5) (in Gaussian units) [49.5, 11, 12]

$$l = \pi a = \pi\left(\frac{d(1-S^*)}{\pi Q}\right) + \pi\left[\left(\frac{d(1-S^*)}{\pi Q}\right)^2 + \left(\frac{2M_r\delta}{H_c}\right)\left(\frac{2d}{Q}\right)\right]^{1/2}, \qquad (49.6)$$

where

$$S^* = 1 - \frac{M_r/H_c}{\left[(\mathrm{d}M/\mathrm{d}H)_{(H=-H_c)}\right]}$$

quantifies the slope of the hysteresis curve of the recording medium at the coercive field and

$$Q = \frac{(\mathrm{d}H_x/\mathrm{d}x)}{H_c/\mathrm{d}};$$

Q is the normalized slope of the horizontal component of the write-head magnetic field at the value $H = H_c$. A typical value of Q is 0.7.

M_r (emu/cm^3) is the remanent magnetization of the recording medium with thickness δ.

From (49.4) and (49.5) it is seen that for reduced transition length (and hence higher linear density) it is necessary to increase the coercive field (H_c) of the recording medium but to do so requires a concomitant increase in the magnetization of the write head (49.4).

Materials with magnetization larger than Permalloy ($4\pi M_s = 10$ kG ($M_s = 1$ T)) and which are magnetically soft (large permeability) are listed in Table 49.1.

The most common material used in disk drives is electroplated $Ni_{45}Fe_{50}$, with a saturation magnetization of $4\pi M_s = 16$ kG ($M_s = 1.6$ T) [49.13].

Table 49.1 Properties of materials used in magnetic recording heads

Material	B_s (kG) ($T \times 10^{-1}$)	μ_r	ρ ($\mu\Omega$ cm)	H_c (Oe)(A/m)	References
$Ni_{80}Fe_{20}$ (Permalloy)	10	1500–3000	20	0.1–0.5 (7.95–39.78)	*Bozorth* [49.10]
Sendust (FeSiAl)	10	8000	85	0.25 (19.9)	*Comstock* [49.5]
$Ni_{45}Fe_{55}$	16	1700	48	0.3 (23.87)	*Robertson* [49.13]
$Co_{48}N_{22}Fe_{30}$ electroplated with pulsed plating	21.5	1000	–	$\approx$ 1 (79.6)	*Liu* [49.14]
$Co_{56}Ni_{13}Fe_{31}$ (C) electroplated with carbon impurity	19	–	$\approx$ 70	$\approx$ 2 (159.2)	*Osaka* et al. [49.15]
FeAl(2% Al)N (N/Fe 6.6%) sputtered	20	2450	38	3.8 (302.4)	*Liu* et al. [49.16]
FeTiN sputtered	18–21	–	–	$\approx$ 5 (397.9)	*Ding* et al. [49.17]

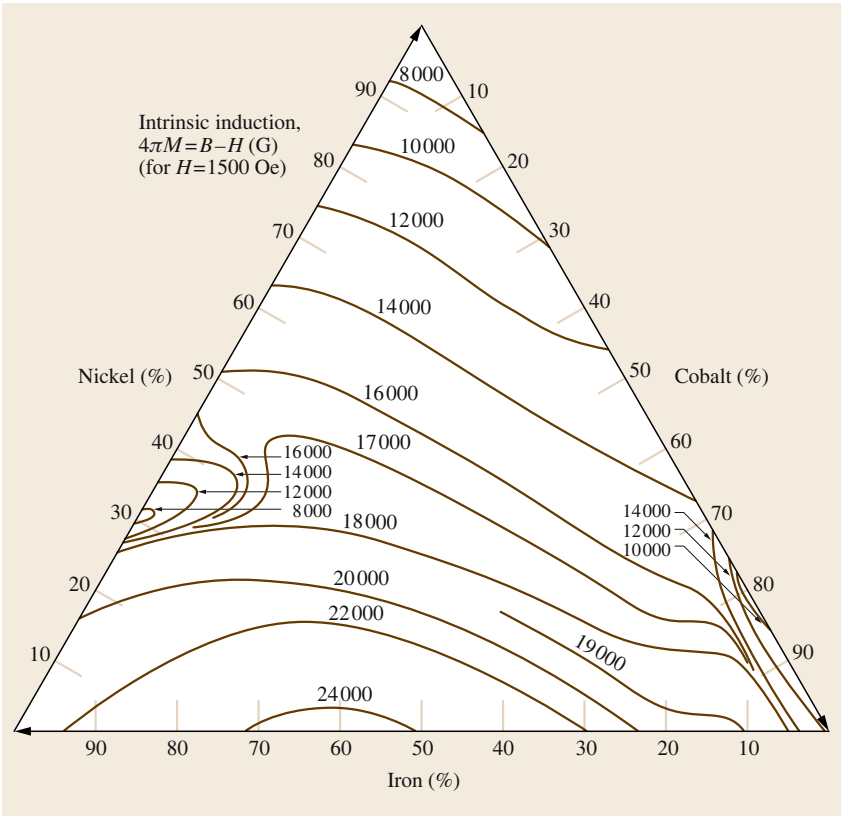


Fig. 49.6 FeCoNi saturation magnetization contours. (After [49.6])

Ternary alloys of Ni, Fe and Co also can be electroplated with saturation magnetization of $4\pi M_s \geq 20$ kG ($M_s = 2$ T). The composition region that results in the largest value of saturation magnetization but also large permeability is: $Co_{48-56}Fe_{30-31}Ni_{22-13}$ (Table 49.1). A useful guide to the selection of the composition of the ternary FeCoNi alloys is the composition diagram shown [49.6] as Fig. 49.6.

The data for Fig. 49.6 were accumulated in bulk materials. In a study of electroplating FeCoNi alloys for write heads [49.8] care was taken to avoid any impurities in the plating bath and the results showed the possibility of achieving $4\pi M_s$ values in the range of 21–21.5 kG and $H_c < 2$ Oe in the same region of the composition diagram (approximately $Fe_{35}Co_{45}Ni_{20}$).

However, experiments in a larger cell in conditions closer to those in a manufacturing plating system and with higher Fe content and an organic additive ascorbic acid used to reduce the formation of Fe^{3+} showed even larger values of $4\pi M_s$ (> 2.2 kG) but with increased coercivity [49.6] (> 8 Oe). The composition of ferromagnetic alloys can be selected to result in zero values for the magnetostriction constant and the Permalloy composition of nickel-iron alloys ($Ni_{80}Fe_{20}$)

has a zero magnetostriction constant (Fig. 49.3). The magnetostriction constant is frequently chosen to be small and negative so that the magnetoelastic energy defines an easy axis at right angles to the direction of the stress ($\theta = 90°$ in (49.2)).

In the case of thin-film write heads the stress is along the axis of the poles (upward in the patterned yoke in Fig. 49.4) and with negative magnetostriction constant the easy axis is across the pole width, which means that the direction of flux flow is along the hard axis with the desired linear relation between the magnetization and the magnetic field (Fig. 49.4). In the case of the ternary alloys of FeCoNi an alloy composition that results in zero magnetostriction is $Fe_{13}Co_{31}Ni_{56}$, in which the nickel content is too high to be a useful write-head material [49.8]. Alloys of FeAl (2% Al) [49.16] and FeTi ($\approx$ 2% Ti) [49.17] sputtered using a mixed N and Ar working gas offer desirable magnetic properties, as shown in Table 49.1, but have not replaced the electroplated alloys because of increased costs of sputtering.

A material of interest for even higher values of saturation magnetization is the α'' phase of $Fe_{16}N_2$, which has the potential for a saturation magnetization of 29 kG; however, this phase is metastable [49.7]. The actual values of saturation magnetization with this phase in thin-film form was 20 kG and required annealing at 200 °C [49.18].

To record on recording media with increased coercivity is not the only issue with the magnetic materials used in write heads. It is also important that the write heads have high efficiency. Efficiency (η) in this case is defined as the ratio

$$\eta = \frac{H_g g}{N_w I} , \tag{49.7}$$

where H_g is the value of the magnetic field in the gap of the write head and I is the amplitude of the write current pulse. High efficiency is important to allow write-current amplitudes that are easily supplied from integrated circuits. For high efficiency it is necessary that a large percentage of the magnetomotive force ($N_w I$) results in a magnetic field across the gap. However, the permeability of the yoke influences the efficiency

$$\eta = \frac{\frac{g}{A_g}}{\frac{g}{A_g} + \frac{l_c}{\mu_r A_c}} , \tag{49.8}$$

where A_g and A_c are the area of the gap and core respectively; l_c is the length of the core and μ_r is the relative permeability of the core or yoke material. As shown in Fig. 49.3, the relative permeability (μ_r) of nickel-iron alloys drops for compositions different from the Permalloy composition. This decrease in the permeability of nickel-iron alloys with composition different from that for Permalloy is the direct result of the increase in crystalline anisotropy and magnetostriction, as shown in Fig. 49.3. The lower permeability of the FeCoNi alloys can be obviated in a design of the write head in which the high-magnetization films are used only in the gap region in a design referred to as a stitched head. The values of coercivity of the recording medium (H_c) that can be recorded with recording heads with the saturation magnetization of Permalloy ($4\pi M_s = 10$ kG ($M_s = 1$ T)) and materials with saturation magnetization 2.5 times that for Permalloy ($4\pi M_s = 25$ kG ($M_s = 2.5$ T)) are shown versus the magnetic spacing (d) in Fig. 49.7.

In 2006, magnetic spacing values were 15 nm (carbon overcoat for head: 3 nm, pole tip recession tolerance: 3 nm, flying height: 5 nm, and media carbon overcoat: 4 nm).

In today's technology magnetic spacing values are 13 nm (carbon overcoat for head: 3 nm, pole tip recession tolerance: 3 nm, flying height: 3 nm, and media carbon overcoat: 4 nm).

The trend in coercivity (H_c) and flux ($M_r\delta$) versus areal density in commercial disk media is shown in Fig. 49.8 [49.7].

Areal density is determined by bit size where size is length in LMR whereas it corresponds to diameter in PMR.

Areal density plays a major role in magnetic recording similarly to minimal feature length F in electronic circuits (Table 49.15); additionally, most magnetic spacings scale with respect to it (Sect. 49.1).

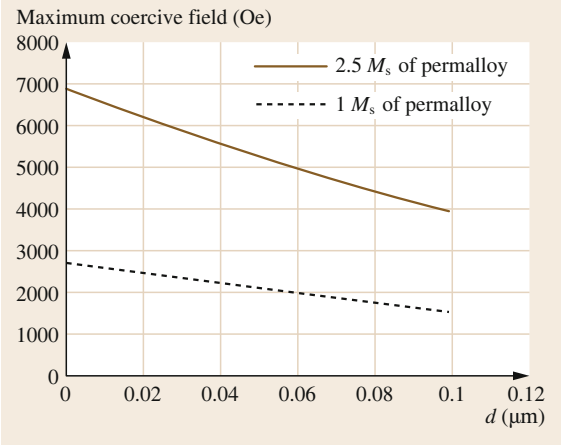


Fig. 49.7 Maximum coercivity of recording media versus magnetic spacing (d) for write heads fabricated from Permalloy and from a magnetic material with a saturation magnetization equal to 2.5 M_s of Permalloy

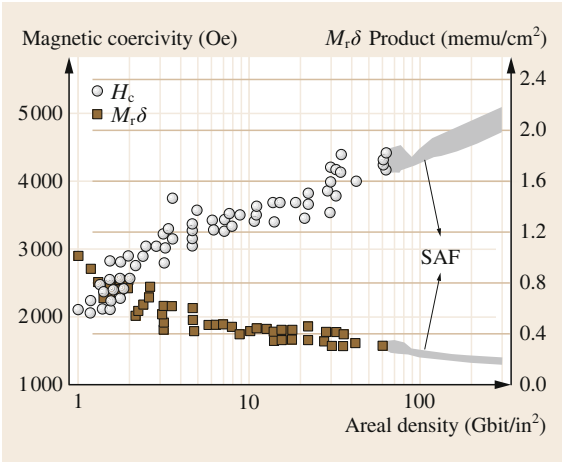


Fig. 49.8 Trend in coercive field (H_c) and media flux ($M_r\delta$) versus areal density in Gbit/in^2 for disk drives. (After [49.7])

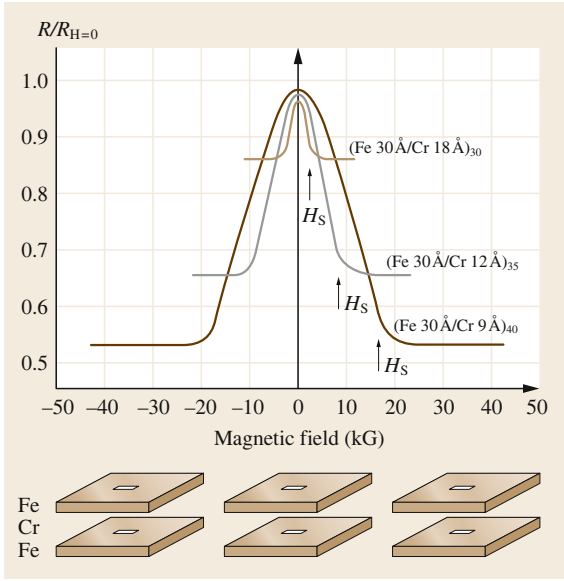


Fig. 49.9 The magnetoresistance of three (Fe/Cr)$_n$ superlattices at 4.2 K. The thickness of the layers in Angstroms is shown. (After [49.20])

We discuss synthetic antiferromagnetic media (SAF media) further below.

With the increased linear density associated with increased areal density in disk drives and with increases in the rotation speed of the disks, the rate at which data are recorded (and read back) has increased over time. This increase in data rate leads to a requirement that the magnetic fields in the gap of a write head switch in ever decreasing times. The switching time of thin magnetic films as used in write heads is limited by three factors [49.19]:

1. The fundamental switching time resulting from the fact that magnetic moments are always associated with angular momentum that cannot be changed instantaneously
2. The eddy currents associated with the changing magnetic fields both decrease the magnetic fields and slow down the switching speed
3. The inductance of the coil.

Data rates depend on the form factor of the disks and with high-end 3.5 in drives the data rates are larger than 50 Mb/s, while with 2.5 in drives the data rates are roughly one-half as large.

49.1.3 Spin-Valve Read Head

The technology that has evolved for sensing or reading the flux emanating from transitions in thin-film disk media is the spin valve. This technology is based on the giant-magnetoresistive (GMR) effect discovered in France and published in 1988 [49.20]. GMR results when thin ferromagnetic films are coupled by an even thinner spacer film of a nonmagnetic transition metal. A curve of resistance normalized to that with no magnetic field applied versus magnetic field for arrays of ferromagnetic films of Fe with thickness of 30 Å (3 nm) coupled by Cr with thickness varying from 9–18 Å (0.9–1.8 nm) is shown in Fig. 49.9 [49.20].

The number subscripted to the parenthesis characterizing the magnetic and transition-metal films is the number of such films, which are referred to as superlattices. With no applied magnetic field, the two films were antiferromagnetically coupled; that is, the magnetizations in the two films are equal in magnitude and oppositely directed. The application of a large in-plane magnetic field rotated the magnetization of the coupled films to a parallel configuration, reducing the resistance (to be discussed).

The current used to measure the magnetoresistance, the change in resistance with applied magnetic field, is applied in the plane of the films. This orientation of current is referred to as current-in-plane (CIP). Later we will discuss an alternative orientation of the current, the current-perpendicular-to-plane (CPP) case.

The initial discovery of GMR was done with a temperature of 4.2 K and the magnetic fields applied were measured in kOe ($10^6\,4\pi$ [A/m]).

The large change in resistance with magnetic field was in contrast to that observed in the anisotropic magnetoresistance (AMR) effect, which is the change in resistance observed when the magnetization in a single ferromagnetic film is rotated from an easy toward a hard axis. This latter technology, for which the resis-

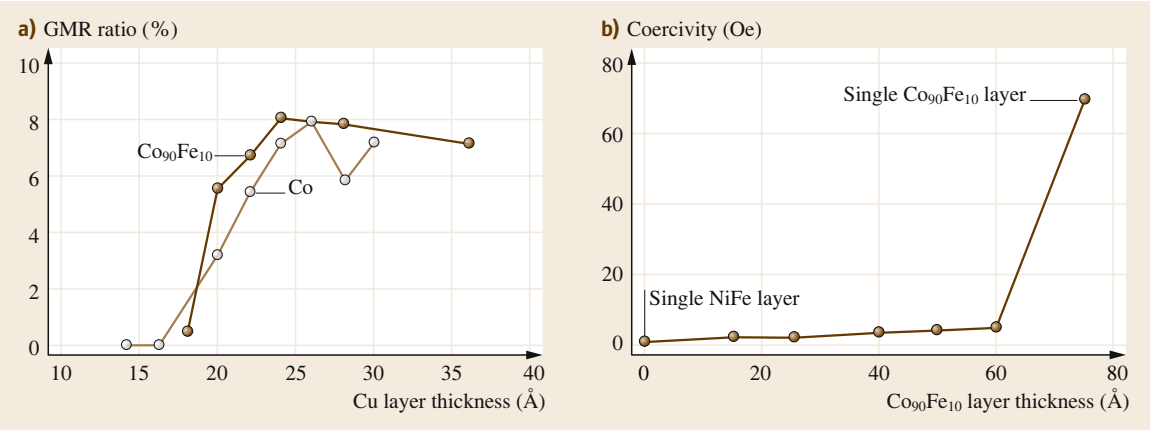


Fig. 49.10 **(a)** Comparison of GMR ratio ($\Delta\rho/\rho$) as a function of copper layer thickness in $Co_{90}Fe_{10}$ and Co films (see text). **(b)** Coercivity as a function of $Co_{90}Fe_{10}$ layer thickness in Ta (50 Å)/[$Ni_{80}Fe_{20}$/$Co_{90}Fe_{10}$] (75 Å) films. (After [49.21])

tance change with Permalloy was of the order of 2–3%, was being used in read heads during the time of research on the GMR effect.

By the early 1990s it was found to be possible to observe GMR in other film materials with increased spacer film thickness at 300 K and with magnetic fields of tens of Oersteds. For example, the GMR ratio for two coupled films of an alloy of cobalt and iron ($Co_{90}Fe_{10}$ for which the magnetostriction constant is zero) with a variable thickness of the spacing layer of copper is shown in Fig. 49.10 [49.21].

It is seen that the GMR ratio is considerably reduced from that for the Fe/Cr superlattices at 4.2 K (Fig. 49.8) but the ratio is larger than that for the AMR effect.

The discovery of the GMR effect led to the invention of the spin-valve read head, in which one of two ferromagnetic films coupled by a thin transition-metal film is maintained constant in direction (the pinned film) and the orientation of the other film (the free film) is allowed to rotate (in the plane of the magnetic film), resulting in a decrease in resistance [49.22]. We will discuss the technology of pinning later. By free, it is meant that the magnetization in the film is easily rotated, for example by the magnetic field arising from transitions in the recording medium.

It is surprising that cobalt is found to be useful in this application since cobalt is magnetically hard (pure cobalt metal has a high coercivity and is used in recording surfaces) and one of the requirements for the GMR effect is that the free film has high permeability and high efficiency (49.6).

The reason for this result is shown in Fig. 49.10, where the coercivity of a free film consisting of two films: $Ni_{80}Fe_{20}$ (60 Å (6 nm)) and $Co_{90}Fe_{10}$ is plotted versus the thickness of the $Co_{90}Fe_{10}$ film [49.21]. Even though the cobalt alloy would have a large coercivity in the bulk, the coercivity for thin films, together with the underlying $Ni_{80}Fe_{20}$ film, is low for thicknesses less than about 60 Å.

The advantage of the cobalt-iron film is the increased GMR ratio over $Ni_{80}Fe_{10}$. In order to increase the GMR ratio even more it has been found advantageous to deposit an interfacial nanolayer of cobalt metal on one or both sides of the free film [49.23]. We will learn in a discussion of the origin of GMR that cobalt has the potential for large GMR ratios.

A schematic of the spin-valve read head is shown in Fig. 49.11 [49.22]. The magnetization of the free film is shown rotated to an angle θ_1 and the pinned layer is shown with a fixed angle of θ_2. The pinning effect

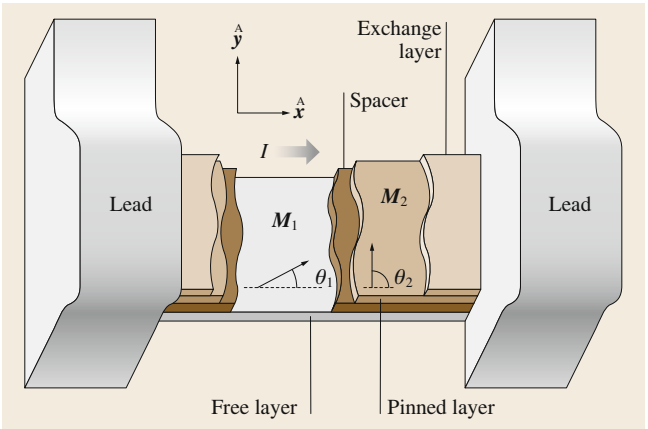


Fig. 49.11 Schematic of spin-valve read element. The magnetization of the free layer is shown at an angle θ_1 to the easy axis (and direction of the current) and the magnetization of the pinned layer is fixed at 90° to this axis. (After [49.22])

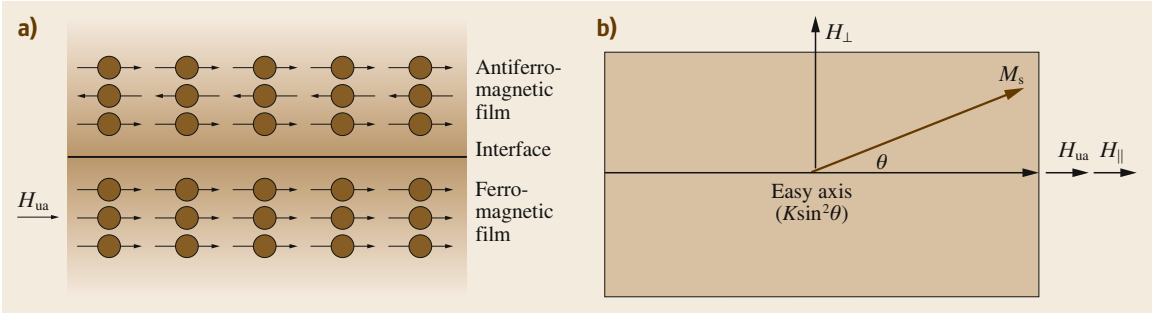


Fig. 49.12 **(a)** Schematic of the coupling of an antiferromagnetic to a ferromagnetic film. **(b)** Geometry of the ferromagnetic film showing the induced easy-axis field H_{ua}, the external magnetic fields $H_\parallel$ and $H_\perp$ and the magnetization of the film at an angle θ to the easy axis. (After [49.5])

is due to the coupling of the pinned film to an underlying antiferromagnetic film as we will discuss. The nonmagnetic spacing layer separates the two ferromagnetic films. The maximum output voltage for the spin valve with no external resistance (and no reduction in the flux due to shielding – to be discussed) is

$$\Delta V = \eta J W_r \Delta\rho \frac{(1-\sin\Delta\theta_1)}{2}\,, \qquad (49.9)$$

where η is the efficiency of the read head, which accounts for the fact that the angle θ_1 is not uniform over the height of the free film, J is the density of the current flowing through the films, W_r is the width of the read head (always less than the width of the written track) and $\Delta\rho$ is the change in resistivity of the stack of films with a change of θ_1 from $-90-+90°$, again with θ_2 fixed. The factor of 2 in the denominator is the result of the actual rotation of θ_1 being limited to $0-90°$.

The device senses magnetic field in the plane of the free film resulting from transitions in the recorded media ($H_{y,\,disk}$) through the rotation of the magnetization in the free film

$$\sin\theta_1 = \frac{H_{y,\,disk}}{H_k}\,, \qquad (49.10)$$

where H_k is the anisotropy field of the free film (see analysis of the hard-axis magnetization curve in Sect. 49.1.1). In the absence of any other magnetic fields, the magnetization will lie along the easy axis. The hard axis is perpendicular to the easy axis. The voltage is measured across the lead terminals using a sense current and the width of the track is just the spacing between the leads. Typically, when areal density is higher than 100 Gbit/in^2, in spin valves, peak output voltage per unit of track width must be larger than [49.24] about 10 mV/μm. Spin-valve read heads are shielded on each side of the stack of films to realize a narrow pulse width. The total width between the shields is G and the thickness of the stack of films comprising the spin valve is t. The read pulse is characterized by the amplitude but also by the width at one-half amplitude (PW_{50} in distance units) (discussed in [49.5, 12])

$$PW_{50} \approx \sqrt{4\left(d+a+\frac{\delta}{2}\right)^2 + \frac{(G^2+t^2)}{2}}\,, \qquad (49.11)$$

where G is the total spacing between the shields and δ is the thickness of the recording layer.

The read-head peak output voltage between shields [49.12] with no external resistance can be approximated by

$$V_{GMR}\ (\text{peak}) = \frac{\eta I \Delta R_{max} M_r^{rec}\delta}{\pi t M_s}\left(\frac{g+t}{g}\right) \times 2\tan^{-1}\frac{g}{2(d+a)}\,,$$
$$I\Delta R_{max} = J W_r R_\square t_{sv}\left(\frac{\Delta R_{max}}{R_0}\right), \qquad (49.12)$$

where η is read efficiency, defined as the fraction of the flux sensed by the head to the total flux, M_r^{rec} is the magnetization of the recording layer, M_s the saturation magnetization of the free layer, g the half-gap between the shields, $R_\square$ the head square resistance ($R_\square = \rho_r/t_r$, with ρ_r, t_r the head resistivity and thickness respectively), W_r the head width and t_{sv} the free layer thickness (modified from *Mallinson* [49.25]).

$$\left(\frac{\Delta R_{max}}{R_0}\right)$$

ratio is a typical measure of resistance variation used for GMR performance (see further below).

Pinning with an Antiferromagnetic Film

For a spin valve to function, the pinned ferromagnetic film must have its magnetization perpendicular to the

Table 49.2 Comparison of alternative antiferromagnetic films for transverse bias of Permalloy films

Property	NiFe/FeMn [49.26]	NiFe/NiMn [49.26]	NiO/NiFe [49.26]	CoNiO/NiFe [49.27]	NiFe/IrMn [49.27]	NiFe/CrMnPt [49.28]	NiFe/Pd-Pt-Mn [49.29]
Corrosion resistance	Low	High	Very high	Excellent	Moderate	Good	High
Coupling field H_{ua} Oe (kA/m)	≈ 77 (6.12)	≈ 206 (16.4)	≈ 46 (3.66)	≈ 45 (3.58)	≈ 350 (27.84)	60 (4.77)	$(Pd_{0.6}Pt_{0.4})_{0.5}Mn_{0.5}$ 80 (6.36)
Coercivity H_c Oe (kA/m)	≈ 6 (0.48)	≈ 116 (9.23)	≈ 35 (2.78)	≈ 23 (1.82)	11 (0.88)	8 (0.64)	16 (1.27)
Blocking temperature (°C)	≈ 150	≥ 450	≈ 200	150	250	320	300
Annealing	Not required	Required	Not required	Not required	Not required	Required	Required

recorded track, as shown in Fig. 49.11 (the vertical magnetic field from the transitions in the disk is in the y-direction). This orientation of the magnetization in the pinned film is achieved by coupling the ferromagnetic film to an antiferromagnetic film and the coupling results from the layer of spins in the antiferromagnetic film adjacent to the top layer of spins in the ferromagnetic film and is characterized by a coupling or exchange field of value H_{ua}. The geometry for the coupling is shown in Fig. 49.12 and the ferromagnetic film geometry with a magnetic field along the easy axis ($H_{||}$) and the hard axis ($H_{\perp}$) as Fig. 49.12.

The coupling results in shifting the easy axis hysteresis curve ($M_{||}/M_s$ versus $H_{||}$) of the ferromagnetic film by an amount H_{ua}, as shown in Fig. 49.13 [49.5]. The magnetization of the pinned film is held constant at $-M_s$ over the range of magnetic fields from large negative fields to $H_{ua} + H_k$. Note that the magnetization could be held at $+M_s$ by reversing the direction of the lower layer of spins in the antiferromagnetic film. The coupling field depends on the temperature and vanishes at the blocking temperature. The direction of the coupling is established during an annealing step by heating the structure above the blocking temperature, where the film is paramagnetic, in the presence of a magnetic field and then cooling the film below the blocking temperature.

The choice of antiferromagnetic films to be used in this application is extensive as shown in Table 49.2 [49.26–29]. The parameters listed in Table 49.2 are the strength of the coupling field (H_{ua}), the blocking temperature and the susceptibility of the antiferromagnetic and ferromagnetic films to corrosion. Corrosion is of concern in recording heads because of the mechanical lapping process used to form the surface on the bottom of the slider on which the air-bearing contour is etched [49.4].

One antiferromagnetic film with particularly favorable corrosion propensity is NiO. This film is typically used in a bottom spin valve in which the antiferromagnetic film is closest to the substrate. The other

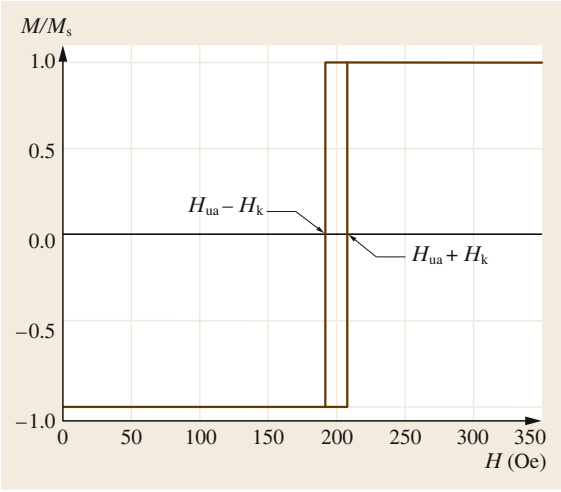


Fig. 49.13 Magnetization curve with the magnetic field along the easy axis ($M_{||}/M_s$ versus $H_{||}$) for coupled antiferromagnetic and ferromagnetic films. (After [49.5])

antiferromagnetic films are typically formed on top of the stack of films. Another advantageous film is the alloy of $Pd_{30}Pt_{20}Mn_{50}$ [49.29].

This ternary alloy has a favorable coupling strength H_{ua} but also a high blocking temperature (300 °C) and favorable corrosion resistance. $Pd_{30}Pt_{20}Mn_{50}$ does require annealing the stack at 220–240 °C. The advantage of using an antiferromagnetic film for the pinning film instead of a permanent magnetic film is that the antiferromagnetic film does not itself respond to any external magnetic field since the film has a zero net magnetization.

Synthetic Antiferromagnets

The magnetic coupling between magnetic films closely spaced by certain nonmagnetic metal films oscillates between antiferromagnetic and ferromagnetic coupling. Using this fact, it is possible to design film structures that have a desired magnitude and sign of the magnetic coupling [49.30]. A schematic of a structure of

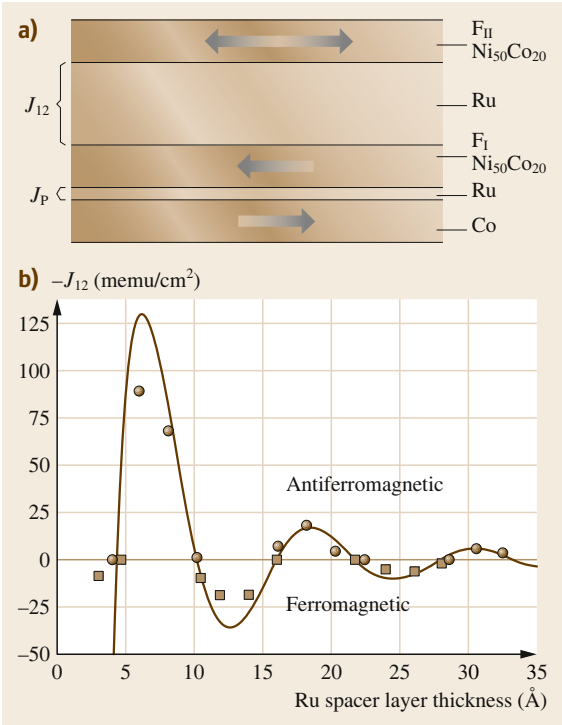


Fig. 49.14 **(a)** Geometry of ferromagnetic films coupled with thin layers of Ru. **(b)** Magnetic coupling constant per unit of film area between layers F_I and F_{II} of the films shown in **(a)** versus the thickness of the Ru layer. (After [49.30])

films that exhibits this variable coupling is shown in Fig. 49.14.

Here the strong antiferromagnetic coupling of Co through Ru to soft $Ni_{80}Co_{20}$ is used to pin the magnetization in the adjacent $Ni_{80}Co_{20}$ film (F_I) to the direction opposite to the Co film and the thickness of the second layer of Ru can be adjusted to change the sign of the coupling between the two soft-magnetic films (F_I and F_{II}). Ruthenium is selected for the coupling film since it has a strong and oscillating coupling between Co, $Ni_{80}Fe_{20}$ and $Ni_{80}Co_{20}$. Ruthenium has the hexagonal close-packed (hcp) crystal structure with ($a = 0.271\,\text{nm}$ and $c = 0.428\,\text{nm}$) while cobalt also has the hcp structure with ($a = 0.250\,\text{nm}$ and $c = 0.410\,\text{nm}$) and therefore the two metals are expected to grow epitaxially.

The experimental coupling constant (J_{12}) for this configuration of layers, where the soft-magnetic film is Ni_{80} Co_{20}, is shown in Fig. 49.14. The units of the coupling constant are $10^{-3}\,\text{emu/cm}^2$ or memu/cm^2 (10 A in SI units) (coupling per unit of film area). The coupling constant can be either antiferromagnetic ($J_{12} < 0$) or ferromagnetic ($J_{12} > 0$) and can have a magnitude varying between $+25 - -75\,\text{memu/cm}^2$ ($+250 - -750$ A). By varying the thickness of the Co layer it is possible to adjust the total magnetic field at the layer F_{II} to be zero.

An overall spin valve consists of a sequence of thin films such as: MnFe 21 Å/Co 30 Å/Ru 7 Å/Co 30 Å/Cu 25 Å/Co 30 Å/$Ni_{80}Fe_{20}$ 28 Å, where the first four films (MnFe/Co/Ru/Co) form a synthetic antiferromagnetic film since the final Co film of these four films is pinned to a direction opposite to that for the first Co film. This Co film then acts as a pinned film in a spin valve together with the final three films (Cu, Co, and $Ni_{80}Fe_{20}$). The thickness of ruthenium listed in the sequence of films is optimum for antiferromagnetic coupling between the Co films. The last Co film is a nanolayer, which increases the magnetoresistive coefficient, as we have discussed [49.23]. The advantages of the synthetic antiferromagnet for pinning the pinned film compared with just an antiferromagnetic film (e.g., FeMn, NiO, NiMn or IrMn) are:

1. The pinning coupling field is an order-of-magnitude larger.
2. The magnetostatic coupling field from the pinned layer to the free layer is reduced since the two Co films in the synthetic antiferromagnet are antiferromagnetically coupled and hence their magnetostatic fields at the free film cancel.

The synthetic antiferromagnet contributes to ensuring that the only magnetic field at the free film is that from the transitions in the recording medium. One last requirement for the proper operation of the spin valve is that there is abutted on the two ends of the structure a permanent magnetic film referred to as the hard bias film (discussed in [49.12]). The permanent magnetic films are similar to the cobalt alloy magnetic films used for disk recording surfaces (to be discussed).

The purpose of the hard bias is to reduce the presence of magnetic domains in the free film and to reduce Barkhausen jumps of the domain walls in the presence of the external magnetic fields from the transitions in the recording medium.

The stack of films used in a spin valve has been discussed. The films used in Fig. 49.15, which we used to illustrate the GMR effect in a $Co_{90}Fe_{10}/Ni_{80}Fe_{20}$ free film, was: Ta (50 Å)/$Ni_{80}Fe_{20}$ (60 Å)/$Co_{90}Fe_{10}$ (55 Å)/Cu (variable)/$Co_{90}Fe_{10}$ (55 Å)/FeMn (100 Å)/Ta (70 Å) [49.21]. In this case the antiferromagnetic film was FeMn (typically about 50%, 50%) and the Ta films are used to protect the rest of the stack.

Origin of GMR

The physical origin of the GMR effect lies in the differential scattering of electrons with spins in the same

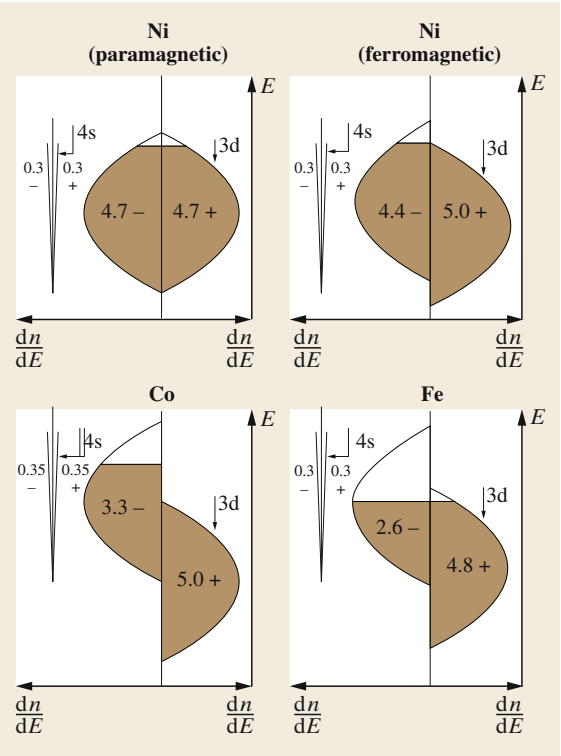


Fig. 49.15 Schematic of density-of-states curves for the ferromagnetic transition elements (Ni, Co and Fe). Both the 4s and 3d bands are shown. The magnetization of the elements is due to the difference between the spins in the spin-up and spin-down bands and the bands are full of electrons to the Fermi level (the *shaded region*). (After [49.31])

direction as the local magnetization (spin-up electrons) or opposite to this direction (spin-down electrons) at the interfaces and in the bulk part of the thin films comprising the GMR stack of films. The ferromagnetic transition elements (Fe, Co and Ni) have different 3d band structures, as shown in Fig. 49.15 [49.31] and discussed by [49.5]. This figure illustrates the energy of electrons in the two half-bands for spin-up and spin-down electrons versus the density of states $D(E)$ (number of states per unit energy per unit volume) in the 3d band for the ferromagnetic transition metals. The 4s bands are also shown. Electrons in the 4s bands make only a small contribution to the net spin, however, due to the larger mobility (inversely proportional to the curvature of the bands), they are largely responsible for the current.

The bands are filled to the Fermi energy at 0 K, and states above, but close to, the Fermi level are occupied at room temperature (illustrated by the shaded region in Fig. 49.15). The conductivity of a metal is given by $\sigma = e^2 v_F^2 \tau D(E_F)/3$, where e is the electron charge, v_F is the velocity of electrons at the Fermi energy, τ is the mean scattering time and $D(E_F)$ is the density of states at the Fermi energy [49.32]. If the mean scattering time (τ) is held constant then it is seen that the conductivity increases with the density of states at the Fermi energy; however, the scattering of the 4s electrons with a given spin is large when there are a large number of scattering centers and empty states with the same spin in the 3d band, i. e., the density of states at the Fermi energy $D(E_F)$ in the 3d band is large.

In ferromagnetic transition metals this is the larger effect [49.33] leading to $\rho = \frac{1}{\sigma} \propto D(E_F)$.

The scattering is due to spin-spin scattering, and spin-orbit scattering is believed to be small [49.34, 35]. The magnetic moment per atom can be deduced as the difference in the number of electrons/atom with spin-up (the majority) versus those with spin-down. For example, the magnetic moment per atom for iron is $4.8-2.6 = 2.2$ Bohr magnetons (μ_B) per atom. From Fig. 49.15 we conclude that the differential scattering will be large in cobalt since for this element the difference in the density of states at the Fermi level is large (approximately 8–10 to 1 [49.31]). A schematic of the density of states versus energy for the three films comprising the spin-valve structure from the two extreme cases of the orientation of the ferromagnetic films (ferromagnetic configuration and antiferromagnetic configuration) is shown [49.31, 36] in Fig. 49.16. All films have same Fermi energy. *Mathon* [49.31] described the GMR effect by discussing the differential scattering of holes in the 3d bands, but by concentrating on the scattering of electrons from the 4s to the 3d band we can make the same conclusions [49.34, 36].

It is assumed that the mean free path for the electrons is large compared with the thickness of any layer in the superlattice. Consider the ferromagnetic configuration case (top of Fig. 49.16): here the majority electrons have spin-up and the scattering of 4s electrons is low in both ferromagnetic films since both have a low density of states at the Fermi level. The nonmagnetic film has a larger density of states and will scatter electrons regardless of the orientation of the ferromagnetic films, which is why this film must be thin. Now consider the antiferromagnetic configuration case in Fig. 49.16: from the density of states at the Fermi level in the three films it is seen that the scattering is large in the right-hand ferromagnetic film for the electrons leaving the left-hand ferromagnetic film with spin-up (top density-of-states curve), resulting in large resistivity for this class of electrons. Similarly, the scattering of spin-down electrons is large for the left-hand ferromagnetic film.

The GMR ratio (%) resistivity for the spin-up and spin-down electrons for the three films (ferromag-

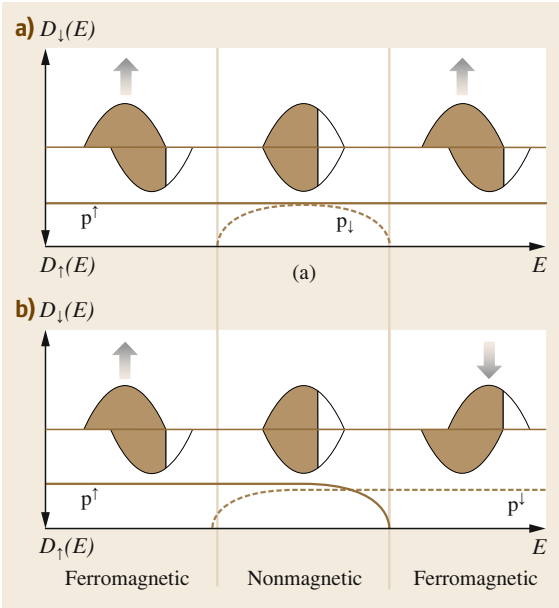


Fig. 49.16 Density-of-states curves for the spin-valve structure in the ferromagnetic spin configuration **(a)** and the antiferromagnetic spin configuration **(b)**. The resistivities of the layers are shown on the *bottom* part of the figures. (After [49.31])

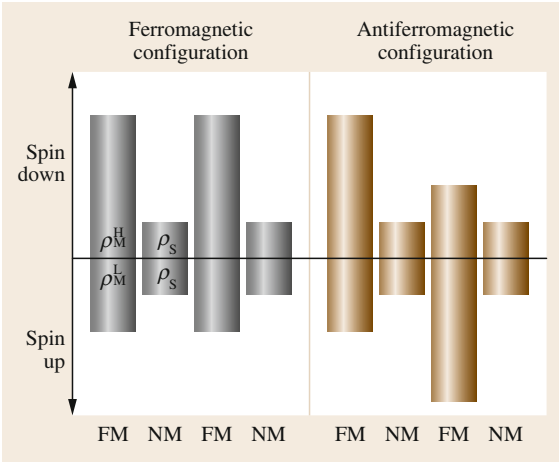


Fig. 49.17 Distributions of local resistivities for the spin-up and spin-down electrons in the ferromagnetic and antiferromagnetic spin configurations. (After [49.31])

netic – FM, nonmagnetic – NM and ferromagnetic – FM) in the ferromagnetic and antiferromagnetic orientation cases is shown as a schematic in Fig. 49.17 [49.31], where it is shown that the ferromagnetic orientation of the ferromagnetic films has a (relatively) low resistivity path for the spin-up electrons while the resistivity for the antiferromagnetic

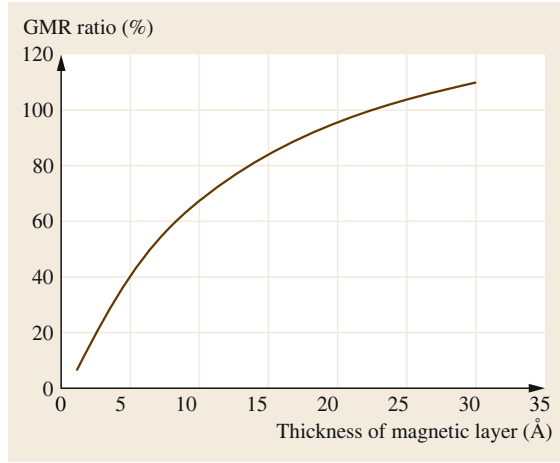


Fig. 49.18 $\Delta R/R$ versus the thickness of the magnetic layer for a nonmagnetic layer thickness of 10 Å for a Co/Cu spin valve. (After [49.31])

orientation of the ferromagnetic films is high for both spin-up and spin-down electrons.

The change in resistance of the spin-valve structure can be analyzed in terms of the different resistivities for the ferromagnetic and antiferromagnetic orientation of the two ferromagnetic films shown in Fig. 49.16 [49.31]

$$\frac{\Delta R_{\max}}{R_0} = \frac{(\alpha-\beta)^2}{4\left(\alpha+\frac{t_{NM}}{t_M}\right)\left(\beta+\frac{t_{NM}}{t_M}\right)}\,, \qquad (49.13)$$

where

$$\alpha = \frac{\rho_M^H}{\rho_s}\,, \quad \beta = \frac{\rho_M^L}{\rho_s}$$

with t_{NM} the thickness of the nonmagnetic film (spacer) and t_M the thickness of the magnetic film. ρ_M^H and ρ_M^L are respectively the high and low resistivities for two spin orientations of the magnetic (M) films, whereas ρ_s is the nonmagnetic (NM) spacer resistivity.

Figure 49.18 shows the variation of $\Delta R_{\max}/R_0$ for $t_{NM} = 10$ Å versus t_M for assumed values of $\alpha = 8.0$ and $\beta = 1.0$, which is appropriate for a Co/Cu superlattice [49.31].

When magnetic film thickness is beyond approximately 10 Å, the change in $\Delta R/R$ decreases. Figure 49.19 shows the variation of $\Delta R/R$ for $t_M = 20$ Å versus t_{NM}.

The decrease in $\Delta R/R$ is rapid with increasing thickness of the NM layer.

This combination of films is not used in spin valves since the cobalt film is not magnetically soft; the usual choice is Permalloy with a variety of nonmagnetic transition metals and the resulting $\Delta R/R$ is lower.

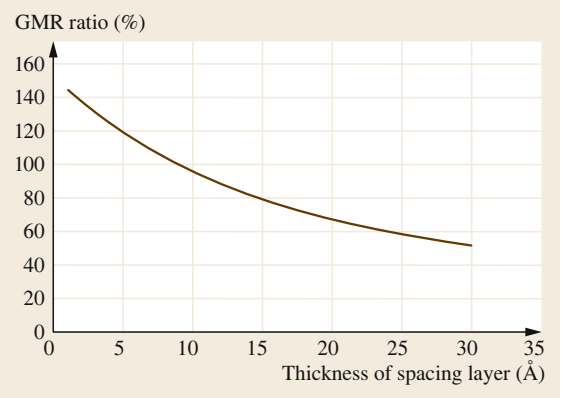


Fig. 49.19 $\Delta R/R$ versus the thickness of the nonmagnetic layer for a magnetic layer thickness of 20 Å for a Co/Cu spin valve. (After [49.31])

49.1.4 Longitudinal Recording Media (LMR)

Magnetic recording media used in disk drives are all sputter-deposited thin films of cobalt alloys for high coercivity and media flux ($M_r\delta$). Coercivities of 4000–5000 Oe are used (2006) and are limited by the ability of the write head to record on the media. The alloys that have evolved are four- or five-element alloys of CoPtCrX, where X is Ta or B or both. A typical atomic concentration of the four-element media is $Co_{64}Cr_{24}Pt_8B_4$. It has been found necessary to form a microstructure in the magnetic film that consists of a phase of small magnetic grains isolated by a nonmagnetic phase. The isolation of the magnetic grains results in less exchange coupling between grains and reduced noise – as we will discuss. Chemical segregation of $Co_{100-x}Cr_x$ alloys forms a nonmagnetic phase (Cr-rich) and a magnetic phase (Co-rich) and the nonmagnetic phase forms at the grain boundaries. The microstructure is controlled by undercoats that promote grain growth with the axes of the grains in the plane of the film. Platinum results in higher crystalline anisotropy and boron promotes the formation of an amorphous alloy of CoCrB, which grows at the grain boundaries, further reducing exchange coupling.

Figure 49.20 shows the microstructure for a magnetic recording surface, illustrating the separation of the grains by the amorphous CoCr phase and, in the lower image, the c-axis stacking of the hcp grains [49.37]. In five-element alloys (CoCrPtBTa), Ta assists the segregation of Cr at the grain boundaries [49.38]. Some recording media use as many as three magnetic layers with the saturation magnetization increasing towards the read head, so that the average saturation magnetization is reduced while the magnetization nearest to the read head is increased.

The layer of films comprising a recording layer at the areal densities for a 130 Gbit/in^2 case were: NiP-plated aluminum substrate with mechanical texturing in the data-recording zone/Cr seed layer/a CrMoX (X = grain-boundary segregation-enhancement element) underlayer/a CoCrT intermediate layer/two ferromagnetic layers (one with high Cr/low B concentration and the second with low Cr/high B concentration) and a double-layer amorphous carbon overcoat [49.39].

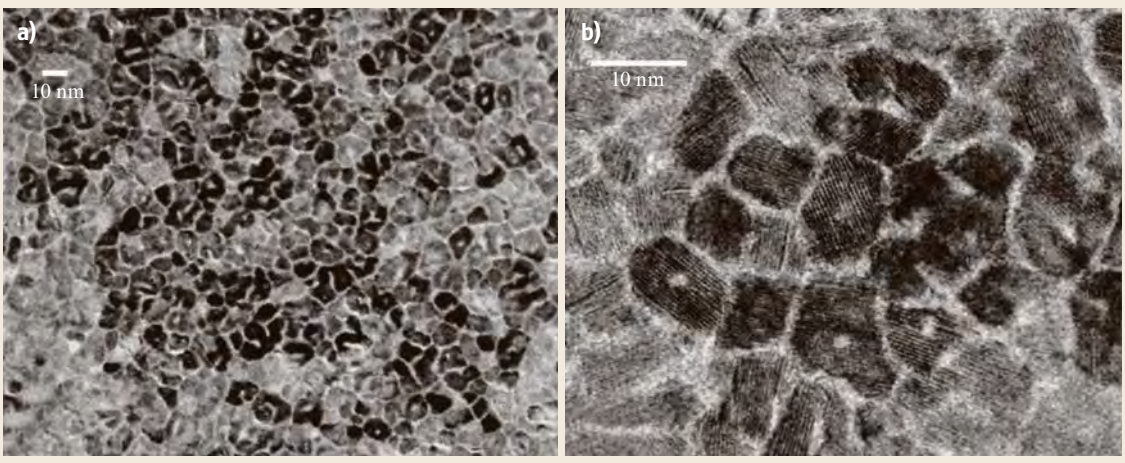


Fig. 49.20a,b Plan-view electron microscope images ((**a**) is lower resolution and (**b**) is higher resolution) of the CoPtCrB magnetic alloy layer used in the 35 Gbit/in^2 case made by IBM. The amorphous grain boundaries are seen as *white* and the *c*-axis stacking of the hcp grains can be seen within the grains in the higher-resolution image. (After [49.37])

The substrate used for the samples in Fig. 49.20 was glass. Glass substrates have several advantages over aluminum with electroless Ni(P) (for hardness) substrates, including increased stiffness, which leads to reduced flutter, increased hardness and reduced propensity to corrosion. However, aluminum substrates continue to be used in nonportable applications because of reduced cost.

The amorphous carbon is required to protect the magnetic film and to give a mechanically rigid surface on which the slider can fly at less than 25 nm.

Media Noise

In digital recording systems the dominant noise voltage, which appears simultaneously with the desired read signal at the terminals of the read head, is that from the recording media, assuming that the resistance from the read head is not excessive. Noise in the read channel is caused by random variations in the location of the recorded transitions.

Transitions are not straight across the recorded track, as we assumed in our discussion of the transition parameter, but instead form zig-zag patterns in order to reduce the large demagnetizing field and demagnetizing energy associated with the longitudinal transition (discussed in [49.5, 12]). Transitions also tend to follow the grain boundaries in the polycrystalline media. To achieve low media noise, it is required to have a small grain size so that the random nature of the grains is averaged over the width of the read head (the track width). An analysis of the signal-to-noise ratio, defined as the base-to-peak signal voltage amplitude (S_0) over the media noise voltage amplitude (N) has been published [49.40]

$$\left(\frac{S_0}{N}\right)^2 = 0.31\left(\frac{B}{a}\right)^2\left(\frac{\mathrm{PW}_{50}}{B}\right)\frac{W_\mathrm{r}}{W_\mathrm{c}}\,, \tag{49.14}$$

where W_r is the width of read head, PW_{50} is the width of the read pulse at 50% of the peak amplitude, W_c is the cross-track correlation width, which has been approximated as the grain or cluster size of the recording media. The grain size has a log-normal distribution and W_c is the mean of the distribution, a is the transition parameter discussed above, and B is the length between bits.

For a reliable low error rate the value of $20\log S_0/N$ is required to be above 26–27 dB and this has required an increasingly smaller grain size as the width of the read head has decreased, resulting in higher track densities. For a 10 Gbit/in^2 system, the value of W_r was 0.505 μm and the required mean of the grain size was 12 nm [49.41].

The ratio B/a has been ≈ 3 to avoid severe distortion of the transition across the track, referred to as percolation (discussed in [49.5]) and the ratio PW_{50}/B has been in the region 2.5–2.8 for the type of detection channels used in disk drives (partial response maximum likelihood – PRML).

Therefore, to maintain constant S_0/N the ratio of the track width to the grain size must be kept constant and, with the decreased track width, the grain size must decrease in proportion. Increased areal density has required decreased mean grain size but, equally important, tighter distributions of grain size. The distributions are closely matched by log-normal distributions of the form

$$y = \frac{1}{\sigma_\mathrm{Z}\sqrt{2\pi}}\exp\left(-\frac{(Z-\bar{Z})^2}{2\sigma_\mathrm{Z}^2}\right), \tag{49.15}$$

where Z is the $\log_{10}$ value of the grain diameter, $\bar{Z}$ is the mean value of Z and σ_Z is the standard deviation of Z.

Grain-size distributions are measured on foils by tunneling electron microscopy (TEM). Figure 49.20 shows the microstructure of a typical recording surface, emphasizing the segregation of the grains by the nonmagnetic constituents of the alloy.

The distribution of the grain sizes in recording media for 20 GB/platter (counting both sides of the platter) and the higher density 80 GB/platter are shown in Table 49.3 [49.7]. Both the mean and standard deviation for the 80 GB/platter are reduced from those of the 20 GB/platter media.

We will discuss the decay rates shown in Table 49.3 in the next section. The trend in the mean grain size and grains per bit for recording films versus areal density is shown in Fig. 49.21 [49.7]. A serious problem with this trend is the fact that the ratio of magnetic to nonmagnetic volume in the film is decreasing faster than the average grain size, which results in the number of grains per bit decreasing faster than the average grain size. The price to be paid for the increasingly small grain size is the stability of the recorded data over time as we discuss in the next section.

Table 49.3 Comparison of 20 and 80 GB/platter disk characteristics. (After [49.7])

	20 GB/90 mm disk	80 GB/90 mm disk
H_c (Oe)	2700	4200
$M_\mathrm{r}\delta$ (memu/cm^2)	0.46	0.32
Average grain size (nm)	10.8	9.1
Distribution sigma	0.41	0.24
Amplitude decay rate (%/decade)	0.7	0.4

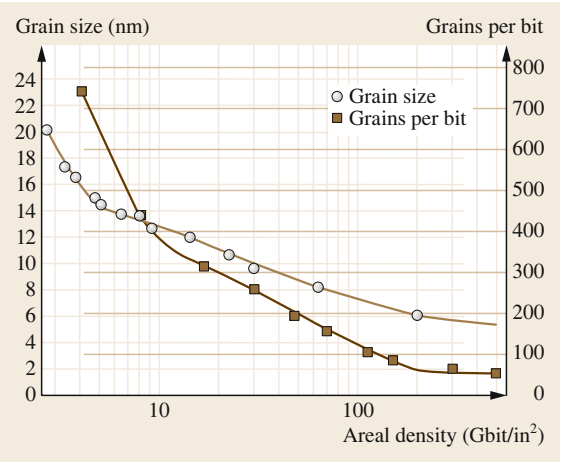


Fig. 49.21 Grain size and grain/bit for sputtered cobalt alloys as a function of areal density in Gbit/in^2. (After [49.7])

To investigate how far the areal density can be extended with LMR, use has been made of the signal-to-noise ratio calculation to see if an areal density of 200 Gbit/in^2 is feasible. To ensure that the signal-to-noise ratio does not degrade when extending the areal density from 60 to 200 Gbit/in^2, it is required that: (a) the transition parameter be reduced from 12.8–6.2 nm, to achieve which the coercive field (H_c) must be increased from 4000–5000 Oe, and (b) the product of the magnetization and thickness ($M_r\delta$) be reduced from 0.32–0.2 memu/cm^3. In addition, the HMS must be reduced from 30–15 nm [49.7].

Superparamagnetic Effect

To maintain the growth in areal density as shown in Fig. 49.1, it is necessary to reduce the physical dimensions of the recording system components and to maintain the signal-to-noise ratio at levels required for satisfactory error rates. This includes reducing the mean grain size of the recording media. However, at some grain size the thermally driven fluctuations of the orientation of the magnetization in the isolated grains result in increased probability of the magnetization being switched from the desired orientation, as initially recorded, to the opposite direction. This instability of the magnetization is the superparamagnetic effect. The magnetic energy in an isolated grain is K_uV, where K_u is the uniaxial anisotropy consisting of crystalline, magnetostriction and shape contributions whereas $V = (\pi(D/2)^2\delta)$ is the volume of the particle or grain with diameter D.

Thermal energy supplied from the environment to the particle is k_BT, where k_B is the Boltzmann constant (0.8619×10^{-4} eV/K). An analysis of

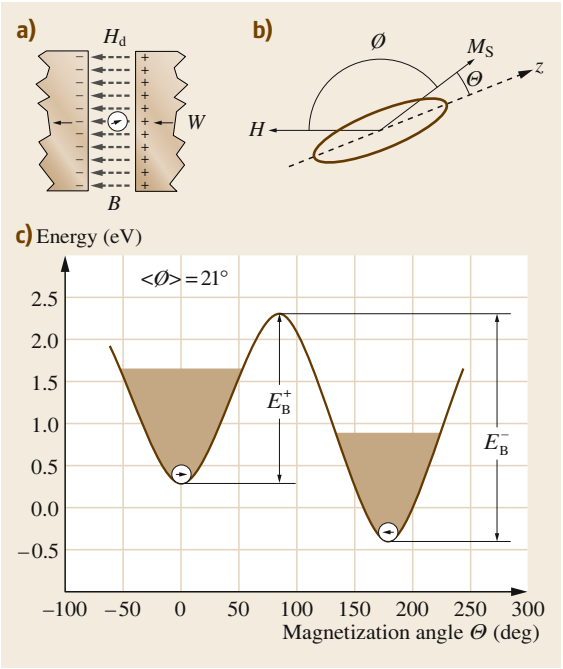


Fig. 49.22 Calculated energy **(c)** of an isolated single-domain particle **(b)** at the center of an 80 nm-long dibit **(a)** (double transition). The properties of the particle are listed in Table 49.4. The ratio of the barrier height to thermal energy is 43 (see text) and the height $43k_BT$ is shown *shaded*. There is margin between the top of the shaded region and E_B^+ and the particle is stable for the time t_x of 10 y. (After [49.42])

the superparamagnetic effect using an isolated small single-domain asymmetrical particle is summarized in Fig. 49.22 [49.42]. Here, the magnetic particle is assumed to be in the center of two transitions and the energy contour results from an analysis using only shape and crystalline anisotropy (combined into K_u) and magnetostatic energy terms

$$E(\theta,\phi) = K_uV\sin^2\theta + HM_s\cos(\theta-\phi)\,. \qquad (49.16)$$

It is assumed in Fig. 49.22 that the angle ϕ has a mean value $\phi = 21°$. The energy contour shown in Fig. 49.22 results from the application of (49.15) with the parameters listed in Table 49.4. The energy required to switch the particle from the initially magnetized direction is shown as E_B^+. In the absence of an external or demagnetizing field, the energy barrier is $E_B^+ = K_uV$.

With a demagnetizing field (H_d) the switching field is reduced since the field is in the direction of the switched magnetization

$$E_B^+ = K_uV\left(1-\frac{|H_d|}{H_0}\right)^{\frac{1}{2}}\,. \qquad (49.17)$$

Table 49.4 Energy of single-domain particle in a magnetic field of two transitions. K_u is particle anisotropy energy, consisting of shape and crystalline anisotropy contributions, H_d is the demagnetizing field between transitions and ϕ is mean field angle with respect to grain or particle axis. A dibit is a double transition corresponding to a double change in magnetization orientation. (After [49.42])

Parameter	Value
B (dibit length)	80 nm
K_u	1.5×10^6 erg/cm^3 (1.5×10^5 J/m^3)
$M_r\delta$	0.39 memu/cm^2 (3.9 mA)
M_s	375 emu/cm^3 (375×10^3 A/m)
D (grain/particle diameter)	15 nm
δ (media thickness)	13 nm
V (grain/particle volume)	$\pi(D/2)^2\delta$
H_d	780 Oe (62 kA/m)
ϕ	21°

With a random two-dimensional (2-D) system $H_0 = 2K_u/M_s$. The magnetization of the particle will switch when $H_0 = H_d$. When the magnetization of the particle is switched, the new energy barrier is E_B^- and is increased in depth since the demagnetizing field is now in the same direction as the magnetization. Using the set of parameters listed in Table 49.4, the depth of E_B^+ (with a demagnetizing field of 780 Oe (62 kA/m)) is approximately 1.9 eV.

The time constant for the thermally driven switching of the magnetic particle in the presence of a magnetic field, which in the case under consideration is the demagnetizing field acting on the magnetic particle, can be found using an Arrhenius–Néel model

$$\tau = \tau_0 \exp\left(-\frac{E_B^+(H)}{k_B T}\right), \tag{49.18}$$

where $\tau_0 = 1/f_0$ and f_0 is an *attempt frequency*, usually taken as $\approx 10^9$ Hz [49.42]. The energy barrier depends on the thermal energy and the fraction x ($0 < x < 1$) of the retained magnetization and the storage time t_x

$$E_B^+(H) = k_B T \ln\left(\frac{t_x f_0}{|\ln x|}\right),$$
$$t_x = |\ln x|\, \tau_0 \exp\left(\frac{E_B^+(H)}{k_B T}\right). \tag{49.19}$$

As an example, if the required storage time is $\approx 10^8$ s (10 years) and the percentage loss of data amplitude is 5%, then $x = 0.95$ and $|\ln 0.95| = 0.05$ and

$$\frac{E_B^+(H)}{k_B T} \approx 43\,.$$

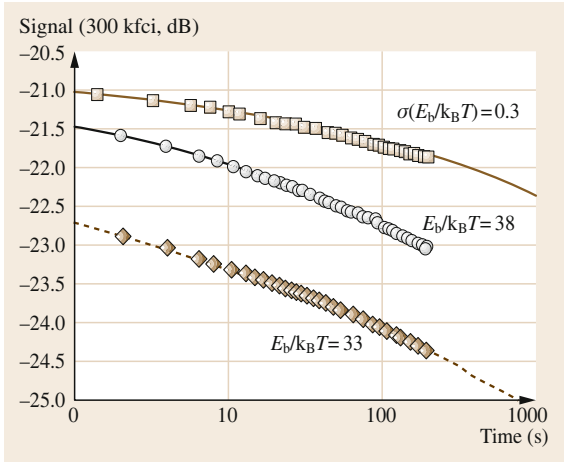


Fig. 49.23 Signal loss versus time at room temperature and 300 kflux changes per inch (kfc/in) for three magnetic recording surfaces: The first two (*from the top*) are SFM or AFC media with layer thicknesses of: [12 nm/0.7 nm/4 nm] (*open square*), [12 nm/1.4 nm/4 nm] (*open circle*). The lower curve is for a 12 nm single layer (*solid diamond*). (After [49.43])

The thermal energy $k_B T$ at 350 K (the assumed operating temperature of a typical disk drive) is 0.03 eV, 43 $k_B T \approx 1.29$ eV and the shaded region in Fig. 49.22 has this energy level. Since there is some margin between the thermally driven contribution to switching and the top of the barrier, the particle is stable (over the 10 year period). However, for smaller volumes (V) the value of E_B^+ decreases and the particle could become unstable.

For example, if the thickness of the recording media is reduced from that assumed in Fig. 49.22 (13 nm) to 6 nm, the value of E_B^+ is reduced to ≈ 0.9 eV and the thermal energy would easily switch the particle. A dramatic loss of data over time is observed with recording media with media flux ($M_r\delta$) less than about 0.25 memu/cm^2. Figure 49.23 shows amplitude loss of recorded data for three different media at a linear density of 300 kfc/in (kflux changes per inch) at room temperature [49.43]. The three media are: a single-layer of CoCrPtB (lowest curve with a thickness of 12 nm); and two recording surfaces we will discuss in the next section. The signal decay rate R is defined as

$$R(H_d, T) = 100\frac{\left(\frac{A(t)}{A(t_0)} - 1\right)}{\log\left(\frac{t}{t_0}\right)}, \tag{49.20}$$

where the signal amplitude of a square-wave pattern is $A(t)$. With $t_0 = 1$ s, the decay rate per decade for the

media shown as the lower curve in Fig. 49.23 is found to be −7.19%.

The decay rate for the other media will be discussed in the next section. The decay rates for the two recording films shown in Table 49.3 are −0.7%/decade and −0.4%/decade for the 20 GB/platter and 80 GB/platter media respectively. In a conventional disk recording system this superparamagnetic effect was judged to limit the areal density to somewhere in the range 36–100 Gbit/in^2 (reviewed in [49.42]). Formally the superparamagnetic limit to areal density in LMR is set at 100 Gbit/in^2.

The superparamagnetic effect and the requirement for low media noise result in recording-system designers being caught in two inconsistent requirements: first, a small grain size for the recording media is required for adequate signal-to-noise ratio, so that the number of grains per bit remains constant (49.13), and second, a large grain size is required for small data loss over time because of the superparamagnetic effect ((49.16) and (49.17)). The requirement to reduce the thickness of the recording media can be deduced from the need for low values of $M_r\delta$ for short transition lengths (49.5). The approach of increasing the anisotropy of the grains to increase K_uV to resolve the superparamagnetic effect is limited by the ability of write heads to record on the media with the resulting increased media coercivity.

Table 49.3 shows that the decay rates observed with the smaller grain size/higher density media is actually smaller than that for the larger grain size/lower density media. A possible explanation is that the K_uV values for the smaller grains are higher than for the larger grains [49.7]. In the next section we discuss a novel approach that has extended the progression in areal density.

Antiferromagnetically Coupled (AFC) or Synthetic Ferrimagnetic (SFM) Media

The invention of a novel recording media decreasing the impact of the superparamagnetic effect was accomplished simultaneously by teams of researchers at the Fujitsu and IBM Almaden Research Laboratories [49.44, 45].

The recording media consists of two ferromagnetic layers coupled by a thin layer of ruthenium (antiferromagnetically coupled – AFC or synthetic ferrimagnetic media – SFM), which results in an antiferromagnetic coupling of the two films, similar to that used in the spin valve for a synthetic antiferromagnet. In order for the coupling between the two ferromagnetic layers to be antiferromagnetic, the thickness of the ruthenium layer, referred to by IBM as *pixie dust*, is required to be approximately 0.6 nm [49.45]. A schematic of the antiferromagnetic coupled media and the resulting hysteresis curve is shown in Fig. 49.24 [49.43]. The sharp reduction in the remanent magnetization at ≈ 80 Oe is due to the switching of the thinner lower layer (M_rt_2) and the lowered value of the remanence is advantageous for reduced transition noise and transition length.

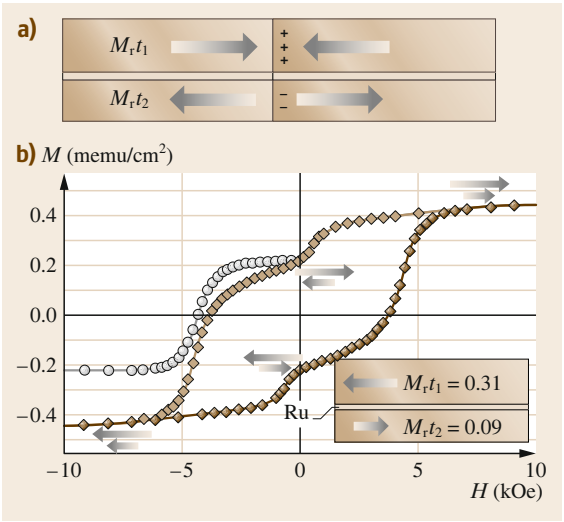


Fig. 49.24 (a) Schematic of the antiferromagnetically coupled (SFM or AFC) media showing the two ferromagnetic layers coupled by a thin layer of ruthenium (*pixie dust*) and with a single magnetic transition for which the magnetizations in the two layers are oppositely directed. **(b)** Magnetic hysteresis curve for the SFM or AFC media with remanent thickness products for the two ferromagnetic layers shown in the inset and measured at 350 K. The *filled triangles* are the major hysteresis curve and the *filled circles* are the remanence hysteresis curve. The *arrows* show the directions of the magnetization in the two films. (After [49.43])

The curve with solid circles is a remanence hysteresis curve measured by subjecting the sample to a negative reversing field and then increasing the field in a positive direction to zero and plotting the remanent magnetization. This measurement is an attempt to simulate what happens when a recording medium leaves the vicinity of the recording head.

By analyzing the read pulse from the transition in the upper layer combined with the read pulse from the transition with reversed sign from the lower layer (the layers are antiferromagnetically coupled), it can be shown that no degradation of the resolution (PW_{50}) of the system results from having two as opposed to one recording layer [49.45]. The advantage of the AFC or SFM configuration is that the medium is in fact thicker than would be required for the given recording density using a single layer and hence is less susceptible to the superparamagnetic effect (49.17).

The effective volume is bounded by $K_u V_1 < K_u V_{\text{eff}} < K_u V_1 + K_u V_2$. In the case of tight coupling and alignment of grain axes, the upper limit can be approached. The effective remanent magnetization-thickness ($M_r\delta$) product while close to $M_r t_1 - M_r t_2$, reduces the pulse width over that of a single-layer recording medium (49.10). Because of the antiferromagnetic coupling, an exchange magnetic field $H_{ex,2}$ acts on the lower magnetic film

$$H_{ex,2} = \frac{J_{ex,2}}{M_2 t_2}$$

and a typical value for the antiferromagnetic exchange constant $J_{ex,2}$ is $0.06\,\text{erg/cm}^3$ ($6\,\text{mJ/m}^3$) [49.46].

For the magnetization of the lower layer to assume the opposite direction to that of the top layer, before the field H is reduced to zero, it is required that $H_{ex,2} > H_{c2}$. The antiferromagnetic exchange constant can be increased by adding thin layers of cobalt or cobalt alloys known as *E layers* to both sides of the ruthenium layer [49.47]. The increase in the antiferromagnetic exchange constant is rapid with increasing thickness of the *E layers* and we show in Table 49.5 a listing of J_{ex} for several *E layers* [49.47].

The largest increase in J_{ex} is with cobalt, but this material suffers from significant degradation on overwrite and poor signal-to-noise ratio compared with cobalt alloys. To illustrate that the AFM or SFM media configuration actually results in reduced amplitude loss over time, we show in Fig. 49.23 the amplitude loss with two SFM media: 12/0.7/4 (open squares) and 12/1.4/4 (open circles), where the numbers are the thickness of the three layers in nm and the value of E_B/k_BT is 38 [49.43]. The signal decay rate is -4.14% per decade for the top curve (open squares) and -6.45% per decade for the middle curve (open circles) [49.43].

The coupling between the two magnetic films is antiferromagnetic for the top curve and ferromagnetic for the middle curve and the lower curve is a single magnetic layer. The AFM or SFM media clearly reduces the decay rate. The improvement is even more significant with a larger value for E_B/k_BT and values of R less than 1%/decade with $M_r t$ values lower than $0.2\,\text{memu/cm}^2$ have been reported by IBM researchers using static magnetization measurements with a reversing magnetic field to mimic the demagnetizing field [49.45].

AFM media started out being used in IBM portable disk-drive applications [49.45]. Areal densities larger than $100\,\text{Gbit/in}^2$ have been used with LMR using AFM media [49.48].

The magnetic recording layers were alloys of CoCrPtB with an average grain size of 9.4 nm. The write head was trimmed on the air-bearing surface to the final track width using a focused ion beam and the track density was 149 ktracks/in and the bit density was 720 kbits/in [49.48].

The data shown in Table 49.3 also shows that it is possible to achieve low decay rates with single recording layers [49.7]. Analysis comparing SAF and single-layer media has shown that the single layer media is preferred unless the media flux drops below about $0.25\,\text{memu/cm}^2$ with H_c values constrained below 4300 Oe. If recording heads are able to record with adequate overwrite on media with coercivity greater than 4300 Oe, then the single-layer media is preferred [49.49].

Table 49.5 Comparison of materials for *E layers*: J_{ex} values and degradation in overwrite and signal to media noise. Note that negative dB values of overwrite degradation indicate improved overwrite. Both degradations in overwrite and S/N_m are considered without *E layers*. (After [49.47])

Material for *E layers* (all 1 nm thick)	Co	Co-Cr	Co-Cr-X
J_{ex} (erg/cm^2)	0.73	0.59	0.24
Degradation in overwrite (dB) with respect to SFM media	10	≈ 3	≈ -6
Degradation in $(S/N)_m$ (dB) with respect to SFM media	≈ 4.5	≈ 1.6	≈ 1.5

Dynamic Coercivity

Another manifestation of thermal effects on magnetic properties is a change in coercivity with switching time. With long periods of time the probability that a particle will switch is increased by thermal excitation and with short periods this probability is decreased, leading to a change in the coercivity of the particle. An experimental study of this effect was done on a test stand in which a stationary write head is driven by a current pulse of variable duration over a previously magnetically saturated recording medium. The current pulse records two transitions (a dibit). The width of each transition is essentially the gap of the write head and the length of the reversed magnetization region varies with the pulse width (t_p). The dibit is read with a magnetoresistive sensor moved across the recorded transitions resulting in a dibit pulse. By measuring the amplitude of the dibit signal as a function of the pulse width and varying the magnitude of the reverse field and measuring $H_{head}(I)$ (head field versus current) it is possible to deduce the remanent coercivity (H_{CR}) as a function of t_p. The resulting curves of H_{CR}/H_{CR} (1 ns) are shown in Fig. 49.25 [49.42]. The experimental slope C increases as the thickness decreases.

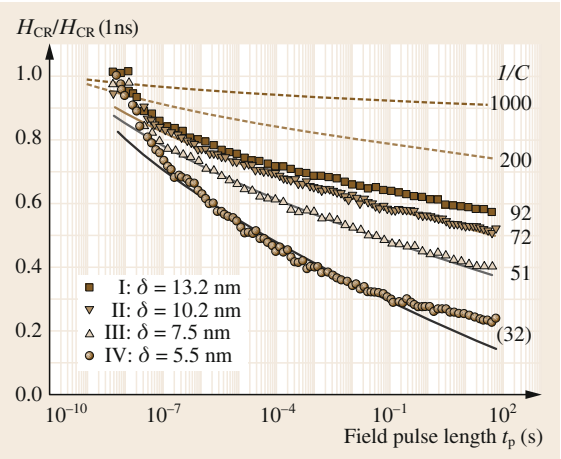


Fig. 49.25 Dynamic coercivity $H_{CR}(t)$ normalized by its value at $t = 1$ ns for a series of $Co_{64}Pt_{14}Cr_{22}$ films with variable thickness. (After [49.42])

The inverse slope parameter is given by

$$\frac{1}{C} \approx \frac{K_u V}{k_B T} ,$$

which is the superparamagnetic stability parameter discussed above and it is observed that, even with a stability factor larger than 51, for which the superparamagnetic region is avoided, the coercivity increases significantly for short switching times. This is the actual coercivity that must be exceeded by the write-head magnetic fields to switch a recording film.

49.1.5 Perpendicular Magnetic Recording

The technology that supplanted LMR is perpendicular magnetic recording (PMR). Perpendicular recording offers a way out of the conundrum caused by the need for thinner recording media for improvements in recording density and the need for larger particles to reduce the impact of superparamagnetism.

While smaller particles are needed to increase areal density in LMR, particle size in PMR has kept its value around 7–9 nm since it was introduced for the first time in 2006.

In perpendicular recording the magnetization in the recording media is held perpendicular to the surface of the recording media by a perpendicular anisotropy large enough to overcome the large demagnetizing field with this orientation of the magnetization. A schematic of a perpendicular recording system including a probe write element and a magnetically soft underlayer (SUL) is shown in Fig. 49.26. It is also possible to have a perpendicular recording system with a ring write head and a single-layer media, but this approach has not been developed.

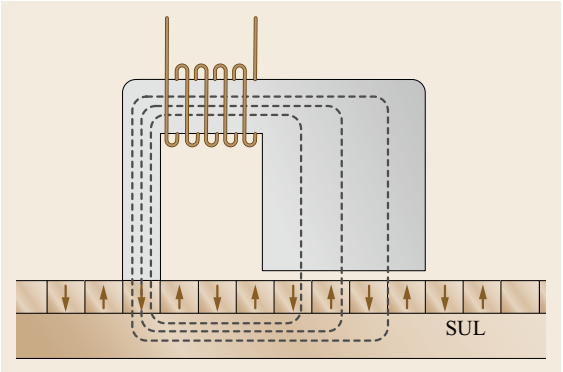


Fig. 49.26 Schematic of a perpendicular recording system. (After [49.5])

In contrast to LMR, the SUL is required for a low-reluctance path for the magnetic flux generated by the coil on the write element. One advantage of perpendicular recording is that the magnetic field for recording is larger than that for LMR since the magnetic field from the write pole is in the gap of the write head, in contrast to LMR where the field leaks from the write gap. A schematic of the total geometry for LMR and PMR is shown in Fig. 49.27 [49.1]. Reading for PMR is accomplished with a GMR sensor as with LMR evolving later into tunnel magnetoresistance (TMR).

Historically, when areal density was around 100 Mbit/in^2 (at the beginning of the 1990s) the read sensor was based on simple magnetoresistance. In the mid-1990s it evolved into a spin-valve GMR sensor for areal densities larger than 1 Gbit/in^2. In 2006 TMR valves were used for areal densities larger than 100 Gbit/in^2 and in 2011, CPP-GMR read heads started to be used for areal densities close to 1 Tbit/in^2. Typically a TMR head is about 70 nm wide whereas a CPP-GMR head is about 30 nm in width.

The major difference in the write head is that it uses a single pole to generate the perpendicular magnetic field with a wider return pole (to reduce reluctance). The write head for perpendicular recording is imaged in the SUL and this results in an increase in the strength of the magnetic field by approximately a factor of two. In addition, due to imaging effects, the effective number of turns on the write head is doubled for perpendicular recording. It is important for perpendicular recording that the magnetization of the SUL magnetic material be larger than that for the pole of the write head [49.50].

This is the case to prevent the SUL from saturating before the pole tip saturates, resulting in a decrease in the gradient of the head field, which would result in an increased transition length (to be discussed). A large

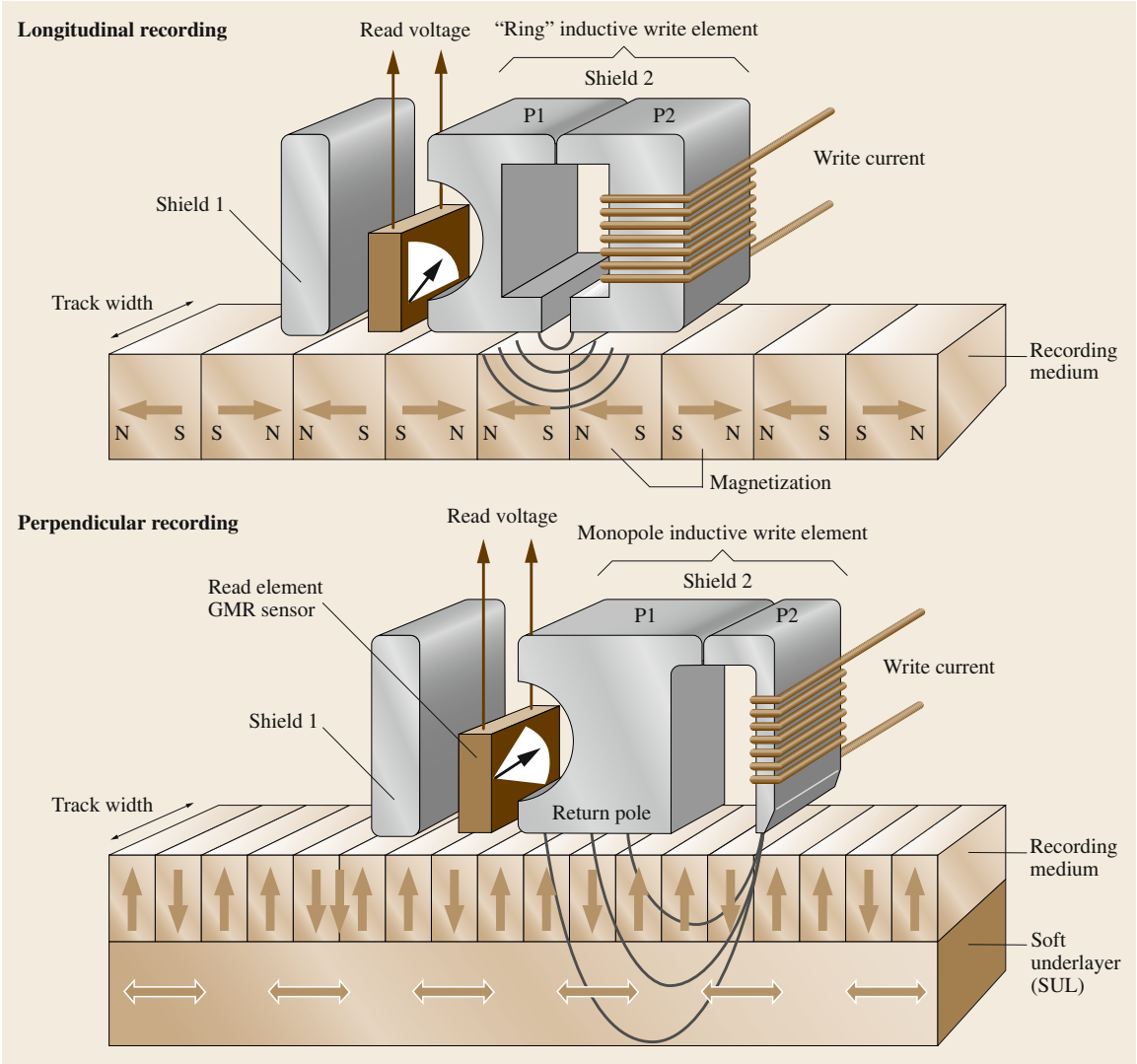


Fig. 49.27 Schematic of longitudinal and perpendicular recording systems. (After [49.1])

magnetic field and gradient of magnetic field from the write head can record media with a larger coercive field than for LMR, which results in a reduced transition length. Perpendicular-recording media can be made thicker than for longitudinal media for the same areal density. The media thermal stability factor ($K_u V/k_B T$) is increased both by an increase in K_u by virtue of the increased coercive field but also by the increase in the volume (V) with the thicker media. Reading of the magnetic transitions with perpendicular recording is done with a spin-valve head as with LMR – the difference being that with perpendicular recording the response is a step instead of a pulse (to be discussed).

A significant development for perpendicular write heads is the addition of a trailing-edge shield to the head that results in an improvement in the concentration of the writing flux and an increase in the gradient of the head field (dH_y/dx), which reduces the transition length [49.51, 52]. A schematic of a write head with a trailing-edge shield is shown in Fig. 49.28 [49.53]. The shield connects to the larger return structure for the head shown in Fig. 49.27.

The particular geometry and saturation magnetization of the main pole and the front yoke or trailing-edge shield were used in a two-dimensional finite element model to calculate the head fields and shape of the recorded transition [49.53]. Results of the simulation of transition shape are shown in Fig. 49.29, where it is seen that the recorded transition is considerably sharper with the head with the front yoke compared with the head

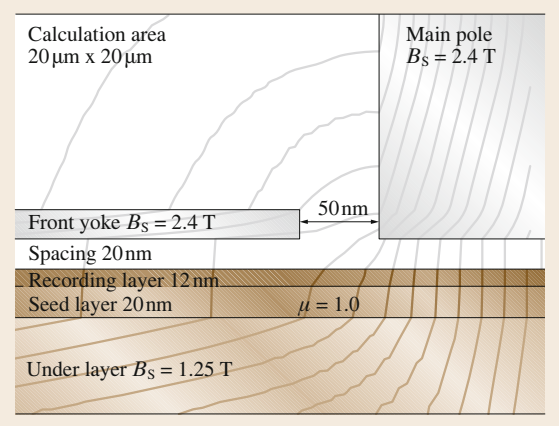


Fig. 49.28 Schematic of a perpendicular recording head including the main pole and the front yoke or trailing-edge shield. (After [49.53])

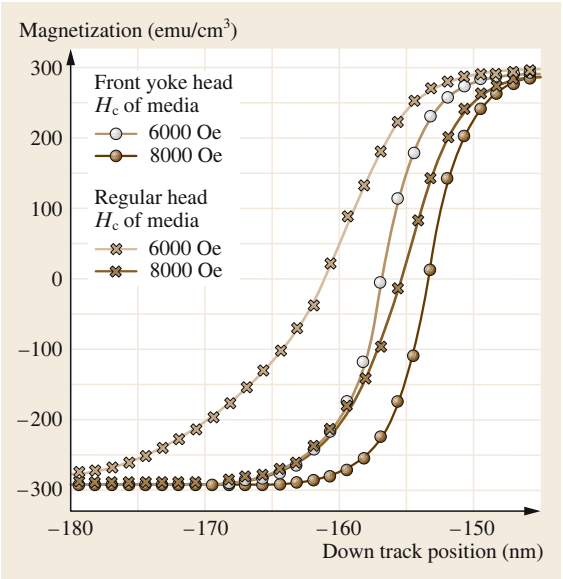


Fig. 49.29 Magnetic transitions in perpendicular recording with and without the front yoke or trailing-edge shield. (After [49.53])

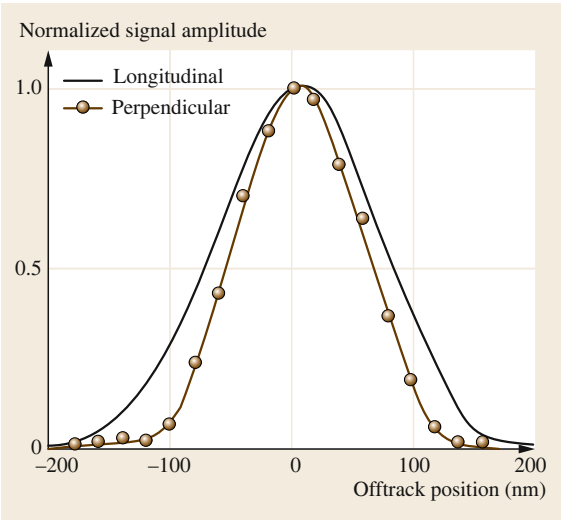


Fig. 49.30 Cross-track profile of the signal with perpendicular and longitudinal recording. (After [49.52])

without the shield, particularly with the higher coercivity media ($H_c = 8000$ Oe).

The perpendicular magnetic field has the amplitude reduced to one-half at a distance of $0.8d_0$ with the shield and $1.3d_0$ without the shield (assuming that the spacing of the shield is separated from the main pole by a distance d_0) [49.52]. The shield shunts some of the write flux, reducing the amplitude of the write field, so it is important to keep the height of the shield short. Another advantage of the trailing-edge shield is that a horizontal component of the field is added. This field reduces the field for switching since, if the angle of the field is θ, the switching field is

$$H_{\text{switch}} = \frac{H_k}{\left[(\cos\theta)^{\frac{2}{3}} + (\sin\theta)^{\frac{2}{3}}\right]^{\frac{3}{2}}},$$

where H_k is the anisotropy field of the perpendicular medium. The denominator of this equation is always larger than one and hence the switching field is reduced for angles θ larger than zero degrees. Note that this equation is the switching astroid equation used to explain the operation of the MRAM.

A significant advantage of PMR as compared with LMR is the narrower off-track read profile as shown in Fig. 49.30 [49.52]. This improvement in read profile leads directly to increased track density. There are limitations to the achievable track density in disk drives where there is significant skew of the head with respect to the track center line. Head skew results from the fact that disk drives use rotary actuators and the angle of the written transition with respect to the track center line varies as the actuator is rotated across the disk, being largest at the inside and outside diameters. With head skew it is possible for the write element of the head to be placed over the guard band or adjacent track of a selected track, resulting in unwanted recording in these areas. This problem is reduced by using trapezoidal poles [49.54]. Shields can be added to the side of the head to reduce this erasure of adjacent track information; however, the processing of such shields is complex.

Media for Perpendicular Recording

The magnetization and anisotropy in a perpendicular recording film is shown in Fig. 49.31. The energy of the magnetization in the presence of the demagnetizing, anisotropy and magnetostatic energies (Gaussian units) is

$$E = 2\pi M_s^2 \cos^2\theta + K_u \sin^2\theta - H_{ext} M_s \cos\theta \,, \tag{49.21}$$

where the demagnetizing factor for a thin film magnetized normal to the surface is 4π. The equilibrium angle is given by the solution to

$$\frac{dE}{d\theta} = -2\pi M_s^2 \sin 2\theta + K_u \sin 2\theta + H_{ext} M_s \sin\theta = 0 \,. \tag{49.22}$$

For stability

$$\frac{d^2E}{d\theta^2} = -4\pi M_s^2 \cos 2\theta + 2K_u \cos 2\theta + H_{ext} M_s \cos\theta > 0 \,. \tag{49.23}$$

The solution is $\sin\theta = 0$ ($\theta = 0°$ or $180°$) and the stability condition for $H_{ext} = 0$ results in $2K_u > 2\pi M_s^2$.

Alloys under consideration for perpendicular recording include those listed in Table 49.6 [49.55], and Tables 49.7 and 49.8. A review of work on perpendicular recording media was published [49.56]. It is observed from this table that these candidate materials satisfy the condition (49.22). The critical magnetic field below which the $\theta = 0°$ solution is no longer stable is given by the solution to (49.21) with $H_{ext} \neq 0$, resulting in

$$H_e^*(\theta = 0°) = -\frac{(2K_u - 4\pi M_s^2)}{M_s} \,,$$
$$H_e^*(\theta = 180°) = \frac{(2K_u - 4\pi M_s^2)}{M_s} \,. \tag{49.24}$$

The solution to (49.21) with $\sin\theta$ factored out is

$$-4\pi M_s^2 \cos\theta + 2K_u \cos\theta + H_{ext} M_s = 0 \,. \tag{49.25}$$

Thus

$$H_{ext} = \left(4\pi M_s - 2\frac{K_u}{M_s}\right)\cos\theta$$

and

$$\frac{M_\perp}{M_s} = \cos\theta = \frac{H_{ext}}{\left(4\pi M_s - 2\frac{K_u}{M_s}\right)} \,.$$

This solution applies when the magnetization curve is not saturated. Note that H_e^* is not the same as the coercive field where

$$M(H_{ext} = H_c) = 0 \,.$$

The coercive field is found from (49.24) by recognizing that the magnetic field

$$H_{ext} = -|H_e^*| - H' \,,$$

where H' is measured from $-|H_e^*|$. The coercive field is then

$$H_c = -2|H_e^*| \,.$$

The result for a Co-Cr-Pt alloy (Table 49.7) with the following parameters: $K_u = 1.45 \times 10^6$ erg/cm^3, $M_s = 280$ emu/cm^3 is as follows

$$H_e^*(\theta = 0°) = -\frac{(2K_u - 4\pi M_s^2)}{M_s}$$

thus

$$H_e^*(\theta = 0°) = -6.84\,\text{kOe} \quad \text{and}$$
$$H_e^*(\theta = 180°) = 6.84\,\text{kOe} \,,$$

whereas

$$\frac{M_\perp}{M_s} = \frac{(H_{ext} + 2|H_e^*|)}{|H_e^*|} \,,$$
$$H_c = -13.68\,\text{kOe} \,.$$

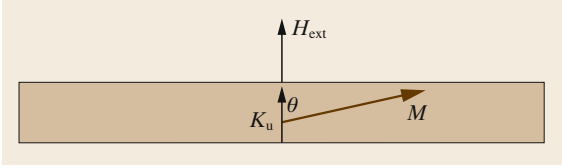


Fig. 49.31 Magnetization and perpendicular anisotropy for a perpendicular recording media. (After [49.5])

Table 49.6 Magnetic properties of selected materials for perpendicular recording. (After [49.55])

Material	H_k (kOe)	M_s (emu/cm^3)	K_u (erg/cm^3 10^6)	$2\pi M_s^2$ (emu/cm^3)^2
Co-Cr	7	270	0.95	0.46
Co-Cr-Pt	10	280	1.45	0.49
Co/Pd	17	350	2.98	0.77
Fe/Pt	28	750	10.5	3.53

Table 49.7 Properties of hard magnetic materials that might be used for PMR [49.57]. All these materials are considered as uniaxial and capable of sustaining $D_p \leq 10\,\text{nm}$ grain sizes over 10 year storage time. Physical parameters include anisotropy field: $H_k = 2K_u/M_s$, domain wall width: $\delta_w = \pi\sqrt{A/K_u}$, domain wall energy: $\gamma_w \approx 4\sqrt{AK_u}$, single particle domain size: $D_c = 1.4\gamma_w/M_s^2$, exchange coupling constant: $A = 10^{-6}\,\text{erg/cm}$, minimal stable grain size: $D_p = (60k_BT/K_u)^{1/3}$ ($\tau = 10$ years) are displayed

Alloy type	Material	K_u (10^7 erg/cm^3)	M_s (emu/cm^3)	H_k (kOe)	T_c (K)	δ_w (Å)	γ_w (erg/cm)	D_c (μm)	D_p (nm)
	CoPtCr	0.20	298	13.7	–	222	5.7	0.89	10.4
Co-alloys	Co	0.45	1400	6.4	1404	148	8.5	0.06	8.0
	Co_3Pt	2.0	1100	36	–	70	18	0.21	4.8
	FePd	1.8	1100	33	760	75	17	0.20	5.0
$L1_0$ phases	FePt	6.6–10	1140	116	750	39	32	0.34	3.3–2.8
	CoPt	4.9	800	123	840	45	28	0.61	3.6
	MnAl	1.7	560	69	650	77	16	0.71	5.1
Rare-Earth	$Fe_{14}Nd_2B$	4.6	1270	73	585	46	27	0.23	3.7
Transition metals	$SmCo_5$	11–20	910	240–400	1000	22–30	42–57	0.71–0.96	2.7–2.2

Table 49.8 Media for perpendicular recording

Media	H_c (kOe)	M_r (emu/cm^3)	α ($4\pi M_r/H_c$)	H_k (kOe)	Magnetic layer thickness (nm)	H_n (kOe)	Comments	References
CoCrNbPt/Ti	2–5	250–450	1.25	6	50	–	Good, poor thermal stability	[49.58]
CoCrPt/Ti/CoZr	2–4	500	2.09	13	25	–	Good SNR, poor thermal stability	[49.59]
$Co_{70.3}Cr_{15}Pt_{14.7}$/Ti/CoZrNb	5.5	460	1.05	–	–	−1.7	Good thermal stability, SNR = 25 dB	[49.60]
$[Co/Pd]_n$	2–10	200–600	0.84	15–30	10–50	–	Good thermal stability, poor SNR	[49.61]
$[Co/PdSi]_{13}$	3.2–5.4	160	1.35	–	–	–	SNR improved	[49.62]
FePt	2–5	800	2.87	> 30	10–50	–	–	[49.63]
CoCrPtB/Ti-CoCrTa/laminated SUL	2.5–3.0	400	1.83	12	–	–	Good SNR, poor thermal stability	[49.64]
CoCrPtO/Ti-Ru/laminated SUL	3.5	520	1.86	–	–	–	Good thermal stability, poor SNR	[49.65]
CoPtCr – SiO_2	> 5	430	1.08	18.6	10–30	–	Good thermal stability, good SNR	[49.66]
Coupled/granular/continuous CoCrPt/CoPt	4–5	400	1.11	–	–	−1 to −3	Good thermal stability, good SNR	[49.67]
Coupled/granular/continuous $CoCr_{18}Pt_{12}$/CoPt	4.02	395	1.23	–	–	−1.8	Good thermal stability. SNR = 31 dB	[49.68–70]

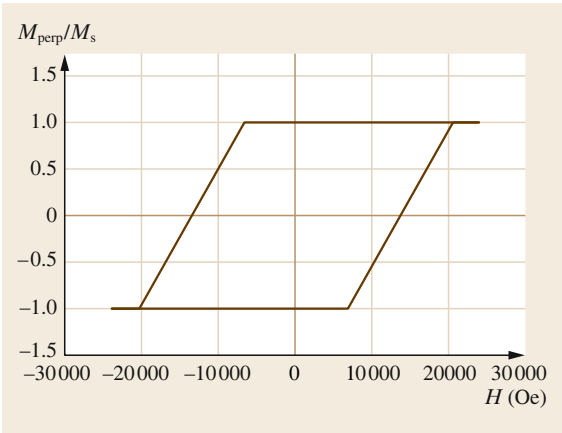


Fig. 49.32 Hysteresis curve for Co-Cr-Pt perpendicular recording media using the analysis in the text

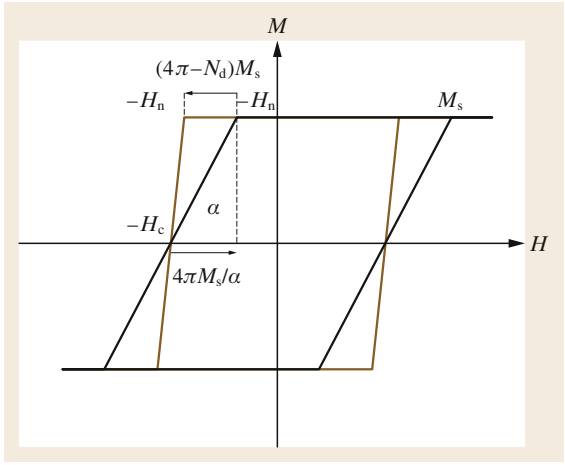


Fig. 49.33 Hysteresis curve with demagnetizing field $(4\pi M_s/\alpha)$ and a reduced demagnetizing factor N' for a perpendicular recording film with loop slope parameter α. (After [49.55])

The magnetization curve for the fourth quadrant is

$$\frac{M_\perp}{M_s} = \frac{(H_{ext} - H_e^*)}{|H_e^*|} - 1 \,.$$

The magnetization curve is shown in Fig. 49.32 and exhibits the desired large coercivity but the curve is sheared due to the large demagnetizing field. In general, the magnetic hysteresis curve (M versus H_{ext}) for the recording medium is as illustrated in Fig. 49.33 [49.55]. Here the slope parameter (α) characterizes the slope of the magnetization curve at $H = -H_c$

$$\alpha = 4\pi \left(\frac{dM}{dH}\right)_{H=-H_c} \,. \qquad (49.26)$$

Finite values of α result from variations in the direction of the perpendicular magnetization, dispersion in the perpendicular anisotropy and variations in the exchange coupling between grains and clusters in the media. The effective demagnetizing field is reduced from $4\pi M$ to $4\pi M/\alpha$. This reduction in demagnetizing field results from the formation of a rectangular checkerboard pattern of domains with high-density linear recording and high track densities [49.55].

The effective anisotropy energy, which assists in overcoming the thermal energy and the superparamagnetic limit, is given in terms of the nucleation field (H_n), at which field the magnetization just starts to reverse, by $K_{eff} = H_n M_s/2$. For high areal density, the reduced demagnetizing factor is $N_d = 0.5\,4\pi M_s$ while the nucleation field is increased to

$$H' = H_n + \frac{4\pi M_s}{2} \,.$$

For high-density recording, large values for α and H_c are required; however, the value of H_c is limited by the ability of the pole head to generate magnetic fields equal to or greater than approximately $1.5\,H_c$, the condition that ensures adequate overwrite. The transition parameter (a) for perpendicular recording can be estimated using the same write-slope criteria discussed for longitudinal recording [49.71]. Here,

$$\frac{dM}{dx} = \frac{dM}{dH}\left(\frac{dH_{head}}{dx} + \frac{dH_d}{dx}\right) \,. \qquad (49.27)$$

This equation is solved by assuming an arctangent transition shape and denoting the slope of the head field by

$$\frac{dH_{head}}{dx} = Q\frac{H_c}{s} \,,$$

where s is the total spacing of the write pole tip to the top of the SUL, consisting of the physical spacing (d), the thickness of the recording media (δ) and the spacing between the SUL and the magnetic recording film (u). H_d is the demagnetizing field. Using H_d for a perpendicular transition [49.72]

$$a = a_1 + \sqrt{\left(a_1^2 + \frac{(1-S^*)\delta s}{\pi Q}\right)} \,,$$

$$a_1 = \frac{4M_t s}{QH_c} + \frac{s(1-S^*)}{\pi Q} - \frac{\delta}{4} \,,$$

$$\alpha = \frac{\frac{4\pi M_t}{H_c}}{(1-S^*)} \,. \qquad (49.28)$$

A comparison between the transition parameters for longitudinal (a_L) and perpendicular (a_P) recording from (49.5) and (49.27) with the parameters listed in Table 49.9 is shown in Fig. 49.34 ((H_c) for longitudinal recording is the dark line and (H_c) for perpendicular

Table 49.9 Parameters for evaluating transition parameters for longitudinal and perpendicular recording

Parameter	a_L	a_P
d (nm)	26	26
t (nm)	–	50
u (nm)	–	0
S^*	0.95	0.98
M_r (emu/cm^3)	200	200
δ (nm)	14	–
Q	0.6	0.6
$H_{c,max}$ (Oe)	3500	8000

recording is dashed). It is seen that the transition parameter for perpendicular recording can be made smaller than that for longitudinal recording if the coercive field for perpendicular recording is larger than 6500 Oe. Note that the thickness of the recording media for perpendicular recording is larger than that for longitudinal recording. The magnetic squareness (M_r/M_s) for perpendicular recording media is larger than that for longitudinal recording.

Gao and *Bertram* have shown a significant advantage for perpendicular recording if the uniaxial anisotropy is not aligned perpendicular to the surface, but oriented at 45° to the perpendicular [49.73].

The geometry for this tilted perpendicular recording is shown in Fig. 49.35. The crystallites comprising the media are still aligned perpendicular but the crystalline anisotropy is assumed to be at 45°. No material system was disclosed that exhibits this tilted anisotropy. With tilted perpendicular recording the crystalline anisotropy field (H_k) can be made significantly larger than if the anisotropy is perpendicular. An example is 15 000 Oe for perpendicular recording and 28 000 Oe for tilted perpendicular recording [49.73]. Some of the advantages of tilted perpendicular recording are: reduced switching field and reduced average grain diameter, resulting in increased signal-to-noise ratio since the number of grains per bit is increased. The requirements for media for ultrahigh-density perpendicular recording have been analyzed [49.74]. Table 49.10 summarizes

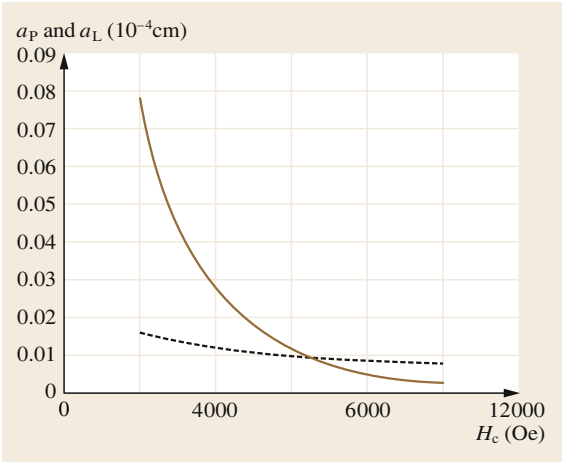


Fig. 49.34 Comparison of transition parameters versus the coercive field (H_c) for perpendicular (*solid line*) and longitudinal (*dashed line*) recording. (After [49.71])

the results for media parameters for areal densities of 100–1000 Gbit/in^2. Several key media parameters are the grains per bit and the coercivity (H_c). An estimate of the signal-to-noise ratio (SNR) using Poisson statistics of additive noise for the magnetic grains is SNR $= -10\log_{10}(N)$, where N is the number of magnetic grains per bit [49.74]. There are two possibilities to realize recording media with a large perpendicular anisotropy and that can approach Table 49.10 parameters: granular media typified by CoCr alloys and continuous multilayer media typified by [Co-Pd]$_n$ multilayers.

CoCr alloys exhibit good SNR but insufficient H_c for ultrahigh recording densities, while the multilayers do not exhibit a satisfying SNR.

Co/PdSi multilayers where the Pd layer is doped with silicon have shown an improvement in SNR [49.62]. The 50+ Gbit/in^2 technology uses a granular CoCr alloy. Tables 49.7 and 49.8 contain several metal alloy used in perpendicular recording. In general, Pt plays a significant role in determining the magnetic

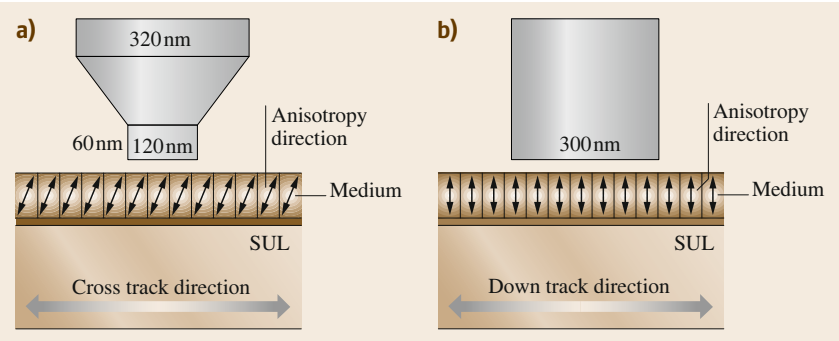


Fig. 49.35 Typical geometry of tilted perpendicular recording. The medium anisotropy directions and not the medium grains are tilted in the cross-track direction. (After [49.73])

Table 49.10 Requirements for perpendicular recording media for ultrahigh recording densities. (After [49.74])

Areal density (Gbit/in^2)	100	250	500	1000
Flying height d (nm)	8	7	6	5
Signal-to-noise ratio (RMS signal/RMS noise)	18	17	16	15
Grains per bit N	63	50	40	32
Average grain diameter D (nm)	9	7	5	4
S^*	1	1	1	1
Media thickness δ (nm)	18	20	22	24
Perpendicular anisotropy K_u (10^6 erg/cm^3)	2	3	4	6
Saturation magnetization M_s (emu/cm^3)	400	550	700	875
Coercivity H_c (kOe)	8.5	11.0	12.5	14.3

anisotropy and Cr, Ta, Nb and B reduce the exchange coupling, resulting in reduced noise.

Another alternative for perpendicular recording media is oxide media [49.66, 75, 76], for example CoCrPtO or CoCrPt-SiO_2.

Oxide media can have their coercivity improved and a negative nucleation field (H_n), a reduced pulse width and reduced signal decay with respect to CoCrPtB media [49.77]. An issue with oxide media is the nonuniformity of the coercivity and oxygen content in the disk surface due to the high gas pressure and oxygen-gas reactive sputtering [49.77]. The last two listings are for a *coupled granular/continuous* (CGC) medium that combines the best properties of the continuous media (high H_c and good thermal stability) and granular media (good SNR) [49.55, 66, 68].

A cross-section of the CGC media is shown in Fig. 49.36 [49.67]. The continuous layer causes an exchange coupling that stabilizes the granular medium; this exchange field is large (≈ 200 kOe) and can result in considerable improvement in thermal stability [49.78].

The initial results for CGC media were encouraging from the standpoint of thermal stability but the high-density SNR was 2 dB less than for the granular media alone. Further optimization of the CGC media using $[CoPt]_n$ multilayers coupled through a Pt buffer layer to a CoCrPt granular layer has shown an improvement in SNR over the granular media alone by 3–5 dB [49.78]. The improvement in SNR is believed to be due to the fact that magnetic recording transitions in the continuous layer are in the form of narrow domain walls and the magnetic interaction between the two layers may result in breaking up the magnetic clusters that form the transitions in the granular medium [49.78].

Magnetic materials used for the soft underlayer (SUL) for a perpendicular recording media must be

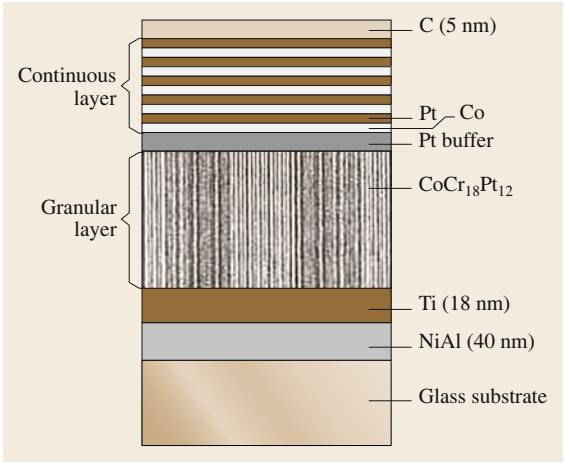


Fig. 49.36 Schematic representation for coupled granular/continuous (CGC) media. (After [49.67])

magnetically soft for low reluctance but also must not exhibit magnetic domains, which can contribute noise to the read signal. The SUL must have a magnetic moment that is larger than that for the pole tip of the write element so that the SUL saturates after the pole tip saturates. This condition assures that there is no excess reluctance directly under the pole tip. The SUL must also be thick enough that the reluctance of the path through which the flux travels allows the maximum magnetic field under the gap. Decreasing the reluctance of the flux path also increases the gradient of the magnetic field from the pole, which reduces the length of the transition parameter (49.20) [49.50]. A guide for selecting the SUL thickness (T_{SUL}) is

$$T_{SUL} \approx \frac{A_{\text{pole tip}}}{C_{\text{pole tip}}} \frac{M_{\text{pole tip}}}{M_{SUL}} , \qquad (49.29)$$

where $A_{\text{pole tip}}$ and $C_{\text{pole tip}}$ are the area and the circumference of the pole tip respectively, whereas $M_{\text{pole tip}}$ and M_{SUL} are the saturation magnetizations of the pole tip of the write head and soft underlayer respectively [49.50]. The effects of changing the thickness of the SUL on the recording parameters of a probe head are shown in Table 49.11 [49.50], where it is seen that the thickness

Table 49.11 Effect of the thickness of the FeAlN soft underlayer (SUL) on recording head parameters. (After [49.50])

Thickness of FeAlN SUL ($4\pi M_s = 20$ kG)	0.1 μm	0.3 μm
Maximum magnetic field (kOe) at recording layer at saturation current (mA) (in parentheses)	16.3 (90)	17.7 (70)
Maximum head field gradient (Oe/nm)	175	550

Table 49.12 Magnetic materials for soft underlayers (SUL). Permalloy has an insufficient magnetic moment for SUL with a field gradient that decreases at maximum field. Sendust also has an insufficient magnetic moment for SUL. FeAlN is sensitive to process conditions and believed to be less useful than FeCoB that must be annealed to reduce stress. CoZrNb has a lower magnetostriction than FeCoB. IrMn antiferromagnetic film is used along with CoZrNb to reduce spike noise due to domain walls

Material	B_s (kG) ($T \times 10^{-1}$)	Relative permeability	References
$Ni_{80}Fe_{20}$ (Permalloy)	10	1500–2000	[49.50]
Sendust (FeSiAl)	10	8000	[49.5]
FeAlN	20	2000	[49.55]
FeCoB	20–24	200–240	[49.83, 84]
CoZrNb	14		[49.84]
$NiFe_{17}Mo_4$			[49.85]
$(Ni\text{-}Fe/Si)_n$			[49.86]
IrMn/CoZrNb	14		[49.87]

and hence reluctance are key to head performance. Candidate materials for the SUL are listed in Table 49.12. A major issue with the soft underlayer is the motion of domain walls driven by the magnetic field from the write head, which can lead to *spike* noise. The last entry in Table 49.12 is for an antiferromagnetic pinning layer to prevent the motion of such domain walls.

Perpendicular Recording Developments

Perpendicular recording is presently used in commercial disk drives. The first was the Toshiba 40 GB, 1.8 in disk drive with an areal density of approximately 75 Gbit/in^2, while later the capacity was increased to 80 GB and the areal density to 133 Gbit/in^2 [49.79].

In 2006, Seagate provided 2.5 in notebook drives with a capacity of up to 160 GB and an areal density of 132 Gbit/in^2. An earlier technology with an areal density of 52.5 Gbit/in^2 was reported in 2001 [49.80] and one of approximately 100 Gbit/in^2 in 2003 [49.81]. Hitachi Global Storage Technologies reported later in 2005 an areal density of 240 Gbit/in^2 using perpendicular magnetic recording and a tunneling current-perpendicular-to-the-plane (CPP) GMR read head (see *Mueller* [49.82]).

We show in Table 49.13 the parameters used for the two areal density configurations.

In both cases a pole write head and a spin-valve read head were used. The parameters for the ≈ 100 Gbit/in^2 case are close to those required for the ultrahigh densities (> 100 Gbit/in^2) as described in Table 49.9, except that the coercivity is about one half that specified.

Table 49.13 Comparison of perpendicular recording parameters for two areal densities using CoCrPt. BER is bit error rate whereas CR is code rate, a measure of coding efficiency given by m/n where m is the number of initial bits and $n > m$ the number of code bits

Areal density	52.5 Gbit/in^2	≈ 100 Gbit/in^2
Type	CoCrPt	CoCrPt
M_s (emu/cm^3)	250	330
H_c (kOe)	2.6	4.94
K_u (erg/cm^3)	H_k = 10 kOe	1.82×10^6
S	0.98	0.92
D (nm)	–	11.9
Underlayer	FeTaC (16 kG) 400 nm	–
Head		
Turns	–	1
Flying height (nm)		6.4
Magnetic spacing (nm)	20	–
Pole		
Material	FeNi (16 kG)	–
Width (nm)	250	105
Thickness (nm)	400	105
Write width (nm)	≈ 250	180
Shield-shield spacing (nm)	80	70
Read width (nm)	200	100
Component performance		
Sensitivity (mV/μm)	3	25
PW_{50}		70
Channel performance		
Channel type	Simulator (CR = 32/33)	Software (CR = 16/17)
Bit density (kb/in)	590	650
Track density (ktracks/in)	88.9	143
Areal density (user/channel) (Gbit/in^2)	52.5 (user)	84 (user) 93 (channel)
On-track BER	10^{-5}	6.7×10^{-5}

Note also that the superparamagnetic limit of PMR is formally set at 1 Tbit/in^2 as it was set previously at 100 Gbit/in^2 for LMR. A cross-section of the write and read elements for the perpendicular recording head used in the 52.5 Gbit/in^2 case is shown in Fig. 49.37 [49.80]. The pole of the write head is identified as the *main pole* and the thick return pole and upper shield of the GMR sensor are identified as the *auxiliary pole/upper shield*. The read signal versus density (roll-off curve) for the 52.5 Gbit/in^2 case is shown in Fig. 49.38 and results from the increasing interference of positive and negative step pulses in a long string of transitions. At low densities a step response is observed in perpendicular recording since the GMR sensor detects the vertical component of the stray magnetization that reverses at

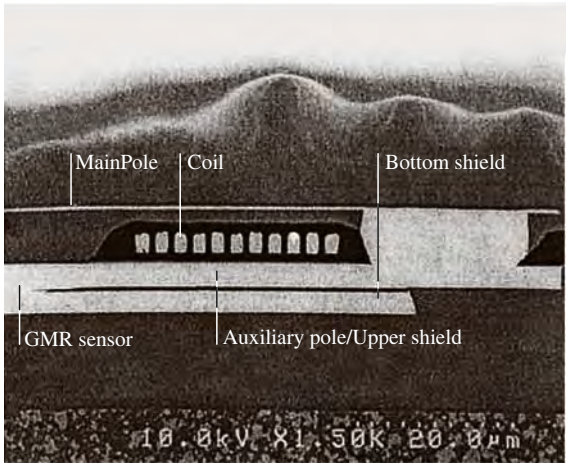


Fig. 49.37 Cross-sectional view of a perpendicular head with a GMR sensor including the bottom and upper shields. The upper shield also serves as the auxiliary pole. (After [49.80])

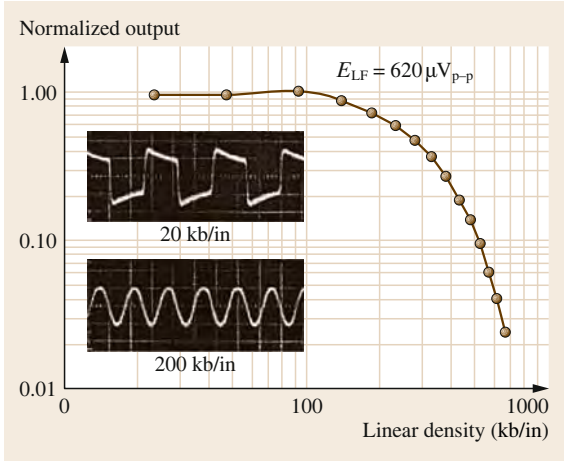


Fig. 49.38 Normalized voltage amplitude for a sequence of read signals from a perpendicular recording system versus linear density in kb/in. (After [49.80])

the transition and at higher densities the response appears as a sinusoid. The decay in the signal at the low-density response is due to the insufficient bandwidth of the preamplifier. The read signal for isolated transitions can be described by the response function

$$h(t) = A \tanh\left(\frac{\ln 3}{T_{50}}t\right), \qquad (49.30)$$

where A is one half of the signal amplitude and T_{50} is the time for the signal to rise from $-A/2$ to $A/2$. The step signals can be detected either by differentiating the signal to produce pulses, as with longitudinal recording, or with equalizer circuits that directly handle the step response. The channel performance as realized with simulator channels is listed in Table 49.12 and approximates the requirements of a practical system. Seagate technology reported an areal density of 100 Gbit/in^2 in November 2002 [49.89]. The linear density was 700 kb/in and the track density was 143 ktracks/in. In 2004 Seagate reported an areal density of 170 Gbit/in^2, crossing the 100 Gbit/in^2 barrier before *Mueller* [49.82] who reported 240 Gbit/in^2 in 2005.

Table 49.14 Progress of areal density as we change material type in PMR [49.88] through optimized selection of composite layers (thickness and material composition), magnetic cap layer (MCL) and granular recording layer (GRL). While grain size stays approximately constant, the number of grains per cluster decreases steadily in progressing from type 1 to 4, decreasing bit size and increasing areal density

Material type	Gains & Clusters	1	2	3	4
MCL + GRL	Grain size (nm)	8.8	8.6	8.8	8.1
	Distribution (%)	57.1	33.8	26.7	26.3
	Cluster size (nm)	38.8	30.0	27.1	18.9
	Distribution (%)	31.4	31.6	25.8	32.6
	Grains per cluster	19.4	12.2	9.5	5.4
GRL only	Grain size (nm)	8.7	8.5	8.8	8.1
	Distribution (%)	54.2	35.0	25.9	26.7
	Cluster size (nm)	24.1	15.6	10.1	10.4
	Distribution (%)	64.3	64.0	35.0	42.1
	Grains per cluster	7.7	3.4	1.3	1.6

Areal density can be further increased in PMR by the use of composite layers. A lower anisotropy layer called magnetic cap layer (MCL) provides continuous reversal and is deposited over a higher anisotropy granular recording layer (GRL). This results in an exchange-coupled-composite (ECC) with a continuous-granular-composite (CGC) that provides lateral exchange in the media [49.88]. In addition, CGC combines high H_c, good thermal stability and good SNR due to the granular nature, as previously discussed. The (ECC/CGC) combination paves the way for exploitation of even higher anisotropy GRL materials. Cluster size and distribution originating from lateral magnetic interaction can be optimized separately through the selection of GRL and MCL materials resulting in steady decrease of the number of grains per cluster, consequently improving areal density (Table 49.14).

Scaling Laws

Areal density controls scaling laws impacting several parameters of LMR and PMR. *Marchon* et al. [49.90] derived the following rules valid from 0.1 Gbit/in^2 to beyond 1000 Gbit/in^2:

- Bit size (in nm) $\approx 167.01\,x^{-0.3715}$, x in Gbit/in^2
- Track width (in nm) $\approx 3890.7\,x^{-0.6313}$, x in Gbit/in^2
- Bit aspect ratio (BAR) $\approx 23.296\,x^{-0.2598}$, x in Gbit/in^2
- Head media separation (HMS) (in nm) $\approx 71.5\,x^{-0.317}$, x in Gbit/in^2.

Additionally, we have the approximate relation [49.90] between HMS and bit size, expected to be valid even when the bit size is smaller (Table 49.14) than 10 nm: HMS $\approx$ (bit size)/2 as long as $10 <$ bit size (nm) < 1000.

49.2 Magnetic Random-Access Memory

The application of the giant-magnetoresistive effect to recording heads that we have discussed assumed that the sense current is in the plane of the recording film.

This is not the only possibility and devices with the current perpendicular to the plane (CPP) have been proposed both for magnetic-recording read heads but also for fast magnetic random-access memory (MRAM).

The performance of MRAM in comparison with other memory technologies is displayed in Table 49.15. SRAM, DRAM and FLASH are electronic memories, and FeRAM is ferroelectric memory similar to DRAM called 1C-1T technology (since a single bit requires a capacitor (C) and a transistor (T)) with the capacitor using a ferroelectric instead of an ordinary dielectric, also called 2C-2T technology because a single bit requires twice as much capacitors and transistors with respect to CMOS-DRAM technology.

The magnetic tunnel junction (MTJ) belongs to CPP technology and is important for both magnetic recording and MRAM developments. It involves a pinned and a free ferromagnetic film spaced by a thin insulating film, usually Al_2O_3 or MgO.

Figure 49.39 shows the schematic of an MTJ used for the MRAM application, including orthogonal bit and word or digit lines and the stack of films comprising a basic MTJ [49.93, 94].

Figure 49.40 shows a more detailed stack of films for an MTJ, including synthetic antiferromag-

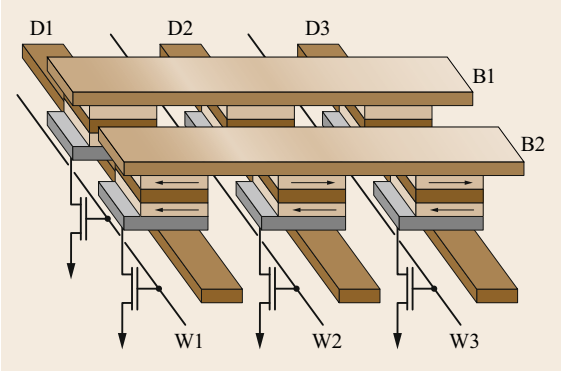


Fig. 49.39 Schematic of MRAM cross-point architecture with bits between orthogonal conductors and each cell defined by one MTJ and one transistor. The top lines in contact with the top electrode of the bits provides hard-axis fields, while the bottom lines are isolated and provide easy-axis fields. Turning on the transistor provides a current path so that the corresponding bit can be sensed. (After [49.93])

Table 49.15 Read and write times are given for the year 2007 and forecast for the year 2022. Cell surface is given in terms of parameter F the minimal feature length for the corresponding technology (current value $F = 14$ nm progressing toward 10 nm in 2017). FLASH memory has a poor endurance (program/erase cycle) of about only 10^5 because erasing uses Fowler–Nordheim tunneling that might leave electrons behind in the dielectric. FeRAM is ferroelectric RAM currently used in smartphones and electronic tablets because of its very low energy consumption, whereas STT-RAM is based on spin-transfer torque [49.91] as discussed in the text. Multilevel cell means that 1 to 3 bit might be stored in a single memory element. (After [49.92])

Memory type	SRAM	DRAM	NAND-FLASH	FeRAM	STT-MRAM
Read (2007)	0.3 ns	1–10 ns	50 ns	45 ns	< 20 ns
Read (2022)	70 ps	0.2–10 ns	10 ns	< 20 ns	< 0.5 ns
Write (2007)	0.3 ns	0.7ns	1 ns	10 ns	< 20 ns
Write (2022)	70 ps	0.2ns	1 ns	1 ns	< 0.5 ns
Retention time	Requires power	64 ms	> 10 years	> 10 years	> 10 years
Cell size	$140F^2$	$6–12F^2$	$5F^2$	$22F^2$	$6–8F^2$
Read voltage (V)	1.1	2	15	0.9–3.3	3
Write voltage (V)	1.1	2.5	2	0.9–3.3	3
Comments			Multilevel cell	Destructive readout	Multilevel cell possible

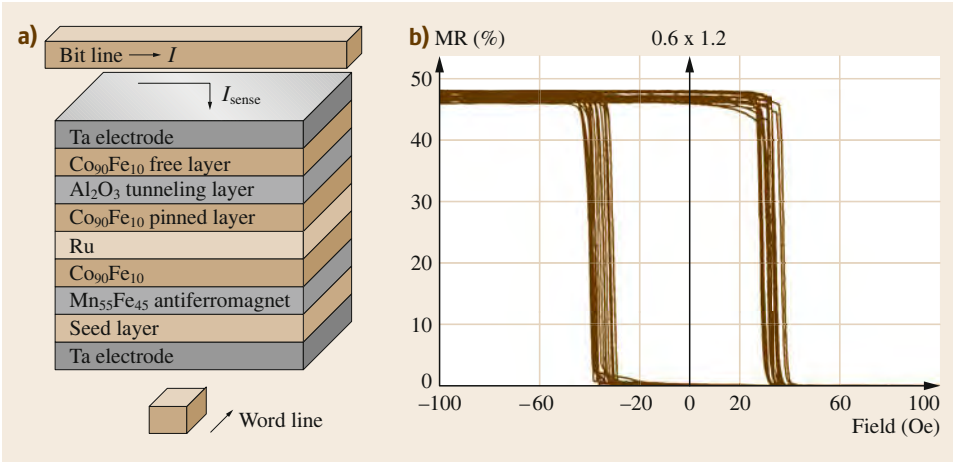


Fig. 49.40 **(a)** Stack of films comprising a typical MTJ. The *arrows* show the direction of the magnetizations in the two coupled films. **(b)** Hysteresis curves of $\Delta R/R$ versus magnetic field for $0.6 \times 1.2\,\mu m$ bits on a 6 in wafer. (After [49.94])

netic films, antiferromagnetic pinning films and the free film.

A typical MTJ has the following thickness values for the key films: antiferromagnet: 300 Å, $Co_{90}Fe_{10}$ ferromagnet: 20 Å, ruthenium: 8 Å, Al_2O_3: 10–15 Å, $Co_{90}Fe_{10}$ ferromagnet: 20 Å. As for the spin valve, the use of $Co_{90}Fe_{10}$ films for the free layer enhances $\Delta R/R$ [49.21]. Figure 49.41 also shows $\Delta R/R$ for a typical MTJ (for a bit area of $0.6 \times 1.2\,\mu m$) versus magnetic field with the multiple curves being taken over a 6 in wafer: the maximum change is greater than 48% [49.94].

The change in resistance (TMR) with orientation of the free layer relative to the pinned layer is significantly larger than for GMR technology (Table 49.15), but a thin insulating layer between the ferromagnetic films is required. The memory architecture that has evolved has a single MOS transistor coupled to the MTJ and the word line. The current in the word line supplies the magnetic field along the easy axis of the free film and is insulated from the MTJ (Fig. 49.39).

Switching of a soft-magnetic film with a magnetic field applied along the direction of an induced uniaxial anisotropy (the easy axis) and simultaneously along a direction orthogonal to the easy axis (the hard axis) has been discussed [49.5].

The switching is described by an astroid given by the following equation and shown in Fig. 49.41

$$(H_{easy})^{\frac{2}{3}} + (H_{hard})^{\frac{2}{3}} = (H_k)^{\frac{2}{3}}, \quad H_k = \frac{2K_u}{M_s}\,.$$

Note that the easy and hard magnetic fields are normalized with respect to the anisotropy field H_k in Fig. 49.41.

Switching of the state of the magnetization will occur whenever the combination of easy- and hard-axis fields is on the outside of the astroid.

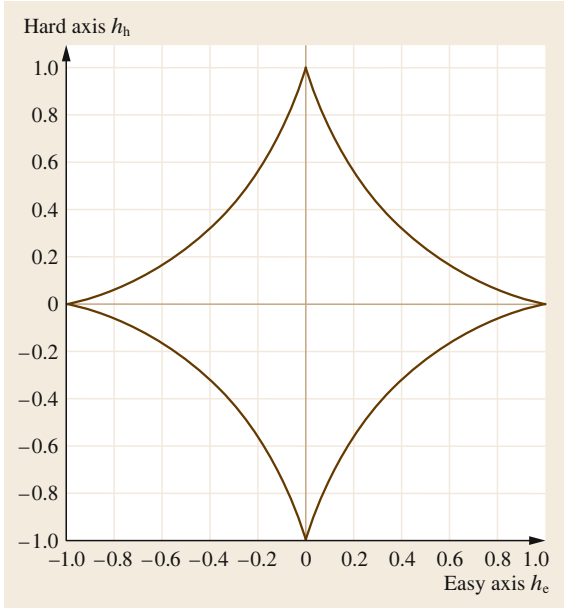


Fig. 49.41 Switching astroid plotted with magnetic fields normalized to the induced anisotropy field ($2K_u/M$). The ferromagnetic film will switch for any combination of easy- and hard-axis fields on the outside of the astroid

The switching field can be significantly reduced if magnetic fields are applied along both the easy- and hard-axis directions and an optimum is to have the fields equal. The bit line is used to supply the hard-axis magnetic field and the word line the easy-axis field. For writing a bit the sense transistor is turned off and current pulses are applied along both orthogonal conductors, generating a magnetic field at the free film sufficient in amplitude to switch the magnetization of the free film.

For reading a bit, the sense transistor is turned on by the word line and pulses of current are driven along the

bit line (with the amplitude of the magnetic field insufficient to switch the free film) and the current is either high when the ΔR is low, corresponding to say a 1 or the current amplitude is low when ΔR is high, corresponding to a 0. Without the sense transistor there would be many sneak paths for the current through other MTJs, reducing the available current for reading the selected bit.

MTJ memory cell technology shows considerable promise for replacing semiconductor SRAM and DRAM memories, primarily because of the nonvolatility of storage in MRAM technology.

The speed for reading data is comparable to that of SRAM and the density is comparable to that for DRAM. The physics of the MTJ is different from that for GMR [49.95]. Ferromagnetic films have two types of electrons: those polarized parallel to the magnetization (↑) and those polarized antiparallel to the magnetization (↓). A polarization ratio (P) for a ferromagnetic material can be defined as

$$P = \frac{D_\uparrow(E_\mathrm{F}) - D_\downarrow(E_\mathrm{F})}{D_\uparrow(E_\mathrm{F}) + D_\downarrow(E_\mathrm{F})} ,$$

where $D_\uparrow(E_\mathrm{F})$ and $D_\downarrow(E_\mathrm{F})$ are the densities of states for electrons at the Fermi energy for the spin-up and spin-down electrons respectively.

The magnetoresistance ratio of two films coupled by an insulator is [49.96]

$$\frac{\Delta R}{R} = \frac{R_\mathrm{a} - R_\mathrm{p}}{R_\mathrm{p}} = \frac{2P_1P_2}{1 - P_1P_2} , \qquad (49.31)$$

where R_a and R_p are the resistances of the coupled films in the antiparallel and parallel configurations respectively, and the spin polarizations are P_1 and P_2 for the two ferromagnetic films. Values of the spin polarization ratio are: $P_\mathrm{Co} \approx 0.35$, $P_\mathrm{Fe} \approx 0.44$ and $P_\mathrm{Permalloy} \approx 0.3$ [49.95]. With cobalt used as one electrode and Permalloy for the other electrode, the predicted value for $\Delta R/R$ is 24%. Conduction through the insulating film is by quantum mechanical tunneling, described by a transmission coefficient (T)

$$T = \frac{|\psi_\mathrm{transmitted}(x)|^2}{|\psi_\mathrm{incident}(x)|^2} ,$$

where $\psi(x)$ is the wavefunction for the electron plane wave and $|\psi(x)|^2$ is the probability of finding an electron at the position x.

The transmission coefficient is a measure of the barrier resistance. The transmission coefficient for electrons with energy E tunneling through an insulating film of thickness a with barrier height V_0, is given by

$$T = T_0 \mathrm{e}^{-2\alpha a} ,$$

where

$$T_0 = \frac{16E(V_0 - E)}{V_0^2} \text{ and } \alpha^2 = \frac{2m(V_0 - E)}{\hbar^2} .$$

This result applies when $\alpha a > 1$. A barrier height $V_0 = 10\,\mathrm{eV}$ applies for CoO.

The exponential dependence of the transmission coefficient on the thickness of the barrier shows the requirement for a thin barrier. A typical value of the thickness of an Al_2O_3 layer is 10 Å and the resistance changes by a factor of 100 million in changing the thickness from 0−10Å.

The reason that the MTJ is more attractive than a spin valve for the memory application is that the MTJ can be made with a resistance of many ohms, while the spin valve is nearly a short circuit. The use of the synthetic antiferromagnet for the MTJ also helps to avoid the superparamagnetic limit for this device, since the films comprising the synthetic antiferromagnet are thicker than the single antiferromagnetic film that was used to pin one of the ferromagnetic electrodes.

A significant problem with the MTJ is the different magnetic states that can be found depending on the size and shape of the bit. A review of an analysis of the states for a rectangular-shaped bit has been published [49.97] and we will quote some of their results to illustrate the problem. The analysis uses finite element analysis to solve the equation of motion for the magnetization (M) with Gilbert damping in the presence of an effective magnetic field ($\mathbf{H}_\mathrm{eff}$)

$$\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t} = \gamma(\mathbf{M} \times \mathbf{H}_\mathrm{eff}) - \frac{\alpha}{M}(\mathbf{M} \times \dot{\mathbf{M}}) . \qquad (49.32)$$

The effective magnetic field $\mathbf{H}_\mathrm{eff} = -\delta G/\delta\mathbf{M}$ can be derived from the magnetic free-enthalpy density (or Gibbs potential) G per unit volume given by the sum of terms pertaining to exchange, anisotropy, demagnetization and external field

$$G = G_\mathrm{exchange} + G_\mathrm{anisotropy} + G_\mathrm{demag} + G_\mathrm{external} .$$

An example of this analysis is a rectangular film of thickness 20 nm, length 1 μm, width 1 μm with magnetic properties shown in Table 49.13. The exchange stiffness constant is $A = S^2a^2J_\mathrm{ex}N_\mathrm{v}/2$, where S is the spin of the atom, a the lattice constant and N_v is the number of nearest-neighbor atoms per unit volume.

The critical length determines the maximum size of the mesh used in the finite element analysis. Figure 49.42 shows three of the six states of the magnetization in the sample.

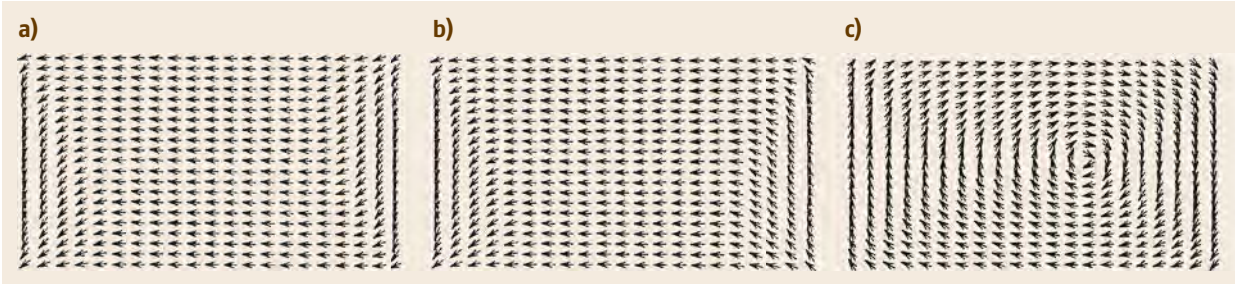


Fig. 49.42a–c Configurations for spins in a Permalloy sample described in Tables 49.16 and 49.17. Figure (a) is in S-state, (b) is in C-state and (c) is in shifted vortex state. (After [49.97])

The first two are high-remanence states referred to as S-state and C-state and can be formed if the starting conditions are a single-domain state. The last state shown is a low-remanence state with an shifted vortex. The last state is formed when the initial conditions are two domains along the long axis of the film. The total energies for the three states are shown in Table 49.17 using parameters of Table 49.16 and it is seen that the lowest energy is the low remanence state.

An issue with the MRAM is the possibility of exciting the different states in the storage cell that can then result in a different switching astroid. Thin films with low BAR bit shapes have been predicted to result in a reduced propensity to excite the vortex structures [49.98]. Changing the shape of the bit may also assist in minimizing the vortex structures. Software for analyzing these problems is the LLG Micromagnetics Simulator developed by M.R. Scheinfein.

One development in MRAM architecture is to pass the write current through the MTJ in order to heat the element and thus lower the writing threshold [49.99]. This is accomplished by turning on the selection transistor (Fig. 49.39) during the write operation. The advantages of this modification to the MRAM architecture are:

1. Since the selection at write is controlled by the temperature, selection errors are reduced.
2. The write threshold is reduced.
3. Thermal instability limits are increased by using materials with larger switching fields at the operating temperature [49.99].

Table 49.16 Parameters for Permalloy ($Ni_{80}Fe_{20}$)

M_s (T)	K_1 (J/m^3)	A (exchange stiffness) (10^{-12} J/m)	Critical length l_s (nm)
1.0	500	13	5.7

Table 49.17 Total energies for different spin configurations in the Permalloy film (described in the text and Table 49.16). (After [49.97])

Structure	G (10^{-18} J)
S-state	169.897
C-state	169.890
Shifted vortex state	88.733

Another important development in MRAM is the use of a spin current called spin-transfer torque (STT) [49.100]. Generally, spin currents [49.101] are either spin-polarized current (SPC) made of free (s-type) carriers or SWC (spin-wave current) carried by localized spins (d-type) undergoing oscillatory motion or STT originating from hybrid (s-d) exchange (called also double exchange) between free (s-type) and localized carriers (d-type) [49.91]. STT-MRAM are extremely fast (2 ns reading and writing speeds), completely electrically controlled with no moving parts and with unlimited endurance (number of program/erase cycles). In contrast, electronic FLASH memory has typically 10^5 endurance (Table 49.15).

49.2.1 Tunneling Magnetoresistive Heads

Tunneling magnetoresistance (TMR) can be used in a magnetic recording read-head application. These TMR heads use the CPP orientation of the sense current and, as with the MRAM, the tunneling is done through a thin insulating barrier. A schematic at the air-bearing surface in the simplest form of such a TMR head is shown in Fig. 49.43 [49.102]. The insulating barrier (again usually fabricated from Al_2O_3 films) is shown on edge and the current flow is in the plane of the paper.

Longitudinal bias to put the ferromagnetic TMR electrode films in a single-domain state and the top and bottom shields are shown. The GMR ratio with the TMR effect is large but the resistance × area (RA) product, which is the resistivity times the film thickness, may also be large, resulting in excessive shot noise. Shot noise is intrinsic to tunneling with the direct-current (DC) bias current required for GMR head operation and is the dominant noise source in the TMR head.

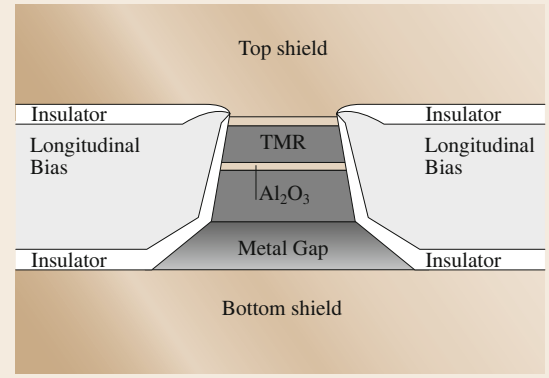


Fig. 49.43 Schematic drawing of the air-bearing surface view of a tunneling magnetoresistive (TMR) head. (After [49.102])

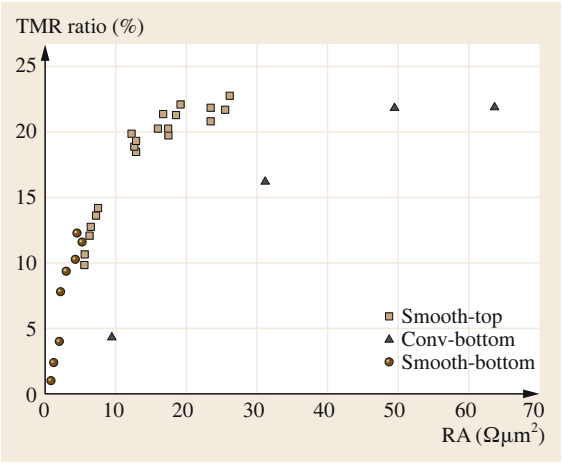


Fig. 49.44 Correlation between TMR ratio and resistance area product (RA) in three structures: *top*- and *bottom*-type MTJ with smoother interfaces and bottom-type MTJ with conventional stacking that has relatively rough interfaces. (After [49.102])

Figure 49.44 is a plot of the TMR ratio in % versus the RA product for three different TMR structures all made from the basic structure: buffer/PtMn/CoFe/Ru/CoFe/AlO_x/CoFe/NiFe/cap [49.102]. PtMn is an antiferromagnet, CoFe/Ru/CoFe is a synthetic antiferromagnet, AlO_x is the insulating barrier (prepared by in situ oxidation) and CoFe/NiFe is the second electrode. The data points for TMR versus RA for this structure are labeled *conv-bottom* (conventional bottom) and the data points for the *smooth-top* structure is for the PtMn layer on top instead of the bottom, while the *smooth-bottom* data points are for a head made with smoother interfaces. The TMR % remains constant at $\approx 20\%$ until a knee, below which the TMR % drops due to pinholes and incomplete insulation coverage.

Typically, room temperature TMR is about 1–70% when the insulating barrier is AlO_x and rises to almost 400% when the insulator is MgO (Table 49.18). At low temperature ($T = 4.2$ K) it can reach a value of 1800% in half-metallic compounds such as $La_{2/3}Sr_{1/3}MnO_3$ (LSMO)/$SrTiO_3$/LSMO, and 600% in Heusler alloys such as $Co_2MnSi/AlO_x/Co_2MnSi$ [49.103]. Perhaps the main issue is thermal stability, and that is why a system such as a perpendicular MTJ consisting of Ta/CoFeB/MgO/CoFeB/Ta possessing high thermal stability, a size as small as 40 nm diameter and a low switching current of 49 μA [49.104] is quite interesting in spite of a TMR ratio of only 120%. The thermal stability factor E_B/k_BT of the free layer should be larger than 40 to retain information (as previously seen), where E_B is the energy barrier. Because the volume V of the free layer decreases with junction size, the anisotropy energy density $M_sH_k/2$ needs to be large enough to ensure high thermal stability.

Table 49.18 TMR ratio increases strongly when AlO_x is replaced by MgO. Materials such as CoFeB/MgO/CoFeB and Co/MgO/Co reach about 400% at room temperature

Insulator	Junction	TMR (%)	Reference
AlO_x		1–70	*Zhu* et al. [49.105]
	CoFeB/AlO_x/CoFeB	70	*Wang* et al. [49.106]
MgO		5–400	*Ikeda* et al. [49.104]
	Fe/MgO (100)/Fe	220	*Parkin* et al. [49.107]

The stack of films is ion-milled and the stripe height is defined by lapping. Magnetic recording results showed the expected large signal (peak-to-peak $V_{pp} = 42$ mV/μm) with a biasing current of 1 mA with only 23.4 dB SNR deemed insufficient to demonstrate the required low error rate [49.102].

A significant problem with the TMR head is that there is a large resistance in parallel with the large capacitance resulting from the tunneling barrier. This combination results in a low-pass filter and care is required to limit the capacitance. In the example discussed here the cutoff linear density is near 580 kfc/in.

Because of the significantly larger value of ΔR for CPP as opposed to the CIP technology, CPP has become the dominant technology for read heads when areal density became much larger than 100 Gbit/in^2, whereas TMR was used when areal density was just around 100 Gbit/in^2 (see *Mueller* [49.82]).

49.3 Extraordinary Magnetoresistance (EMR)

An interesting development in read-head technology is the discovery that narrow band gap and hence high mobility semiconductors can be made to function as detectors of magnetic field [49.108]. The concept is based on the Lorentz force exerted on charge carriers by a magnetic field

$$F = q\mathbf{v} \times \mathbf{B} ,$$

where q is the charge, $\mathbf{v}$ is the velocity of the charge carrier and $\mathbf{B}$ is the magnetic flux density acting on the charge.

Electrons have the highest mobility as compared to holes and are the carriers used.

The electrons are driven by an external source similar to the sense current in GMR sensors in a high-mobility semiconductor, are deflected by the external magnetic field into a trajectory, and the voltage drop in the semiconductor at low locations is used to sense the magnetic field. The III–IV semiconductor InSb was chosen for the initial work on EMS because of the high mobility of electrons of $60\,000\,\mathrm{cm}^2/\mathrm{Vs}$.

However, a major limitation was the ability to maintain the high mobility in thin films ($< 100\,\mathrm{nm}$) required for read-head applications and this was addressed by forming a thin-film quantum well heterostructure [49.108]. An advantage of such sensors is the lack of any magnetic noise since there are no magnetic materials involved in the head structure. A review of EMR as magnetic field nanosensors has been reported by *Solin* [49.109].

49.4 Summary

It was the purpose of this chapter to review the current status of the technology of magnetic recording as used in disk drives. The emphasis was on the magnetic materials used in the application and on some of the technical problems that may limit areal density steady increase. MRAM technology, which has evolved from magnetic recording applications, was also reviewed. It was found that a wide range of magnetic materials is essential for the advance of magnetic recording and the MRAM technology, including high-magnetization soft-magnetic materials for write heads, antiferromagnetic alloys with high blocking temperatures and low propensity to corrosion for pinning films in giant-magnetoresistive sensors, and ferromagnetic alloys with large values of giant magnetoresistance. A significant limitation to magnetic recording was found to be the superparamagnetic effect and advances in multilayer ferromagnetic films to reduce the impact of the effect, but also to allow high-density recording, were discussed.

Perpendicular recording as an alternative to longitudinal recording was reviewed and it is projected by the ASTC (Advanced Storage Technology Consortium) that it will be replaced by HAMR [49.110] (heat assisted magnetic recording) and later by HDMR (heated dot magnetic recording) a combination of bit-patterned recording and HAMR. In the latter technology, large magnetic anisotropy materials are used, such as FePt (Tables 49.7 and 49.8) with $L1_0$ structure (FCT phase), which can provide grains as small as 2–3 nm in diameter yielding areal densities about as large as $100\,\mathrm{Tbit/in}^2$. The coercivity of these materials, being much larger than any magnetic field that can be generated by typical recording heads, can be reduced to zero by heating them above Curie temperature [49.110].

Magnetic quantum dot arrays that will be used in HDMR are individually addressable and adapted to scale integration since submicron ordered dot arrays possess properties such as moment, susceptibility, coercivity and remanence that are significantly enhanced with respect to their bulk counterparts. This originates from the fact they are closer to quantum atomic structures with augmented magnetic properties when fabricated on the nanometer scale. When the $100\,\mathrm{Tbit/in}^2$ level is reached, information storage by hard disks will be competing with human brain capacity, which is estimated to be on the order of several hundred terabytes.

References

49.1 E. Grochowski, R.D. Halem: IBM Systems J. **42**, 338 (2003)

49.2 R.L. Comstock: J. Mater. Sci. Electron. **13**, 509 (2002)

49.3 C.D. Mee, E.D. Daniel, M.H. Clark: *Magnetic Recording: The First 100 Years* (Wiley, New York 1998)

49.4 R.L. Comstock, M.L. Workman: Data storage in rigid disks. In: *Magnetic Storage Handbook*, 2nd edn., ed. by C.D. Mee, E.D. Daniel (McGraw-Hill, New York 1996) p. 655

49.5 R.L. Comstock: *Introduction to Magnetism and Magnetic Recording* (Wiley, New York 1999)

49.6 R. Bozorth: *Ferromagnetism* (Van Nostrand, New York 1951)

49.7 G. Bertero, S. Malhotra, B. Bian, J. Tsoi, M. Avenell, D. Wachenschwanz, T. Yamashita: IEEE Trans. Magn. **39**, 651 (2003)

49.8 X. Liu, G. Zangart, M. Shamsuzzoha: J. Electrochem. Soc. **150**(3), C159 (2003)

49.9 J.M. Harker, D.W. Brede, R.E. Pattison, G.R. Santana, L.G. Taft: IBM J. Res. Dev. **25**(5), 677 (1981)

49.10 P. Andricacos, L. Romankiw: Magnetically soft materials in data storage, their properties and electrochemistry. In: *Adv. Electrochem. Sci. Eng*, Vol. 3, ed. by D. Gerischer, C. Tobias (Wiley, Weinheim 1990)

49.11 R.L. Comstock, M. Williams: AIP Conf. Proc. Magn. Mater. **5**, 738 (1971)

49.12 H.N. Bertram: *Theory of Magnetic Recording* (Cambridge Univ. Press, New York 1994)

49.13 N. Robertson, B. Hu, C. Tsang: IEEE Trans. Magn. **33**, 2818 (1997)

49.14 X. Liu, G. Zangari: IEEE Trans. Magn. **37**, 1764 (2001)

49.15 T. Osaka, T. Yokoshima, T. Nakanishi: IEEE Trans. Magn. **37**, 1761 (2001)

49.16 Y. Liu, V. Harris, M. Kryder: IEEE Trans. Magn. **37**, 1779 (2001)

49.17 Y. Ding, S. Byeon, C. Alexander Jr.: IEEE Trans. Magn. **37**, 1776 (2001)

49.18 S. Byeon, F. Liu, G. Mankey: IEEE Trans. Magn. **37**, 1770 (2001)

49.19 R. Wood, M. Williams, J. Hong: IEEE Trans. Magn. **26**, 2954 (1990)

49.20 M. Baibich, J. Broto, F. Nguyen Van Dan, F. Petroff, P. Etienne, G. Creuzet, A. Friedrich, J. Chazelas: Phys. Rev. Lett. **61**, 2472 (1988)

49.21 H. Kanai, K. Yamada, K. Aoshima, Y. Ohtsuka, J. Kane, M. Kanamine, J. Toda, Y. Mizoshita: IEEE Trans. Magn. **32**, 3368 (1996)

49.22 B. Dieny, V.S. Speriosu, S. Metin, S. Parkin, B.A. Gurney, P. Baumgart, D.R. Wilhoit: J. Appl. Phys. **69**, 4774 (1991)

49.23 S. Parkin: Appl. Phys. Lett. **61**, 1358 (1992)

49.24 H. Fukuzawa, K. Koi, H. Tomita, H.N. Fuke, Y. Kamiguchi, H. Iwasaki, M. Sahashi: J. Magn. Magn. Mat. **235**, 208 (2001)

49.25 J.C. Mallinson: *Magneto-Resistive and Spin-Valve Heads: Fundamentals and Applications*, 2nd edn. (Academic, New York 2002)

49.26 T. Lin, C. Tsang, R. Fontana, J. Howard: IEEE Trans. Magn. **32**, 2585 (1995)

49.27 A. Devasahayam, M. Kryder: IEEE Trans. Magn. **35**, 649 (1999)

49.28 Y. Hamakawa, M. Komuro, K. Watanabi, H. Hoshiya, T. Okada, K. Nakamoto, Y. Suzuki, M. Fuyama, H. Fukui: IEEE Trans. Magn. **35**, 677 (1999)

49.29 H. Kishi, Y. Kitade, Y. Miyaki, A. Tanaka, K. Kobayashi: IEEE Trans. Magn. **32**, 3380 (1996)

49.30 S. Parkin, D. Mauri: Phys. Rev. **B 44**, 7131 (1991)

49.31 J. Mathon: Contemp. Phys. **32**, 143 (1999)

49.32 S.O. Kasap: *Principles of Electronic Materials and Devices*, 2nd edn. (McGraw-Hill, New York 2002)

49.33 N.F. Mott: The electrical conductivity of transition metals, R. Soc. Proc. A **153**, 699 (1936)

49.34 I.A. Campbell, A. Fert: Transport properties of ferromagnets. In: *Ferromagnetic Materials*, Vol. 3, ed. by E.P. Wohlfarth (North-Holland, Amsterdam 1984)

49.35 R.C. O'Handley: *Modern Magnetic Materials Principles and Applications* (Wiley, New York 2000)

49.36 R.L. White: IEEE Trans. Magn. **28**, 2482 (1992)

49.37 M. Doerner, X. Bian, M. Madison, K. Tang, Q. Peng, A. Polcyn, T. Arnoldussen, M. Toney, M. Mirzamaani, K. Takano, E. Fullerton, D. Margulies, M. Schabes, K. Rubin, M. Pinarbasi, S. Yuan, M. Parker, D. Weller: IEEE Trans. Magn. **37**, 1052 (2001)

49.38 T. Arnoldussen: IEEE Trans. Magn. **34**, 1851 (1998)

49.39 G. Choe, J. Zhou, B. Demczyk, M. Yu, M. Zheng, R. Weng, A. Chekanov, K. Johnson, F. Liu, K. Stoev: IEEE Trans. Magn. **39**, 633 (2003)

49.40 H. Bertram, H. Zhou, R. Gustafson: IEEE Trans. Magn. **34**, 1846 (1998)

49.41 M. Madison, T. Arnoldussen, M. Pinarbasi, T. Chang, M. Parker, J. Li, S. Duan, X. Bian, M. Mirzamaani, R. Payne, C. Fox, R. Wang: IEEE Trans. Magn. **35**, 695 (1999)

49.42 D. Weller, A. Moser: IEEE Trans. Magn. **35**, 4423 (1999)

49.43 E. Abarra, B. Acharya, A. Inomata, I. Okamoto: IEEE Trans. Magn. **37**, 1426 (2001)

49.44 E. Abarra, A. Inomata, H. Sato, I. Okamoto, Y. Mizoshita: Appl. Phys. Lett. **77**, 2581 (2000)

49.45 E.E. Fullerton, D.T. Margulies, M.E. Schabes, M. Carey, B. Gurney, A. Moser, M. Best, G. Zeltzer, K. Rubin, H. Rosen: Appl. Phys. Lett. **77**, 3806 (2000)

49.46 M. Schabes, E. Fullerton, D. Margulies: IEEE Trans. Magn. **37**, 1432 (2001)

49.47 A. Inomata, B. Acharya, E. Abarra, A. Ajan, D. Hasegawa, I. Okamoto: J. Appl. Phys. **91**, 7671 (2002)

49.48 Z. Zhang, Y. Feng, T. Clinton, G. Badran, N. Yeh, G. Tarnopolsky, E. Girt, M. Munteanu, S. Harkness, H. Richter, T. Nolan, R. Ranjan, S. Hwang, G. Rauch, M. Ghaly, D. Larson, E. Singleton, V. Vasko, J. Ho, F. Stageberg, V. Kong, K. Duxstad, S. Slade: IEEE Trans. Magn. **38**, 1861 (2002)

49.49 G. Choe, J. Zhou, R. Weng, K. Johnson: J. Appl. Phys. **91**, 7665 (2002)

49.50 S. Khizroev, Y. Lui, K. Mountfield, M. Kryder, D. Litvinov: J. Magn. Magn. Mater. **246**, 335 (2002)

49.51 M. Mallory, A. Torabi, M. Benaki: IEEE Trans. Magn. **38**, 1719 (2002)

49.52 S. Khizroev, D. Litvinov: *Perpendicular Magnetic Recording* (Springer Science, Berlin 2005)

49.53 W. Xia, H. Aoi, H. Muraoka, Y. Nakamura: IEEE Trans. Magn. **40**, 2365 (2004)

49.54 T. Okada, H. Kimura, I. Nunokawa, N. Yoshida, K. Etoh, M. Fuyama: IEEE Trans. Magn. **40**, 290 (2004)
49.55 N. Honda, K. Ouchi, S. Iwasaki: IEEE Trans. Magn. **38**, 1615 (2000)
49.56 K. Ouchi, N. Honda: IEEE Trans. Magn. **36**, 16 (2002)
49.57 D. Weller: IEEE Trans. Magn. **38**, 1637 (2002)
49.58 N. Honda, S. Yanase, K. Ouchi, S. Iwasaki: J. Appl. Phys. **85**, 6130 (1999)
49.59 I. Lee, H. Ryu, H. Lee, T. Lee: J. Appl. Phys. **85**, 6133 (1999)
49.60 T. Kertoku, J. Ariake, N. Honda, K. Ouchi: J. Magn. Magn. Mater. **235**, 34 (2001)
49.61 L. Wu, S. Yanase, N. Honda, K. Ouchi: J. Magn. Soc. Jpn. **21**, 301 (1997)
49.62 X. Qi, B. Stadler, R. Victora, J. Judy, O. Hellwig, N. Supper: IEEE Trans. Magn. **40**, 2476 (2004)
49.63 T. Suzuki, N. Honda, K. Ouchi: J. Magn. Soc. Jpn. **21–S2**, 177 (1997)
49.64 G. Bertero, D. Wachenschwanz, S. Malhotia, S. Velu, B. Bian, D. Stafford, Y. Wu, T. Yamashita, S. Wang: IEEE Trans. Magn. **38**, 1627 (2002)
49.65 T. Oikawa, M. Nakamura, H. Uwazumi, T. Shimatsu, H. Muraoka, Y. Nakamura: IEEE Trans. Magn. **38**, 1976 (2002)
49.66 H. Uwazumi, K. Enomoto, Y. Sakai, S. Takenoiri, T. Oikawa, S. Watanabe: IEEE Trans. Magn. **39**, 1914 (2003)
49.67 Y. Sonobe, D. Weller, Y. Ikedu, K. Takano, M. Schabes, G. Zeltzer, H. Do, R. Yeu, M. Best: J. Magn. Magn. Mater. **235**, 424 (2001)
49.68 Y. Sonobe, K. Miura, Y. Nakamura, K. Takano, H. Do, A. Mosher, B. Yen, Y. Ikeda, N. Supper: J. Appl. Phys. **91**, 8055 (2002)
49.69 H. Muraoku, Y. Sonobe, K. Muira, A. Goodman, Y. Nakamura: IEEE Trans. Magn. **38**, 1632 (2002)
49.70 Y. Sonobe, H. Muraoka, K. Miura, Y. Nakamura, K. Takano, A. Moser, H. Do, B. Yen, Y. Ikeda, N. Supper, H. Weresin: IEEE Trans. Magn. **38**, 2006 (2002)
49.71 M. Williams, C. Rettner, K. Takano, W. Weresin: IEEE Trans. Magn. **38**, 1643 (2002)
49.72 R.M. White: *Introduction to Magnetic Recording* (IEEE, New York 1985)
49.73 K.Z. Gao, H.N. Bertram: IEEE Trans. Magn. **38**, 3675 (2002)
49.74 J. Judy: J. Magn. Magn. Mater. **235**, 235 (2001)
49.75 E. Velua, S. Malhotra, G. Bertero, D. Wachenschwanz: IEEE Trans. Magn. **39**, 668 (2003)
49.76 S. Oikawa, A. Takeo, T. Hikosaka, Y. Tanaka: IEEE Trans. Magn. **36**, 2393 (2000)
49.77 M. Zheng, G. Choe, A. Chekanov, B. Demczyk, B. Acharya, K. Johnson: IEEE Trans. Magn. **39**, 1919 (2003)
49.78 K. Miura, H. Muraoka, Y. Sonobe, Y. Nakamura: IEEE Trans. Magn. **38**, 2054 (2002)
49.79 G.E. Totten (Ed.): *Application and Maintenance*, Handbook of Lubrication and Tribology, Vol. I (Taylor Francis, London 2006)
49.80 H. Takano, Y. Nishida, A. Kuroda, H. Sawaguchi, Y. Hosoe, T. Kawabe, H. Aoi, H. Muraoka, Y. Nakamura, K. Ouchi: J. Magn. Magn. Mater. **235**, 241 (2001)
49.81 W. Eppler, A. Sunder, D. Karns, E. Kurtas, G. Ju, X. Wu, P. van der Heijden, Y. Kubota, H. Zhou, C.H. Chang: IEEE Trans. Magn. **39**, 663 (2003)
49.82 S. Mueller: *Upgrading and Repairing PC's*, 22nd edn. (Que, Indianapolis 2002)
49.83 C. Brucker, T. Nolan, B. Lu, Y. Kubota, M. Plumer, P. Lu, R. Cronch, C. Chang, D. Chen, J. Chen, R. Michel, G. Parker, N. Tabat: IEEE Trans. Magn. **39**, 673 (2003)
49.84 C. Chang, M. Plumer, C. Brucker, J. Chen, R. Ranjan, J. van Elk, J. Yu, D. Karns, Y. Kubota, G. Ju, D. Weller: IEEE Trans. Magn. **38**, 1637 (2002)
49.85 M. Zheng, G. Choe, K. Johnson, L. Gao, S. Liou: IEEE Trans. Magn. **38**, 1979 (2002)
49.86 S. Kong, T. Okamoto, K. Kim, S. Nakagawa: IEEE Trans. Magn. **38**, 1982 (2002)
49.87 S. Takenoiri, K. Enomoto, Y. Sakai, S. Watanabe: IEEE Trans. Magn. **38**, 1991 (2002)
49.88 T. Wang, V. Mehta, Y. Ikeda, H. Do, K. Takano, S. Florez, B.D. Terris, B. Wu, C. Graves, M. Shu, R. Rick, A. Scherz, J. Stohr, O. Hellwig: Appl. Phys. Lett. **103**, 112403 (2013)
49.89 D.B. Bogy, W. Fong, B.H. Thornton, H. Zhu, H.M. Gross, C.S. Bhatia: IEEE Trans. Magn. **38**, 1879 (2002)
49.90 B. Marchon, T. Olson: IEEE Trans. Magn. **45**, 3608 (2009)
49.91 J.C. Slonczewski: J. Magn. Magn. Mat. **195**, L261 (1999)
49.92 S.N. Piramanayagam, T.C. Chong: *Developments in Data Storage: Materials Perspective* (IEEE Wiley, New York 2012)
49.93 S. Parkin, K. Roche, M. Samant, P. Rice, R. Beyers, R. Scheuerlein, E. O'Sullivan, S. Brown, J. Bucchigano, D. Abraham, Y. Lu, M. Rooks, P. Trouilloud, R. Wanner, W. Gallagher: J. Appl. Phys. **85**, 5828 (1999)
49.94 S. Tehrani, B. Engel, J. Slaughter, E. Chen, M. DeHerrera, M. Durlam, P. Naji, R. Whig, J. Janesky, J. Calder: IEEE Trans. Magn. **36**, 2752 (2000)
49.95 R. Meservey, P. Tedrow: Phys. Rep. **238**, 175 (1994)
49.96 M. Jullière: Phys. Lett. **54A**, 225 (1975)
49.97 H. Kronmüller, R. Hertel: J. Magn. Magn. Mater. **215**, 11 (2000)
49.98 K. Lee, W. Park, T. Kim: IEEE Trans. Magn. **39**, 2842 (2003)
49.99 I. Prejbeanu, W. Kula, K. Ounadjela, R. Sousa, O. Redon, B. Dieny, J.-P. Nozières: IEEE Trans. Magn. **40**, 2625 (2004)
49.100 A.D. Kent, D.C. Worledge: A new spin on magnetic memories, Nat. Nanotechnol. **10**, 187 (2015)
49.101 M.Z. Wu, A. Hoffmann (Eds.): *Recent Advances in Magnetic Insulators – From Spintronics to Microwave Applications*, Solid State Physics, Vol. 64 (Academic, New York 2013)
49.102 S. Araki, K. Sato, T. Kagami, S. Saruki, T. Uesugi, N. Kasahara, T. Kuwashima, N. Ohta, J. Sun, K. Nagi, S. Li, N. Hachisuka, H. Hatata, N. Kagotani, N. Takahashi, K. Ueda, M. Matsuzaki: IEEE Trans. Magn. **38**, 72 (2002)

49.103 M. Oogane, S. Mizukami: Phil. Trans. R. Soc. A **369**, 3037 (2011)
49.104 S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H.D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, H. Ohno: Nat. Mater. **9**, 721 (2010)
49.105 J.-G. Zhu, C. Park: Mater. Today **9**, 36 (2006)
49.106 D. Wang, C. Nordman, J. Daughton, Z. Qian, J. Fink: IEEE Trans. Magn. **40**, 2269 (2004)
49.107 S. Parkin, C. Kaiser, A. Panchula, P.M. Rice, B. Hughes, M. Samant, S.-H. Yang: Nat. Mater. **3**, 862 (2004)
49.108 S. Solin, D. Hines, A. Rowe, J. Tsai, Y. Pashkin, S. Chung, N. Goel, M. Santos: Appl. Phys. Lett. **80**, 4012 (2002)
49.109 S. Solin: Sci. Am. Mag. **291**, 71 (2004)
49.110 W.A. Challener, C. Peng, A.V. Itagi, D. Karns, W. Peng, Y. Peng, X.M. Yang, X. Zhu, N.J. Gokemeijer, Y.-T. Hsia, G. Ju, R.E. Rottmayer, M.A. Seigler, E.C. Gage: Nat. Photonics **3**, 220 (2009)

50. High-Temperature Superconductors

Rainer Wesche

The discovery by J. G. Bednorz and K. A. Müller in 1986 that the superconducting state can exist in oxides at temperatures above 30 K stimulated research in the field of superconductivity and opened a new field of research. Within a few years a large number of cuprate superconductors with transition temperatures well above the boiling point of liquid nitrogen have been found. In this chapter, an overview of the major families of high-temperature superconductors and their physical properties is presented.

Starting from the well-known characteristics of conventional superconductors, described in Sect. 50.1, the new phenomena observed in high-temperature superconductors are considered. The complexity of the physical properties of the cuprate superconductors is closely related to the fact that these materials are close to a metal–insulator transition. In Sects. 50.2 and 50.3, the crystal structures, the general trends for the critical temperatures, the anisotropy of the physical properties, and the factors limiting the transport critical current density are discussed. Because of their importance in the field of electronics some features of thin films are presented in Sect. 50.4.

Strictly speaking, the binary compound MgB_2 is not a high-temperature superconductor. Nevertheless, an overview of the physical properties of this interesting metallic superconductor, characterized by a transition temperature as high as 39 K, has been included in the present chapter (Sect. 50.5).

In 2008, H. Hosono et al. discovered that the layered superconductor $La(O_{1-x}F_x)FeAs$ ($x = 0.11$) superconducts at a temperature as high as 26 K. A maximum critical temperature of 58 K has been reported for the iron-pnictide superconductor $Sm(O_{0.74}F_{0.26})FeAs$. The iron-based superconductors are therefore considered as a second class of high-temperature superconductors. Their properties are described in Sect. 50.6.

50.1 **The Superconducting State** 1228
50.1.1 Characteristic Properties of Superconductors 1228
50.1.2 Superconductor Electrodynamics 1229
50.1.3 Superconductivity: A Macroscopic Quantum Phenomenon 1230
50.1.4 Type II Superconductors 1232
50.2 **Cuprate High-T_c Superconductors: An Overview** 1234
50.2.1 Major Families of Cuprate Superconductors 1234
50.2.2 Generic Phase Diagram of Cuprate Superconductors 1235
50.2.3 Crystal Structures 1236
50.2.4 Critical Temperatures 1238
50.3 **Physical Properties of Cuprate Superconductors** 1239
50.3.1 Anisotropic Superconductors 1239
50.3.2 Irreversibility Line 1239
50.3.3 Limitations of the Transport Critical Current 1241
50.4 **Superconducting Films** 1244
50.5 **The Special Case of MgB_2** 1248
50.6 **Iron-Based Superconductors** 1250
50.7 **Summary** 1252
References 1252

After the successful development of helium liquefaction techniques in the laboratory of Heike Kammerlingh Onnes at the University of Leiden, temperatures down to about 1 K became accessible for further research [50.1]. One of the first aspects to be studied was the electrical resistivity of pure metals at very low temperatures. The nearly temperature-independent residual resistivity of platinum and gold was found to decrease with increasing purity of the metals. In 1911, Kammerlingh Onnes found that, in contrast to the behavior of platinum and gold, the electrical resistivity of mercury drops to an unmeasurably small value at

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_50

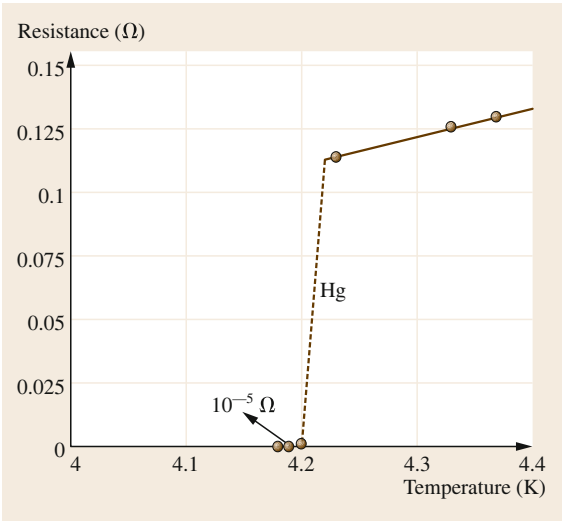


Fig. 50.1 Resistance–temperature plot for mercury obtained by Heike Kammerlingh Onnes

a temperature of 4.2 K, as shown in Fig. 50.1. The remarkable phenomenon of superconductivity had been discovered [50.1].

In addition to zero resistance, the superconducting state is characterized by perfect diamagnetism. *W. Meissner* and *R. Ochsenfeld* found in 1933 that, as soon as the superconducting state is reached, a magnetic field is expelled from the interior of a superconductor cooled in the presence of a static magnetic field [50.2]. In contrast to the shielding of a magnetic field applied to a material in the superconducting state, field expulsion cannot be explained by perfect conductivity. This Meissner effect is therefore indeed an additional characteristic property of superconductors.

The discoveries listed in Table 50.1 provide an overview of the history of superconductivity. The improvement in the highest known critical temperatures in metals, oxides, molecular (organic), and iron-based superconductors is shown in Fig. 50.2. Between 1911 and 1974, the critical temperatures T_c of metallic superconductors steadily increased from 4.2 K in mercury up to 23.2 K in sputtered Nb_3Ge films [50.17, 18]. Nb_3Ge held the record for the critical temperature in metallic superconductors until the unexpected discovery of superconductivity at 39 K in the intermetallic compound MgB_2 [50.15].

The first superconductingxoo oxide $SrTiO_3$ characterized by a transition temperature as low as 0.25 K was discovered in 1964 [50.19]. A remarkably higher critical temperature of 13 K was found for the perovskite $BaPb_{1-x}Bi_xO_3$ in 1975 [50.20]. These results stimulated J. G. Bednorz and K. A. Müller to search for higher critical temperatures in oxides. Their discovery of high-temperature superconductivity in the cuprate $(La, Ba)_2CuO_4$ ($T_c \approx 30$ K) opened up a new field of research [50.12]. Within less than a year a critical temperature well above the boiling point of liquid nitrogen could be achieved in $YBa_2Cu_3O_{7-\delta}$ [50.21]. So far, the highest transition temperature of 135 K at ambient pressure has been found in $HgBa_2Ca_2Cu_3O_{8+\delta}$ [50.13, 14].

It should be mentioned that the phenomenon of superconductivity has also been observed in organic materials [50.22–24] and fullerenes [50.25–27]. Remarkably, high transition temperatures of 19, 33 and 40 K have been found in the fullerenes K_3C_{60}, $RbCs_2C_{60}$, and Cs_3C_{60} respectively.

Since *H. Hosono* and collaborators discovered in 2008 [50.16] that the iron-pnictide $La(O_{0.89}F_{0.11})FeAs$ superconducts at a temperature of 26 K, the cuprates are no longer the only known class of complex high-temperature superconductors. The highest reported critical temperatures of iron-based superconductors are in the range of 56−58 K.

An important advantage of high-temperature superconductors is the possibility of operating devices at

Table 50.1 Milestones in the history of superconductivity

1911	*H. Kammerlingh Onnes* discovers superconductivity in Hg [50.1]
1933	*W. Meissner* and *R. Ochsenfeld* find perfect diamagnetism for the superconducting state [50.2]
1935	London theory of the superconductor electrodynamics [50.3]
1957	*J. Bardeen* et al. develop a quantum theory of superconductivity (BCS theory) [50.4, 5]
1960	*I. Giaever* measures the energy gap by electron tunneling [50.6, 7]
1961	Experimental confirmation of flux quantisation indicating the existence of Cooper pairs [50.8, 9]
1962	*B. D. Josephson* predicts Cooper pair tunneling between two superconductors separated by a thin insulating oxide layer [50.10]
1965	Observation of superconducting quantum interference between Josephson junctions in multiply connected superconducting circuits [50.11]
1986	*J. G. Bednorz* and *K. A. Müller* discover high-temperature superconductivity in the La-Ba-Cu-O system [50.12]
1993	Critical temperature of 135 K in $HgBa_2Ca_2Cu_3O_{8+\delta}$ [50.13, 14]
2001	*J. Akimitsu* et al. discover superconductivity at 39 K in MgB_2 [50.15]
2008	*H. Hosono* et al. discover superconductivity at 26 K in the layered iron pnictide $La(O_{0.89}F_{0.11})FeAs$ [50.16]

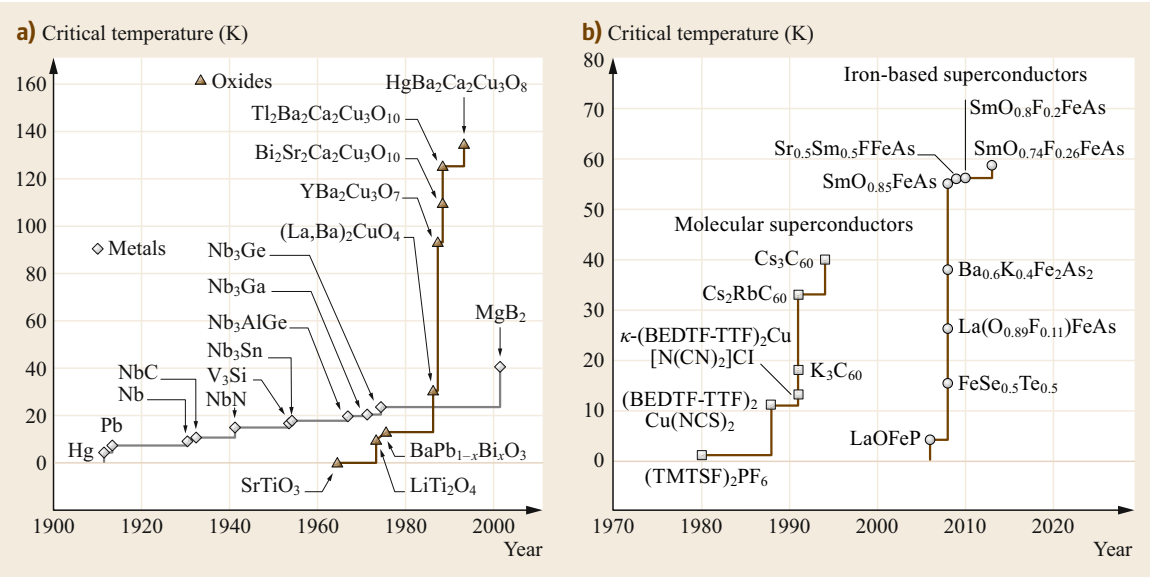


Fig. 50.2a,b Highest known critical temperatures of metallic, oxide, molecular (TMTSF and BEDTF-TTF) and iron-based superconductors. Between 1911 and 1974, the critical temperature of metallic superconductors increased steadily from 4.2 K (Hg) to 23.2 K (Nb_3Ge). In 2001, the unexpected discovery of superconductivity in MgB_2 raised the maximum known T_c of metallic superconductors to 39 K. After the discovery of high-T_c superconductivity in $(La, Ba)_2CuO_4$, in 1986, the maximum T_c of oxides rapidly exceeded 77 K the boiling point of liquid nitrogen **(a)**. In 1980, a first molecular superconductor was discovered. In the fullerene Cs_3C_{60}, a remarkably high T_c of 40 K was found. Between 2006 and 2013, the maximum known T_c of iron-based superconductors increased from around 4 K (LaOFeP) to 58 K ($SmO_{0.74}F_{0.26}FeAs$) **(b)**. (After [50.12–34])

temperatures well above 20 K, leading to a considerable reduction in the required refrigerator input power. In contrast to power applications, the use of liquid nitrogen as a coolant is of less importance for superconducting electronics, which are typically cooled by closed cycle coolers. The economics of closed cycle coolers depends strongly on the operating temperature. Considering an ideal, reversible cooling cycle, the input power necessary to remove a heat load dQ/dt is

$$P_{in} = \frac{1}{\eta_C}\frac{dQ}{dt}\,, \qquad (50.1)$$

where

$$\eta_C = \frac{T_{op}}{300\,K - T_{op}} \qquad (50.2)$$

is the Carnot efficiency. To remove a heat input of 1 W an ideal, reversible refrigerator consumes at room temperature a power of 70 W for $T_{op} = 4.2$ K, whereas this power is only 9 and 2.9 W for operation at 30 and 77 K, respectively. The efficiency η_r of real refrigerators is much smaller than the Carnot efficiency, and depends strongly on the refrigerator size, as illustrated in Fig. 50.3. For small refrigerators able to remove a power of 1 W, the refrigerator efficiency is only a few

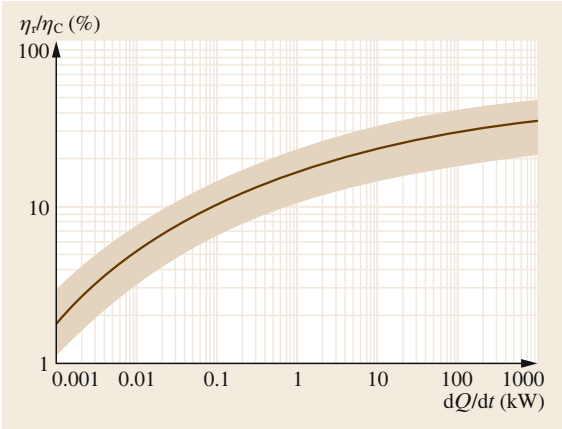


Fig. 50.3 The efficiency of real refrigerators normalized to the Carnot efficiency is typically in the *highlighted area*. The ratio of η_r/η_C increases with increasing refrigeration capacity. (After [50.35])

percent of the Carnot efficiency. On the other hand, large refrigerators with a cooling capacity well above 10 kW may reach 30% of the Carnot efficiency. The ratio η_r/η_C has been found to depend only weakly on the operating temperature.

50.1 The Superconducting State

50.1.1 Characteristic Properties of Superconductors

The discussion of the physical properties of high-temperature superconductors has to be based on the knowledge of the behavior of conventional, metallic superconductors. In this section the superconducting state is therefore described in more detail, starting from the properties of simple, metallic superconductors.

Below a certain temperature, T_c the electrical resistance of a superconducting material vanishes. The critical temperature T_c is characteristic of the superconductor in question. In pure metals, the zero resistance state can be reached within a few mK (Fig. 50.1). In the chemically complex cuprate, high-T_c superconductors the transition to the superconducting state is less sharp than in metallic low-T_c superconductors. The transition width ΔT_c for single-phase cuprate superconductors is typically 1 K. Thus the critical temperature depends slightly on the criterion used to define T_c. Several criteria are illustrated in Fig. 50.4.

At the transition temperature, a resistance drop of several orders of magnitude is typically observed. However, it is in principle not possible to prove experimentally that the resistance in the superconducting state is in fact zero. The most efficient way to determine an upper limit of the resistance is to detect the decay of the magnetic fields generated by currents induced in a superconducting loop. Upper resistivity limits between 2×10^{-18} [50.36] and $7\times10^{-23}\,\Omega$ cm [50.37] have been reported for the high-T_c superconductor $YBa_2Cu_3O_7$, while a value of $3.6\times10^{-23}\,\Omega$ cm has been found for type I low-T_c superconductors [50.38]. These upper resistivity limits are orders of magnitude smaller than the resistivity of $10^{-10}\,\Omega$ cm at 4.2 K achievable in annealed, very pure metals. It is therefore justified to assume zero resistance below T_c for all practical purposes.

The second outstanding feature of the superconducting state is perfect diamagnetism. To understand the magnetic behavior of a superconductor, the two different situations illustrated in Fig. 50.5 have to be considered. In the first case, the superconductor is cooled below T_c without an applied magnetic field (zero field-cooled). In the second case the superconductor is cooled below T_c while a magnetic field is applied (field-cooled). In both cases, the magnetic flux is excluded from the interior of the superconductor. Screening currents flowing in the surface layer of the superconductor generate a flux opposite to that of the applied magnetic field. The magnetic flux density is therefore zero everywhere inside the superconductor. On the other hand, outside the superconducting sphere the magnetic field is enhanced because of the superposition of the flux

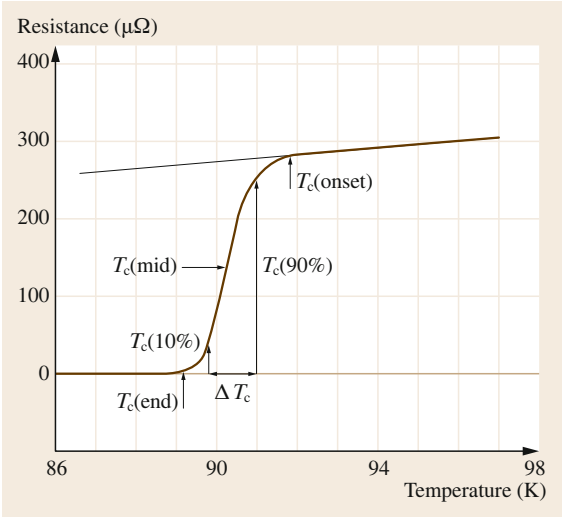


Fig. 50.4 Resistance versus temperature for a $Bi_2Sr_2CaCu_2O_8$/Ag multicore wire. The transition temperatures resulting from different definitions of T_c are indicated. The width of the transition $\Delta T_c = T_c(90\%) - T_c(10\%)$ is ≈ 1.2 K

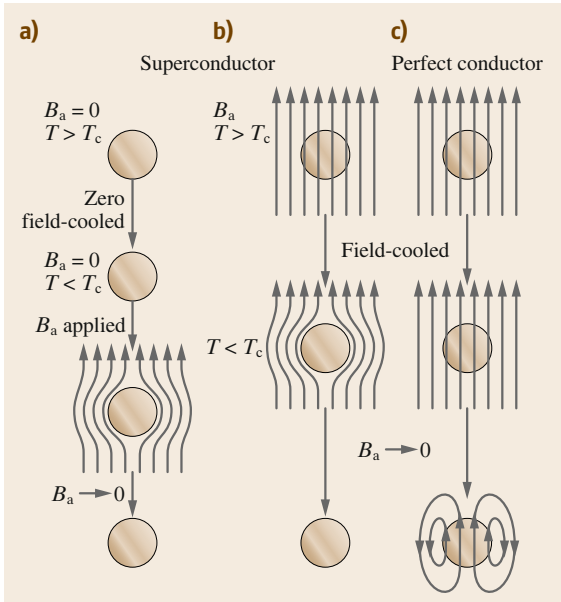


Fig. 50.5a–c The magnetic flux is excluded from the interior of a superconductor **(a)** without field-cooling as well as **(b)** with it. In contrast to this behavior, a magnetic flux would exist in the interior of a field-cooled perfect conductor **(c)**. (After [50.39, p. 20])

of the applied magnetic field and that of the screening currents. In both situations, the superconductor is unmagnetized when the magnetic field is removed. The behavior of a zero field-cooled superconductor may be considered simply as a screening effect caused by perfect conductivity. In contrast to screening, the expulsion of the magnetic flux from a field-cooled superconductor (the Meissner effect) cannot be explained by perfect conductivity. The different behavior of a field-cooled perfect conductor is also shown in Fig. 50.5. The values of the relative magnetic permeability for nonferromagnetic metals are close to unity. Thus the magnetic flux inside the metal is nearly the same as that of the applied magnetic field. Because $dB/dt = 0$, no screening currents are induced, and hence the magnetic flux is not removed from the interior of the perfect conductor at low temperatures. Finally, the removing of the magnetic field ($dB/dt \neq 0$) leads to the magnetization of the perfect conductor.

The superconducting state can be destroyed by a sufficiently large applied magnetic field. The transition from the superconducting to the normal state occurs at a certain critical magnetic field $\boldsymbol{B}_c$, which is a characteristic property of the superconductor considered. The dependence of the critical field on the temperature is typically well described by

$$\boldsymbol{B}_c(T) = \boldsymbol{B}_{c0}\left[1 - \left(\frac{T}{T_c}\right)^2\right], \tag{50.3}$$

where T_c is the critical temperature and $\boldsymbol{B}_{c0}$ is the critical field at zero temperature. The critical field as a function of temperature is shown in Fig. 50.6 for selected metallic superconductors. The $\boldsymbol{B}_c(T)$ curves

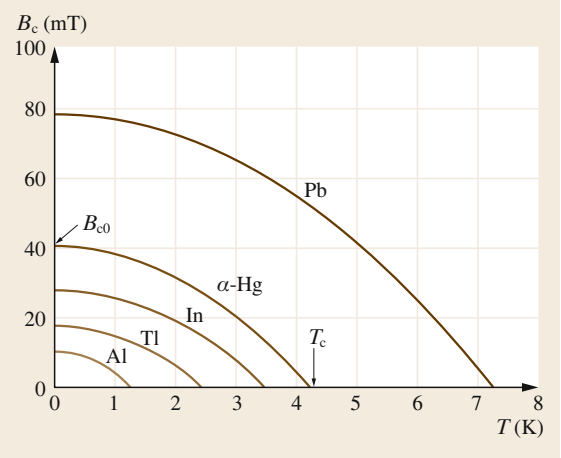


Fig. 50.6 Critical field versus temperature curves for selected metallic superconductors. The B_{c0} values vary from ≈ 10 to ≈ 80 mT. (After [50.39, p. 24])

separate the superconducting from the normal state. The Meissner effect ensures that the properties at any point in the superconducting state are independent of the order in which the final conditions of the applied magnetic field and temperature are reached.

50.1.2 Superconductor Electrodynamics

The electrodynamics of superconductors can be described by the two London equations. As a consequence of the infinite conductivity in the superconducting state, Ohm's law must be replaced by the first London equation

$$\frac{d\boldsymbol{j}}{dt} = \frac{1}{\mu_0 \lambda_L^2}\boldsymbol{E}, \tag{50.4}$$

where $\boldsymbol{j}$ is the current density, $\boldsymbol{E}$ is the electric field, μ_0 is the permeability of free space, and λ_L a constant. The electric field in the superconductor is zero for constant current density. The current density and the magnetic field are related by the second London equation

$$\mathrm{curl}\,\boldsymbol{j} = -\frac{1}{\mu_0 \lambda_L^2}\boldsymbol{B}. \tag{50.5}$$

Using the Maxwell equation $\mathrm{curl}\,\boldsymbol{B} = \mu_0 \boldsymbol{j}$ the following differential equation results for the magnetic field

$$\nabla^2 \boldsymbol{B} = \frac{1}{\lambda_L^2}\boldsymbol{B}. \tag{50.6}$$

For a semi-infinite superconductor with the magnetic field applied along the z-direction, an exponential decay of the magnetic field $\boldsymbol{B}(x) = \boldsymbol{B}_0 \exp(-x/\lambda_L)$ results, as illustrated in Fig. 50.7. The London penetration depth λ_L is therefore the characteristic length for the decay of the magnetic field in the surface layer of the superconductor. Typical values of the penetration depth are 10^{-5}–10^{-6} cm at $T = 0$. The penetration depth is closely related to the number density of the Cooper pairs (Sect. 50.1.3)

$$\lambda_L = \sqrt{\frac{m_C}{\mu_0 n_C q_C^2}}, \tag{50.7}$$

where n_C, q_C, and m_C are the number density, the charge and the mass of the Cooper pairs. Because of the decreasing Cooper pair density, the penetration depth increases rapidly for temperatures close to T_c (see the inset of Fig. 50.7). The temperature dependence is given by

$$\lambda_L(T) = \lambda_{L0}\left[1 - \left(\frac{T}{T_c}\right)^4\right]^{-1/2}, \tag{50.8}$$

where λ_{L0} is the penetration depth at $T = 0$.

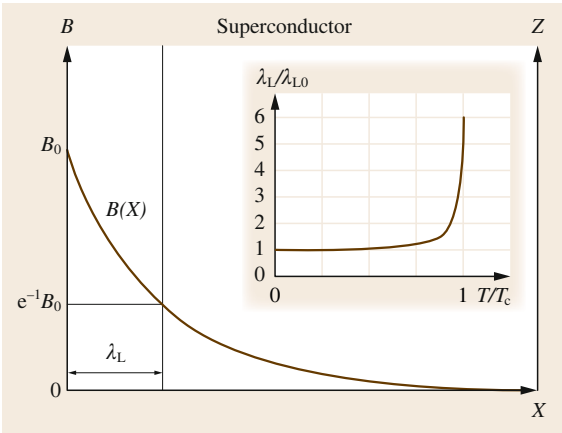


Fig. 50.7 Exponential decay of an applied magnetic field in the surface layer of a semi-infinite superconductor. The normalized penetration depth λ_L/λ_{L0} versus reduced temperature is shown in the *inset*

50.1.3 Superconductivity: A Macroscopic Quantum Phenomenon

Bardeen–Cooper–Schrieffer (BCS) Theory

Next, the microscopic causes of superconductivity in conventional superconductors are briefly described. *J. Bardeen* et al. proposed a microscopic theory of superconductivity [50.4, 5], which is now well known as the BCS theory. At the moment, it seems to be unlikely that the mechanism of superconductivity in the cuprates, which is still not yet understood, can be explained by the BCS theory. The recently discovered iron-based superconductors seem to be a second class of high-temperature superconductors, which show unconventional superconductivity [50.16]. In this chapter, only some fundamental ideas and results of the BCS theory are presented.

Independent of each other, two researchers [50.40, 41] found theoretically that lattice vibrations (phonons) can provide an attractive interaction between two electrons. This can be understood as a consequence of the polarization of the lattice by the electrons. The concentration of positive charges produced by the first electron reduces the potential energy of the second electron. Because of this attractive interaction, electrons with opposite momenta and spins form electron (Cooper) pairs. The Cooper pairs, characterized by zero spin and momentum, have to be considered as new particles. Because their spin is zero, they do not obey Pauli's exclusion principle. The probability that a state is occupied increases with the number of particles populating the same quantum state. Therefore, all Cooper pairs occupy the same quantum state with a single value of the momentum in the presence of an applied electric field. A change of momentum results only for scattering processes that break up some Cooper pairs. Resistance occurs as soon as the energy taken from the electric field exceeds the binding energy of the Cooper pairs. This behavior explains the existence of a critical current.

The electron–phonon interaction leads to an energy gap 2Δ in the density of states for the single electrons. The width of the energy gap at zero temperature is

$$2\Delta(0) = 3.5 k_B T_c \,, \tag{50.9}$$

where k_B is the Boltzmann constant and T_c is the critical temperature. For superconducting elements the measured ratios of $2\Delta(0)/k_B T_c$ are typically 3.2–4.6 [50.42], which is in reasonably good agreement with the value of 3.5 predicted by the BCS theory. The single-electron density of states close to the Fermi energy is

$$D_s(E - E_F) = D_n(E_F) \frac{|E - E_F|}{\sqrt{(E - E_F)^2 - \Delta^2}} \,, \tag{50.10}$$

where Δ is the half-width of the energy gap and $|E - E_F| > \Delta$. The variation of the normal state density of states $D_n(E)$ can be neglected because $E_F \gg 2\Delta$. The single-electron density of states shown in Fig. 50.8 reflects that the energy states present in the normal state within the energy gap are shifted to the edges. The inset of Fig. 50.8 shows the temperature dependence of the energy gap predicted by the BCS theory. The critical temperature is [50.5]

$$T_c = 1.14 \frac{\hbar \omega_D}{k_B} \exp\left(-\frac{1}{D_n(E_F) V_p}\right) , \tag{50.11}$$

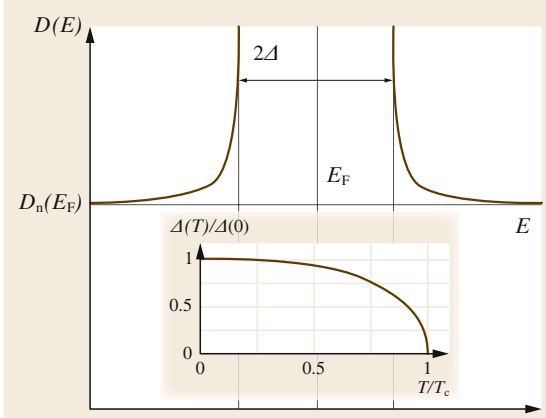


Fig. 50.8 Energy gap and density of states for unpaired electrons in the superconducting state in the vicinity of the Fermi energy E_F. The temperature dependence of the energy gap predicted by the BCS theory is shown in the *inset*

where the electron–phonon interaction is described by the constant V_p. The Debye frequency ω_D is the cut-off frequency of the phonon spectrum in a solid. Considering the atoms in a solid as harmonic oscillators of frequency $\omega = (k_e/M)^{1/2}$, where k_e is the elastic constant resulting from the binding forces, both ω_D and T_c are expected to be proportional to $M^{-1/2}$ for different isotopes of the same element. In many metallic superconductors, this isotope effect has been observed [50.43–46].

All Cooper pairs populate the same quantum state, resulting in a collective wave function

$$\psi = n_C^{1/2} \exp[i\varphi(r)] , \qquad (50.12)$$

where $n_C = \psi^*\psi$ (ψ^*: conjugate complex of ψ) is their concentration. Because the phase coherence of the Cooper pairs is effective at large distances, the phase $\varphi(r)$ is a macroscopically observable quantity. As a consequence, the magnetic flux in a superconducting hollow cylinder of the cross-sectional area F is quantized

$$\phi = \int_F B\mathrm{d}F = \frac{h}{2e} N_\phi = \phi_0 N_\phi . \qquad (50.13)$$

In the above equation, N_ϕ is an integer, e is the electron charge and $\phi_0 = 2.068 \times 10^{-15}$ V s is the flux quantum. The magnetic flux generated by the screening currents flowing in the surface layer of the superconductor is typically small, and has therefore been neglected.

Tunneling Effects

First, single-electron tunneling in a normal–insulator–superconductor (N-I-S) junction will be considered. The two conductors are separated by an insulating oxide layer typically less than 2 nm thick. Tunneling can take place only when empty target energy states are available, otherwise the process is forbidden by the Pauli principle (Fig. 50.9). For a voltage of $U = \Delta/e$ ($-e > 0$), the empty energy states of the superconductor above the energy gap reach the same level as the Fermi energy of the normal conductor, and strong tunneling sets in. On the other hand, for a voltage $U = -\Delta/e$, pair breaking becomes possible. One of the electrons tunnels into the normal conductor and gains an energy Δ available for the excitation of the second electron to an empty state above the energy gap. The resulting N-I-S current–voltage characteristic and the Bose condensation representation of the tunneling process are shown in Fig. 50.9.

In the case of two identical superconductors (S-I-S junction), strong single-electron tunneling sets in at an applied voltage $|U| \geq |2\Delta/e|$. This voltage is high

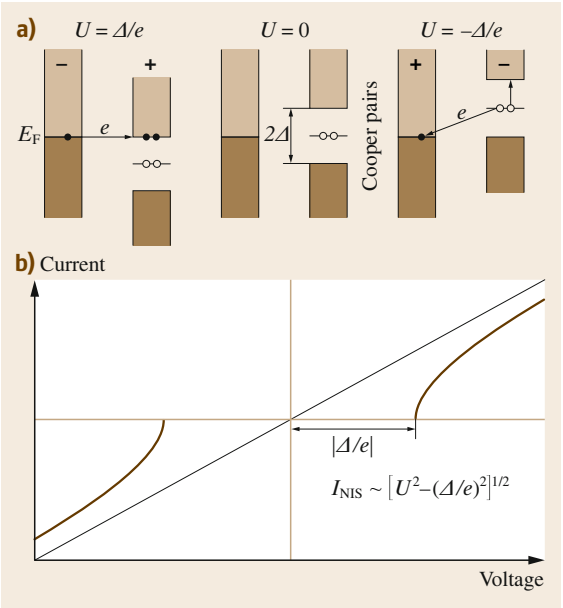


Fig. 50.9a,b Bose condensation representation of N-I-S tunneling **(a)** and the corresponding current voltage characteristic at $T = 0$ **(b)**

enough to allow the empty energy states above the gap of the first superconductor, and the filled energy states below the gap of the second superconductor, to overlap; pair breaking becomes possible. Figure 50.10 shows the semiconductor representation of the tunneling process and the resulting S-I-S current–voltage characteristic. The weak tunneling currents for $|U| < |2\Delta/e|$ are a consequence of a few quasi-particles being thermally excited above the energy gap for $T > 0$ (Fig. 50.10).

So far, only single-electron tunneling has been considered. In 1962, *B. D. Josephson* predicted that Cooper pair tunneling through a very thin insulating layer is possible [50.10]. Even for zero voltage across the insulating layer of an S-I-S junction, Cooper pairs can tunnel through the barrier. This process has to be considered as the tunneling of the wave function collectively describing the Cooper pairs. The DC current through the contact is

$$I = I_c \sin\delta , \qquad (50.14)$$

where I_c is the critical current of the Josephson junction. The Josephson DC current through the contact is between $-I_c$ and I_c, depending on the phase difference δ of the two Cooper pair wave functions. This phase difference is time independent for zero voltage across the insulating layer. Applying a DC voltage U across the insulating layer of the Josephson contact causes a time-

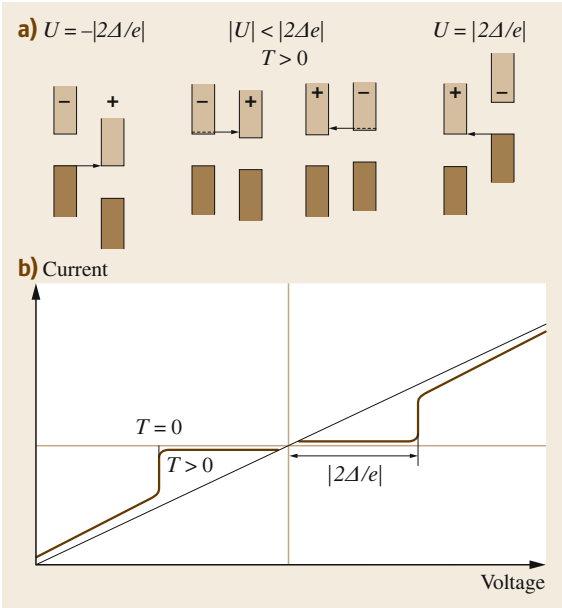


Fig. 50.10a,b S-I-S tunneling. The semiconductor representation of the process **(a)** shows that strong tunneling sets in for $U = |2\Delta/e|$. A few electrons (quasi particles) are thermally excited into states above the gap at $T > 0$. The corresponding current–voltage characteristic **(b)** shows that these electrons are responsible for the weak tunneling current at low voltages. (After [50.39, p. 42])

dependent phase difference between the two Cooper pair wave functions. As a consequence, there flows an AC current

$$I_{AC} = I_c \sin\left[2\pi\nu_J t + \delta(0)\right], \tag{50.15}$$

where

$$\nu_J = \frac{2\,|e|\,U}{h} = \frac{U}{\phi_0}\,. \tag{50.16}$$

The AC Josephson current I_{AC} oscillates with a frequency of 483.6 MHz/μV. The tunneling of the Cooper pair through the barrier is connected to the emission or absorption of a photon with energy $E = h\nu_J$. The AC Josephson current flows in addition to the current resulting from single-electron tunneling. A more detailed description of the AC Josephson effect can be found in [50.42].

Finally, macroscopic quantum interference effects in Josephson junctions will be briefly considered. These interference effects are closely connected to the phase shifts of the Cooper pair wave functions caused by a magnetic field applied transversely to the current direction. The applied magnetic field gives rise to a spatial variation of the phase difference of the two Cooper

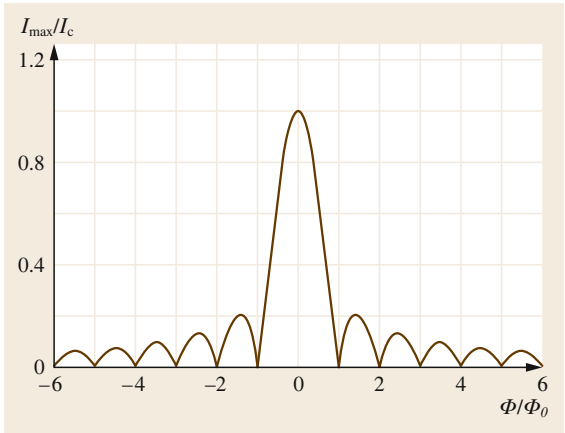


Fig. 50.11 Josephson–Fraunhofer diffraction pattern for the normalized maximum Josephson DC current I_{max}/I_c. Minima of the current occur for Φ/Φ_0 values of 1, 2, 3 etc. (After [50.39, p. 49])

wave functions within the contact area. The resulting Josephson DC current for a rectangular barrier is

$$I = I_c \sin\delta_0 \frac{\sin \pi\phi/\phi_0}{\pi\phi/\phi_0}\,, \tag{50.17}$$

where I_c is the critical current of the Josephson junction, ϕ is the total flux enclosed by the barrier, and ϕ_0 the flux quantum. The phase shift δ_0 is unknown. The maximum Josephson DC current

$$I_{max} = I_c \left| \frac{\sin \pi\phi/\phi_0}{\pi\phi/\phi_0} \right| \tag{50.18}$$

is reached for $\sin\delta_0 = 1$, corresponding to a phase difference of $\pi/2$. In Fig. 50.11, the resulting interference pattern for the Josephson DC current is shown. The Josephson–Fraunhofer interference of the two Cooper pair wave functions indicates that superconductivity is a macroscopic quantum phenomenon (for more details see [50.39]).

50.1.4 Type II Superconductors

A quantitative description of the behavior of type II superconductors is provided by the Ginzburg–Landau theory (see for example [50.42]). In the London theory, the Cooper pair density is assumed to be constant. The only characteristic length scale is the penetration depth λ_L describing the decay of a magnetic field inside a superconductor. Taking into consideration, a variation of the Cooper pair density leads to a second characteristic length scale, namely the coherence length ξ, the shortest distance within which a considerable change

Table 50.2 Coherence length ξ and Ginzburg–Landau parameter κ for selected superconductors at $T = 0$. (After [50.39, p. 70])

Material	T_c (K)	ξ (nm)	$\kappa = \lambda_L/\xi$	Reference
Al	1.18	550	0.03	[50.42]
In	3.41	360	0.11	[50.42]
Cd	0.52	760	0.14	[50.42]
Sn	3.72	180	0.23	[50.42]
Ta	4.5	93	0.38	[50.47]
Pb	7.2	82	0.48	[50.42]
Nb	9.25	39	1.28	[50.42]
NbTi	9.5	4	75	[50.48]
Nb_3Sn	18	3	21.7	[50.48]
Nb_3Ge	23.2	3	30	[50.48]
Rb_3C_{60}	29.6	2	123.5	[50.49]
$YBa_2Cu_3O_{7-\delta}$	93	2	95	[50.42]

of the Cooper pair density can occur. At the interface of a normal and a superconductor, the Cooper pair density increases continuously from zero to the value n_C deep inside the superconductor. The ratio of the coherence length and the penetration depth is the Ginzburg–Landau parameter κ

$$\kappa = \frac{\lambda_L}{\xi} . \qquad (50.19)$$

Depending on whether the parameter κ is smaller or larger than $1/\sqrt{2}$, the superconductor in question is of type I or II, respectively. Values of the coherence length and the Ginzburg–Landau parameter of selected metallic superconductors are listed in Table 50.2. The cuprate high-T_c superconductors are extreme type II superconductors with κ values of ≈ 100.

In Fig. 50.12, the phase diagrams of both types of superconductors are compared. In type I superconductors, there exist two different phases, namely the

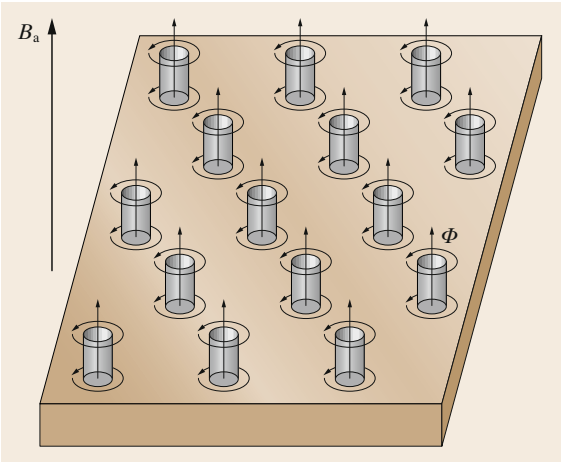


Fig. 50.13 Type II superconductor in the mixed state. Each vortex contains just a single fluxoid $\Phi_0 = h/2|e|$. (After [50.39, p. 74])

superconducting and the normal, which cannot coexist. On the other hand, three different phases can exist in type II superconductors. Below the lower critical field $B_{c1}(T)$, the type II superconductor is in the superconducting phase characterized by perfect diamagnetism, where it behaves like a type I superconductor. Above the upper critical field $B_{c2}(T)$, the material is in the normal state. In the mixed state between the lower and upper critical fields superconducting and normal regions can coexist in a type II superconductor. The mixed state shown in Fig. 50.13 is characterized by an array of normal cores within the superconductor, which allows magnetic flux to enter into the superconductor. The supercurrents encircling the normal cores (vortices) cause a repulsive vortex–vortex interaction.

In the mixed state of a type II superconductor, the transport currents can flow throughout the conductor. In the presence of a magnetic field perpendicular to

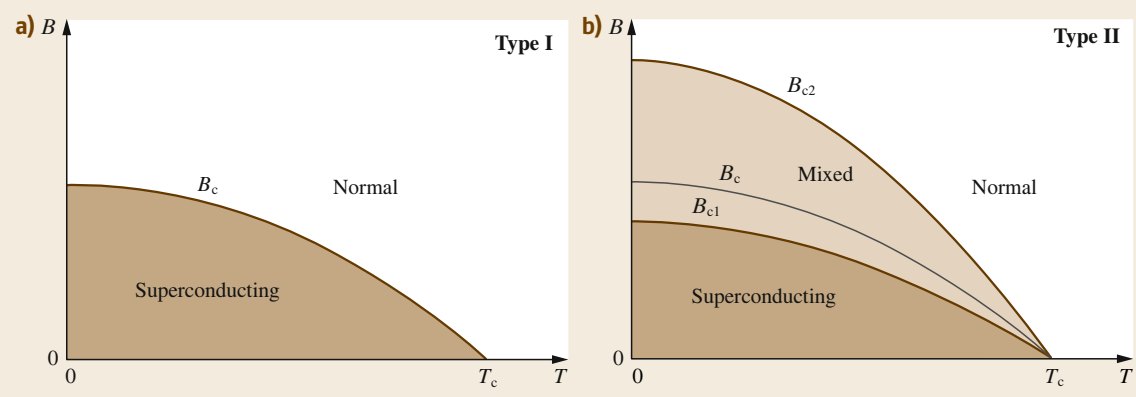


Fig. 50.12a,b Comparison of the phase diagrams for type I **(a)** and II superconductors **(b)**. (After [50.39, p. 74])

the current direction, a Lorentz force $F_L = j\phi_0$ acts on the vortices, which leads to a motion of the vortices perpendicular to the current and the field direction. The variation of the magnetic flux leads to a flux flow resistance, and hence the critical current would be zero. However, in a real type II superconductor defects are always present, which act as pinning centers. As long as the Lorentz force is smaller than the pinning force, flux motion is suppressed, resulting in zero resistance.

50.2 Cuprate High-T_c Superconductors: An Overview

50.2.1 Major Families of Cuprate Superconductors

At present, around 100 cuprate superconductors are known. The major families of cuprate superconductors are listed in Table 50.3. Common features of the cuprate superconductors are layered crystal structures. In the cuprates, conducting copper oxide blocks and insulating charge-carrier reservoirs alternate with each other along the crystallographic c direction. As a consequence of the layered structures, the physical properties of the cuprates are highly anisotropic. The electrical resistivity along the crystallographic a- and b-directions within the conductive CuO_2 planes is much lower than along the c-direction, where the current has to cross the insulating charge-carrier reservoirs. Figure 50.14 shows as an example the layering scheme of Tl-22$(n-1)n$ superconductors. The three thallium-based compounds are distinguished by the different number n of CuO_2 planes in the copper oxide blocks. In the compounds with $n \geq 2$, adjacent CuO_2 planes are separated by Ca layers. In the Bi-22$(n-1)n$ counterparts, Tl and Ba are replaced by Bi and Sr, respectively. In both families of cuprate superconductors the copper oxide blocks are embedded between TlO or BiO double layers. An analogous layering scheme results for the Tl-12$(n-1)n$ compounds. However, the copper oxide blocks of these compounds are separated only by single TlO layers. The highest critical temperature is typically reached in the compounds with three CuO_2 layers ($n = 3$) in the copper oxide blocks.

The physical properties of the cuprate superconductors depend strongly on their chemical composition and the resulting carrier concentration. As in semiconductors, the carrier concentration in the cuprates can be changed by doping. For example, let us consider the superconductor $La_{2-x}Sr_xCuO_4$. The possible oxidation states of copper are +1, +2, and +3. In the insulating parent compound La_2CuO_4, the oxidation state of copper is +2. The valence of La is +3 and that of Sr +2. Substitution of Sr^{2+} for La^{3+} leads to the formation of Cu^{3+} or O^- defects (holes), which are expected to be mobile. Thus, an increasing number of holes is created in the CuO_2 planes with increasing Sr concentration x.

The cuprates are complex nonstoichiometric oxides. Heat treatments in reducing or oxidizing atmosphere

Table 50.3 Major families of superconducting cuprates. (After [50.39, p. 79])

Family	Symbol	Maximum T_c (K)
$(La_{1-x}M_x)_2CuO_4$ [a]	La-214	39
$YBa_2Cu_3O_7$ [b]	Y-123	92
$YBa_2Cu_4O_8$	Y-124	80
$Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4}$	Bi-22$(n-1)n$	122
$Tl_2Ba_2Ca_{n-1}Cu_nO_{2n+4}$	Tl-22$(n-1)n$	128
$TlM_2Ca_{n-1}Cu_nO_{2n+3}$ [a]	Tl-12$(n-1)n$	122
$HgBa_2Ca_{n-1}Cu_nO_{2n+2}$	Hg-12$(n-1)n$	135

[a] M = Sr or Ba, [b] Y can be replaced by rare-earth elements

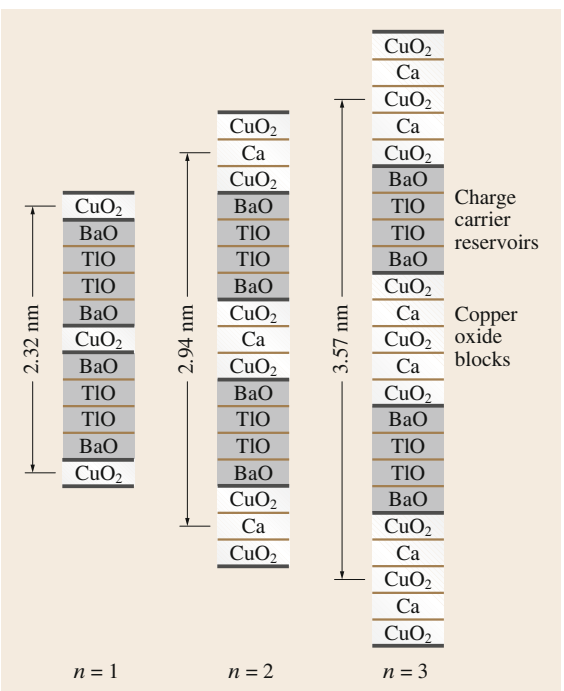


Fig. 50.14 Layering scheme of $Tl_2Ba_2Ca_{n-1}Cu_nO_{2n+4}$ superconductors. (After [50.39, p. 81])

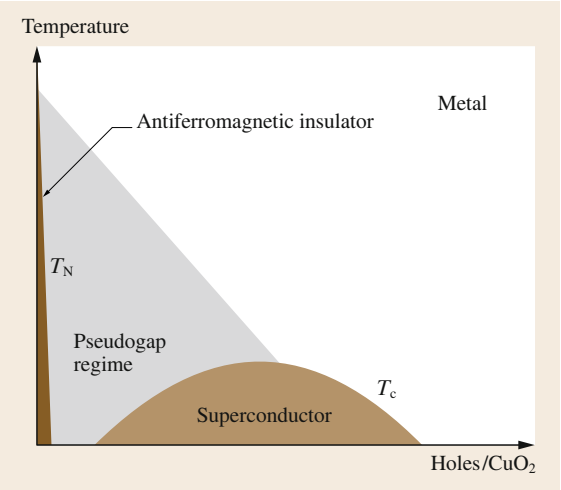


Fig. 50.15 Generic phase diagram for cuprate superconductors indicating that these materials are close to a metal–insulator transition. The physical properties of the cuprates vary considerably with increasing hole concentration ◄

therefore lead to considerable changes in the oxygen content. Incorporation of excess oxygen is a further way to enhance the hole concentration in the CuO_2 planes of the cuprates.

Finally, it should be noted that, in addition to the major families of cuprate superconductors listed in Table 50.3, there exist a large number of high-temperature superconductors which can be synthesized only under high pressure (see for example [50.39]).

50.2.2 Generic Phase Diagram of Cuprate Superconductors

It is beyond the scope of this chapter to discuss all the new physical phenomena related to cuprate superconductors. To give an impression of the complexity of these materials, the generic phase diagram is briefly described. The one shown in Fig. 50.15 indicates that the cuprate superconductors are close to a metal–insulator transition. Their parent compounds are antiferromagnetic Mott insulators, of which La_2CuO_4 and $YBa_2Cu_3O_6$ are examples. The Néel temperature T_N, at which magnetic ordering sets in, reaches values of 250–400 K in the parent compounds. The Néel temperature decreases rapidly with increasing hole concentration. A transition from an insulating to a superconducting phase is observed for further enhanced doping levels. The critical temperature increases with increasing number of holes and reaches a maximum at ≈ 0.18 holes per CuO_2 unit before decreasing again. At doping levels well above 0.3 holes/CuO_2, the material behaves as a nonsuperconducting metal.

In the last few years it could be clarified that the energy gap in cuprate superconductors is of d-wave symmetry, whereas it is of s-wave symmetry in conventional superconductors. The two different symmetries are illustrated in Fig. 50.16. For s-wave symmetry, the superconducting energy gap is constant irrespective of the direction of the wave vector $\boldsymbol{k}$. In the case of d-wave symmetry, there exist positions on the Fermi surface for which the energy gap is zero. At these positions, the sign of the order parameter changes from + to − or vice versa. As a consequence, a finite tunnel conductivity $[\mathrm{d}I/\mathrm{d}V \propto D(E)]$ is observed within the energy gap (Fig. 50.16).

Next, let us consider the temperature dependence of the energy gap. The tunneling conductance of under-

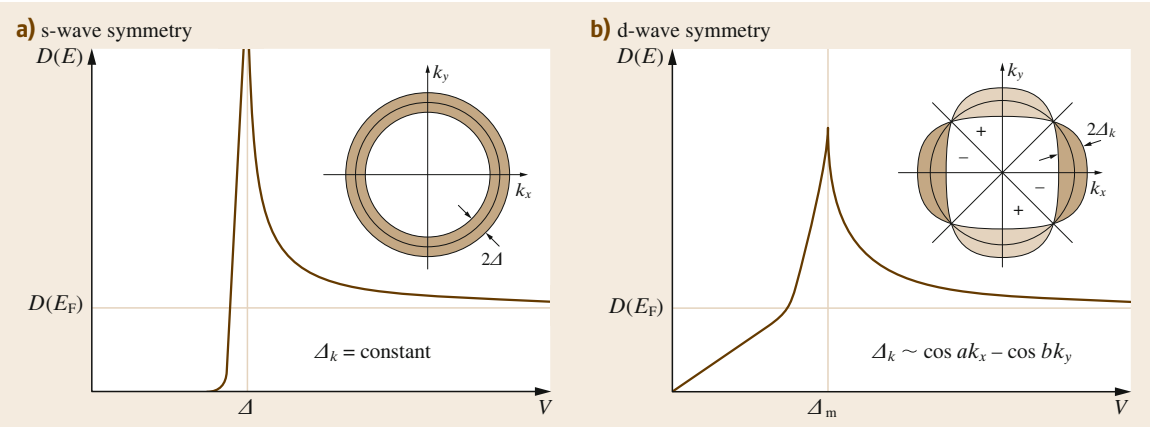


Fig. 50.16a,b The two different symmetries of the energy gap and the corresponding density of states. For s-wave symmetry **(a)**, the width of the energy gap is independent of the wave vector $\boldsymbol{k}$. In the case of the d-wave symmetry **(b)**, the energy gap depends on the wave vector. At four positions on the Fermi surface, the energy gap is zero. (After [50.50])

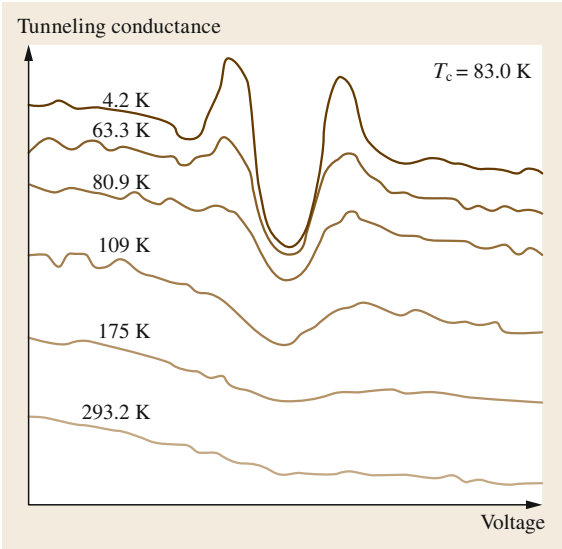


Fig. 50.17 Schematic illustration of the tunneling conductance of under-doped Bi-2212 providing the density of states in the vicinity of the energy gap. For clarity the spectra are offset vertically. A gap-like feature (pseudogap) can be observed even at a temperature as high as 175 K, well above T_c. (After [50.51])

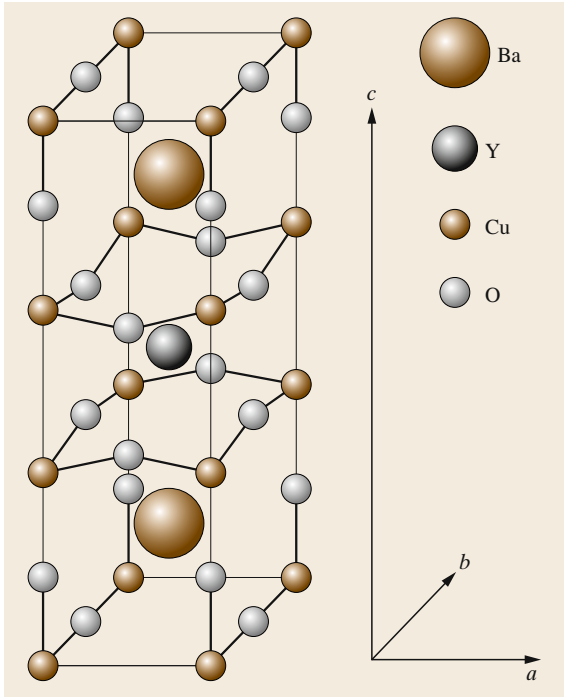


Fig. 50.18 Orthorhombic unit cell of $YBa_2Cu_3O_{7-\delta}$. A special feature of this cuprate are CuO chains in the *b*-direction. (After [50.39, p. 100])

doped Bi-2212 as shown in Fig. 50.17 indicates that the energy gap of cuprate superconductors is independent of temperature. This behavior is in contrast to the predictions of the BCS theory. In under-doped cuprate superconductors, the energy gap has been found to increase for decreasing number of charge carriers in spite of a reduced critical temperature. On the other hand, both the energy gap and the critical temperature of over-doped cuprates decrease with increasing carrier concentration. An outstanding feature of cuprate superconductors is the existence of an energy gap even at temperatures well above T_c. This pseudogap is found in the region of the phase diagram where a transition from an antiferromagnetic insulator to a superconductor occurs.

A potential explanation for the existence of the pseudogap is the formation of Cooper pairs at temperatures well above T_c. However, because of the low Cooper pair density, they are not macroscopically correlated. As a consequence, the material is not able to carry a supercurrent above T_c.

The Nernst effect is a thermo-magnetic phenomenon, in which a transverse voltage originates from a transverse magnetic field and a thermal gradient along the sample. In the presence of vortices, the Nernst effect is strongly enhanced. Measurements of the Nernst effect [50.52, 53] provided evidence for pairing well above the critical temperature in under-doped cuprate superconductors. However, the maximum temperature for the existence of vortices seems to be well below the temperature at which the pseudogap vanishes.

Newer results seem to favor the hypothesis that the pseudogap phenomenon competes with the gaps associated with superconductivity. A combined scanning tunneling microscopy and angle-resolved photoemission study of $Bi_2Sr_{2-x}La_xCu_{6+\delta}$ showed two distinct energy gaps. In the nodal direction, a small gap was found, which vanished at the critical temperature, while in the anti-nodal direction a large one was observed to exist well above the critical temperature. The two gaps seem to coexist in the antinodal direction [50.54]. In the nodal direction the energy gap of a d-wave superconductor vanishes.

New theories have predicted competing orders in the pseudogap regime of under-doped cuprate superconductors. As an example, let us consider charge stripes. Spin and charge are separated, resulting in antiferromagnetic domains and charge stripes located at the antiphase boundaries [50.55]. Experimental evidence for this scenario has been found from neutron scattering [50.56].

50.2.3 Crystal Structures

Next, let us consider the crystal structures of the cuprates in more detail. For $YBa_2Cu_3O_{7-\delta}$ (Y-123)

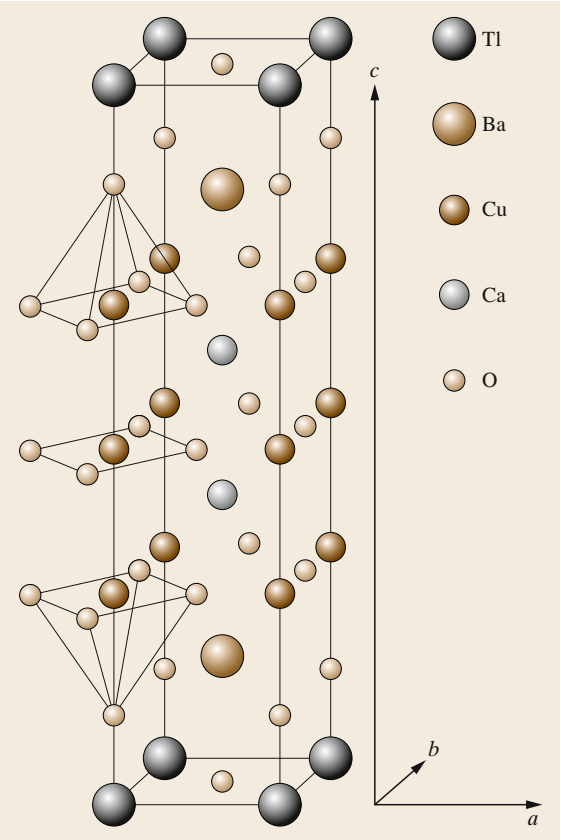


Fig. 50.19 Primitive tetragonal unit cell of $TlBa_2Ca_2Cu_3O_9$. (After [50.39, p. 108])

an orthorhombic structure with lattice parameters $a = 0.38\,\text{nm}$, $b = 0.39\,\text{nm}$ and $c = 1.17\,\text{nm}$ has been deduced from x-ray powder diffraction data [50.57, 58]. Figure 50.18 shows the orthorhombic unit cell of Y-123. The unit cell of $YBa_2Cu_3O_{7-\delta}$ contains double CuO_2 layers with embedded yttrium ions (Fig. 50.18). Slightly different z positions of the Cu and O atoms lead to buckling in the CuO_2 planes. The oxygen atoms occupy perovskite-like anion positions halfway between the copper atoms. These square-coordinated CuO_2 layers are a characteristic feature of all cuprate superconductors. Because metallic conductivity has been found in these layers, they are also called conduction planes. Along the crystallographic c-axis perpendicular to these planes, insulating or semiconducting behavior is frequently observed. Neighboring CuO_2 blocks are separated by two BaO layers. The CuO chains between the two BaO layers in the b-direction are a special feature of the Y-123 compound.

The crystal structures of Tl-12$(n-1)n$ and the analogous Hg-12$(n-1)n$ superconductors are primitive tetragonal. Because of the common CuO_2 planes, the a and b lattice parameters of cuprate superconductors are generally close to 0.39 nm. In both cuprate superconductor families, the c lattice parameter is given approximately by $c \cong (0.63 + 0.32n)\,\text{nm}$. As an example of these crystal structures the primitive tetragonal unit cell of $TlBa_2Ca_2Cu_3O_9$ is shown in Fig. 50.19. The copper oxide blocks of the Tl-1223 compound are formed of three CuO_2 layers. The CuO_2 layers within the copper oxide blocks are separated by

Table 50.4 Lattice parameters of selected cuprate superconductors. (After [50.39, p. 118])

Compound	Symmetry	a (nm) b (nm)	c (nm)
$La_{1.85}Sr_{0.15}CuO_4$	Tetragonal	0.3779	1.3200
$YBa_2Cu_3O_{6.9}$	Orthorhombic	0.3822 0.3891	1.1677
$Bi_{2.2}Sr_2Ca_{0.8}Cu_2O_{8+\delta}$[a]	Orthorhombic	0.5414 0.5418	3.089
$(Bi, Pb)_2Sr_{1.72}Ca_2Cu_3O_{10+\delta}$[a]	Orthorhombic	0.5392 0.5395	3.6985
$Tl_{1.7}Ba_2Ca_{1.06}Cu_{2.32}O_{8+\delta}$	Tetragonal	0.3857	2.939
$Tl_{1.64}Ba_2Ca_{1.87}Cu_{3.11}O_{10+\delta}$	Tetragonal	0.3822	3.626
$Tl_2Ba_2Ca_3Cu_4O_{12-\delta}$	Tetragonal	0.3850	4.1984
$Tl_{1.1}Ba_2Ca_{0.9}Cu_{2.1}O_{7.1}$	Tetragonal	0.3851	1.2728
$Tl_{1.1}Ba_2Ca_{1.8}Cu_{3.0}O_{9.7}$	Tetragonal	0.3843	1.5871
$TlBa_2Ca_3Cu_4O_{11}$	Tetragonal	0.3848	1.9001
$HgBa_2CuO_{4+\delta}$	Tetragonal	0.380	0.9509
$HgBa_2CaCu_2O_{6+\delta}$	Tetragonal	0.3859	1.2657
$HgBa_2Ca_2Cu_3O_{8+\delta}$	Tetragonal	0.3853	1.5818
$HgBa_2Ca_3Cu_4O_{10+\delta}$	Tetragonal	0.3854	1.9006

[a] Using a pseudo-tetragonal unit cell a lattice parameter a (tetragonal) $\cong (a+b)/2\sqrt{2} \approx 0.38\,\text{nm}$ would result

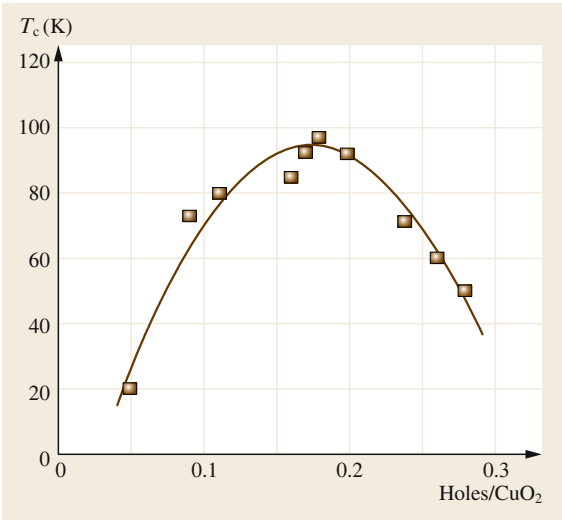


Fig. 50.20 Critical temperature of $HgBa_2CuO_{4+\delta}$ versus hole concentration. The maximum critical temperature is reached for a hole concentration of 0.18 holes/CuO_2. (After [50.39, p. 123], results from [50.59])

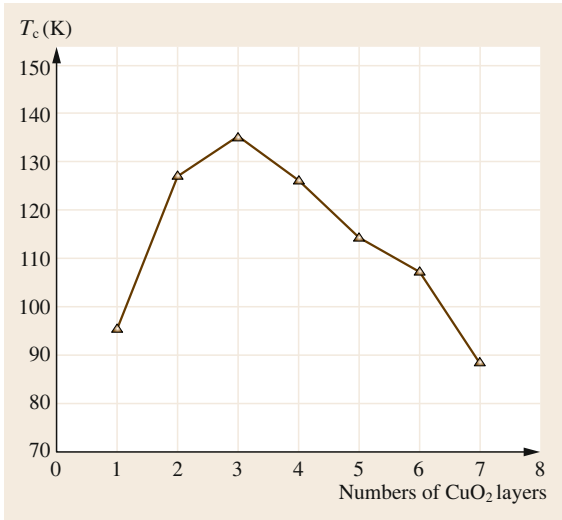


Fig. 50.21 Critical temperature of Hg-12$(n-1)n$ superconductors versus the number n of CuO_2 planes in the perovskite (copper oxide) blocks. (After [50.64])

Table 50.5 Critical temperatures of optimally doped cuprate superconductors

Compound	n	T_c (K)	Reference
Y-123	2	92	[50.21]
Bi-2212	2	96	[50.60]
Bi-2223	3	122	[50.61]
Tl-2201	1	92	[50.62]
Tl-2212	2	118	[50.62]
Tl-2223	3	128	[50.63]
Tl-2234	4	119	[50.61]
Tl-1212	2	90	[50.61]
Tl-1223	3	122	[50.61]
Tl-1234	4	122	[50.61]
Tl-1245	5	110	[50.61]
Hg-1223	3	135	[50.14]

Ca ions. In contrast to the Tl-22$(n-1)n$ compounds (Fig. 50.14) adjacent CuO_2 blocks are separated by only a single TlO layer. Like the TlO double layers of the 22$(n-1)n$ compounds, the single TlO layers of the 12$(n-1)n$ compounds are embedded between two BaO layers. The reduced distance of neighboring copper oxide blocks leads to an enhanced coupling of the conduction planes. This improved coupling is reflected in less anisotropic superconducting properties (Sect. 50.3). Defects in the BaO/TlO/BaO layers forming the charge-carrier reservoirs provide holes for the doping of the CuO_2 planes. In the Tl-12$(n-1)n$ compounds, the BaO layers can be replaced by SrO layers. The substitution of Tl by Hg leads to the structurally analogous Hg-12$(n-1)n$ counterparts. The lattice parameters of selected cuprate superconductors are listed in Table 50.4.

50.2.4 Critical Temperatures

Next, some of the factors determining the critical temperature of cuprate superconductors will be discussed. As shown in the generic phase diagram for cuprate superconductors (Fig. 50.15), the critical temperatures of these materials are closely connected to the hole concentration p in the CuO_2 planes, which is determined by the formal valence $2+p$ of the copper atoms in the conduction planes. Generally, an inverse parabolic dependence of the critical temperature on the number of holes per CuO_2 has been found. The maximum T_c is typically reached for a hole concentration p of 0.15–0.2 holes per CuO_2. For hole concentrations above 0.2 holes per CuO_2 the critical temperature decreases with increasing carrier concentration, leading to a nonsuperconducting metal for $p > 0.3$ holes per CuO_2. As an example of this behavior, the critical temperature of $HgBa_2CuO_{4+\delta}$ as function of the hole concentration is shown in Fig. 50.20.

In addition to the carrier concentration, the number of CuO_2 planes in the perovskite blocks is of importance for the critical temperature. The maximum critical temperatures of selected cuprate superconductors listed in Table 50.5 clearly indicate that the critical temperature increases with increasing number n of CuO_2 planes in the copper oxide blocks, at least up to the compound with $n = 3$. For compounds with four or more CuO_2 planes in the perovskite blocks, the critical tem-

perature decreases with increasing n. In Fig. 50.21, this behavior is illustrated for the Hg-12$(n-1)n$ compounds. A possible explanation for the observed behavior is the fact that in cuprates with $n \geq 3$, the hole concentrations in the central and outer CuO_2 layers may be considerably different. It may therefore be impossible to reach simultaneously the optimum carrier concentration in all the CuO_2 layers for compounds with $n \geq 4$.

Finally, it should be mentioned that the critical temperature of cuprate superconductors has been found to be connected with the lattice parameters. Consequently, the critical temperature depends on applied pressure. Application of a hydrostatic pressure typically enhances the critical temperature of cuprate superconductors. In Hg-1223, superconductivity was observed at temperatures as high as 164 K at a pressure of 31 GPa [50.65]. A more complex situation results for the application of uniaxial pressure. In the case of pressure along the crystallographic a- or b-axis the critical temperature is enhanced, whereas for uniaxial pressure along the c-direction T_c is reduced [50.66, 67].

50.3 Physical Properties of Cuprate Superconductors

50.3.1 Anisotropic Superconductors

As a consequence of the layered crystal structures of the cuprates, the penetration depth depends on the direction of the screening currents. Assuming that the screening currents in the a- and b-directions are equal, two different penetration depths λ_{ab} and λ_c result. The indices give the direction of the screening currents. Thus, two different Ginzburg–Landau parameters $\kappa_{ab}(B||ab)$ and $\kappa_c(B||c)$ can be defined

$$\kappa_{ab} = \left(\frac{\lambda_{ab}\lambda_c}{\xi_{ab}\xi_c}\right)^{1/2} ,$$
$$\kappa_c = \frac{\lambda_{ab}}{\xi_{ab}} , \qquad (50.20)$$

where ξ_{ab} and ξ_c are the coherence lengths in the ab-plane and along the c-direction, respectively. The dimensionless anisotropy parameter is given by [50.68]

$$\gamma_a = \left(\frac{m_c}{m_{ab}}\right)^{1/2} = \frac{\lambda_c}{\lambda_{ab}} = \frac{\xi_{ab}}{\xi_c} = \frac{B_{c2,ab}}{B_{c2,c}} = \frac{B_{c1,c}}{B_{c1,ab}} , \qquad (50.21)$$

where m_{ab} and m_c are the effective masses of the charge carriers for in-plane and out-of-plane currents, respectively. The expressions for the upper critical fields are [50.48, 68]

$$B_{c2,c} = \frac{\phi_0}{2\pi\xi_{ab}^2} ,$$
$$B_{c2,ab} = \frac{\phi_0}{2\pi\xi_c\xi_{ab}} . \qquad (50.22)$$

The upper and lower critical fields of various cuprate superconductors are listed in Table 50.6. The lower critical fields are typically between 10 and 500 mT for magnetic fields applied along the c-direction. The upper critical fields of cuprate superconductors can exceed 100 T even for the unfavorable field direction $B||c$. The data for Tl-2201, Hg-1201, and Hg, Cu-1201 given in Table 50.6 clearly indicate that the critical fields, like the critical temperatures, depend on the actual doping level.

For the characteristic length scales in cuprate superconductors, the following relation has been found

$$\xi_c < \xi_{ab} \ll \lambda_{ab} < \lambda_c . \qquad (50.23)$$

Table 50.7 lists for various cuprate superconductors the penetration depths and the coherence lengths. The in-plane coherence length ξ_{ab} is typically 1–4 nm. The considerably smaller values found for the out-of-plane coherence length ξ_c are comparable to interatomic distances. The small values of ξ_c are responsible for the insulating character of the charge-carrier reservoirs (Fig. 50.14) separating adjacent CuO_2 blocks. The cuprates are therefore nearly two-dimensional superconductors with intrinsic Josephson junctions along the c-direction. Moreover, the cuprates are extreme type II superconductors with κ_c values around 100.

50.3.2 Irreversibility Line

The magnetic phase diagram of cuprate superconductors is much more complicated than that of metallic superconductors. It is beyond the scope of this chapter to describe the details of the magnetic phase diagram of cuprate superconductors. A simplified magnetic phase diagram is shown in Fig. 50.22. An outstanding feature of high-temperature superconductors is the existence of an irreversibility line well below the upper critical field. Above this line, the vortices are movable, and hence the critical current density is zero. The irreversibility line

Table 50.6 Lower and upper critical fields of various cuprate superconductors

Compound	$B_{c1,c}$ (mT)	$B_{c2,c}$ (T)[a]	Comments	Reference
La, Sr-214		53		[50.69]
La-214		36/50	Ba/Sr doped	[50.70]
Y-123		122	$B_{c2,ab} = 674$ T	[50.71]
Y-123	500	29	$B_{c2,ab} = 140$ T	[50.72]
Y-123	150		$B_{c1,ab} = 24$ mT	[50.73]
Y-123	85		$B_{c1,ab} = 25$ mT	[50.74]
Bi-2201		20.2		[50.75]
Bi-2212		107		[50.76]
Bi-2212		31	$B_{c2,ab} = 476$ T	[50.77]
Bi-2212		71.5	$B_{c2,ab} = 542$ T	[50.78]
Bi, Pb-2212		83/95		[50.79]
Bi-2223		39	$B_{c2,ab} = 1210$ T	[50.80]
Bi, Pb-2223		184		[50.81]
Bi, Pb-2223	13.5			[50.82]
Tl-2201		0.85/1.9	$T_c = 9.7/15.7$ K	[50.83]
Tl-2212		99		[50.76]
Tl-2212	30		$B_{c1,ab} = 0.8$ mT	[50.73]
Tl-2223		75		[50.76]
Tl-1223	37	75		[50.84]
Hg-1201		100		[50.85]
Hg-1201	12.4/8.1		$T_c = 82/87$ K	[50.86]
Hg, Cu-1201		67/160	$T_c = 85/96$ K	[50.87]
Hg-1223		190		[50.42]

[a] Including the upper critical field data of polycrystalline samples

Table 50.7 Characteristic lengths for selected cuprate superconductors. (Data from [50.39, 42, 85, 87, 88])

Compound	λ_{ab} (nm)	ξ_{ab} (nm)	ξ_c (nm)
La, Sr-214	250–410	2.1	
Y-123	130–240	1.3–1.6	0.2–0.3
Bi-2201	438	4.0	
Bi-2212		2.7–3.8	0.16–0.18
Bi-2223		2.9	0.093
Bi, Pb-2212	178	2.0	
Bi, Pb-2223	88–245	1.35–2.9	
Tl-2201		5.2	0.3
Tl-2212		2.0–3.1	0.03–0.68
Tl-2223	117–163	1.1–1.36	
Tl-2234		4.5	1
Tl-1223	137	2.1	
Hg-1201	117–216[a]	1.7–3.4[a]	
Hg, Cu-1201	162–283[a]	1.43–2.2[a]	
Hg-1223	130–150	1.3	

[a] Different doping levels

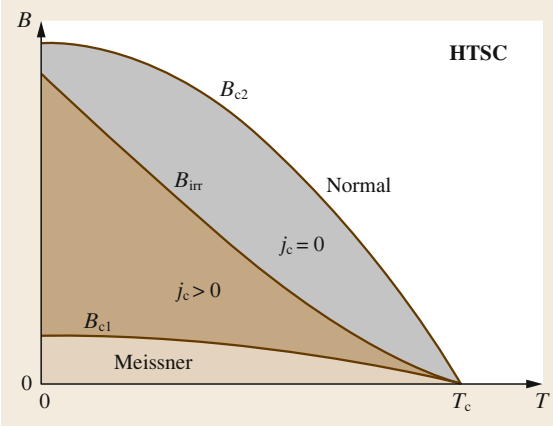


Fig. 50.22 Simplified magnetic phase diagram of cuprate superconductors. Above the irreversibility line, the vortices are movable, and hence the critical current density is zero. (After [50.39, p. 138])

can be described by the expression

$$B_{irr} = B_0 \left(1 - \frac{T_{irr}}{T_c}\right)^{\alpha}, \qquad (50.24)$$

where B_{irr} is the irreversibility field, T_{irr} is the irreversibility temperature, B_0 is a fitting parameter, and the exponent α is characteristic of the superconductor in question. In the case of Y-123, the exponent α is close to 1.5 for magnetic fields perpendicular to the crystallographic c-direction, while values around 1.3 have been found for $B||c$ [50.73].

The existence of the irreversibility line is reflected in the broadening of the resistive transition in the pres-

ence of an applied magnetic field. Figure 50.23 shows the resistive transition of a Bi-2212 single crystal for magnetic fields of 2, 5, and 12 T parallel to the *c*-direction. The data for zero applied field are also shown in Fig. 50.23. The end point of the resistive transition corresponds to the irreversibility temperature for the magnetic field applied. On the other hand, the onset of the resistivity drop may be used to determine the upper critical field.

The existence of an irreversibility line can be explained in two different ways. In the more conventional picture, the occurrence of resistance above the irreversibility line is caused by thermally activated depinning. Within this picture, it can be easily understood that the irreversibility line of cuprate superconductors is shifted to higher fields and temperatures by heavy-ion irradiation [50.90–92]. The irradiation generates amorphous tracks, which are efficient pinning centers. An alternative scenario is the melting of the flux-line lattice. Evidence for this possibility is the latent heat measured at the phase transition [50.93, 94].

The maximum operating field of cuprate superconductors is limited by the irreversibility line for the unfavorable field direction $B||c$. In any comparison of the behavior of different cuprates the irreversibility field as a function of the reduced temperature T/T_c should be considered. The highest irreversibility fields result for Y-123, while the lowest values of B_{irr} have been found for Bi- and Tl-22$(n-1)n$ compounds. Intermediate irreversibility fields result for Tl- and Hg-12$(n-1)n$ compounds [50.39, 95]. The observed order in the irreversibility fields indicates that flux pinning is more efficient in cuprates with less anisotropic physical properties. The largest anisotropy of the physical properties has been observed in Bi- and Tl-22$(n-1)n$ compounds with the largest distance of adjacent CuO_2 blocks. The relatively small anisotropy of Y-123 seems to be the consequence of the short-circuiting of the insulating charge-carrier reservoirs by the CuO chains (Fig. 50.18).

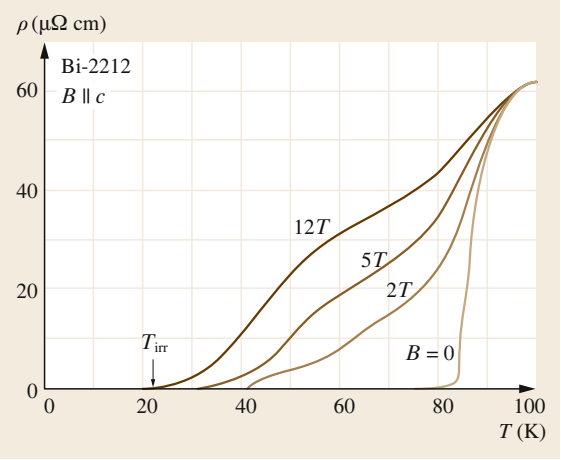


Fig. 50.23 Resistivity versus temperature curves of a Bi-2212 single crystal for magnetic fields parallel to the *c*-direction. (After [50.39, p. 139], results from [50.89])

50.3.3 Limitations of the Transport Critical Current

The electric field generated in a superconductor at the transition to the normal conducting state is typically well described by a power law $E = E_c(I/I_c)^n$, where I is the current, n is the index of resistive transition, and E_c the electric field used to define the critical current I_c. Typically E_c values of 0.1 or 1 μV/cm are used to determine the critical current.

It should be recognized that, unlike the critical temperature and the upper critical field, the critical current density is not an intrinsic material property. The achievable j_c values are determined mainly by the microstructure of the superconductor in question. Grain boundaries can act as weak links, leading to small critical current densities in polycrystalline cuprate superconductors. In the absence of grain-boundary weak links, the transport j_c is mainly determined by flux pinning. Before these effects are discussed, the anisotropy of j_c in individual grains and epitaxial thin films is considered.

Anisotropy of the Critical Current Density

In cuprate superconductors, conductive copper oxide blocks and insulating charge-carrier reservoirs alternate along the crystallographic *c*-direction. The currents can easily flow in the conductive CuO_2 planes, whereas currents along the *c*-direction have to cross the insulating or semiconducting charge-carrier reservoirs (tunneling). In spite of j_c values as high as 10^8 A/cm^2 in the CuO_2 planes of Hg-1212 films, the estimated j_c along the *c*-direction is as low as 5000 A/cm^2 [50.96].

A further anisotropy of the critical current density is related to the direction of the applied magnetic field. Figure 50.24 shows the critical current density of an epitaxial Bi -2212 film as a function of the applied magnetic field for selected temperatures. The critical current density declines rapidly with increasing applied field for $B||c$, while the field dependence of j_c is much less pronounced for $B||ab$. This behavior reflects the considerably different upper critical fields for these two field directions. The angle dependence of j_c in epitaxial Bi-2212 and Bi-2223 films can be described by the expression

$$j_c(B,\theta) = j_{c,c}(B\sin\theta)\,, \qquad (50.25)$$

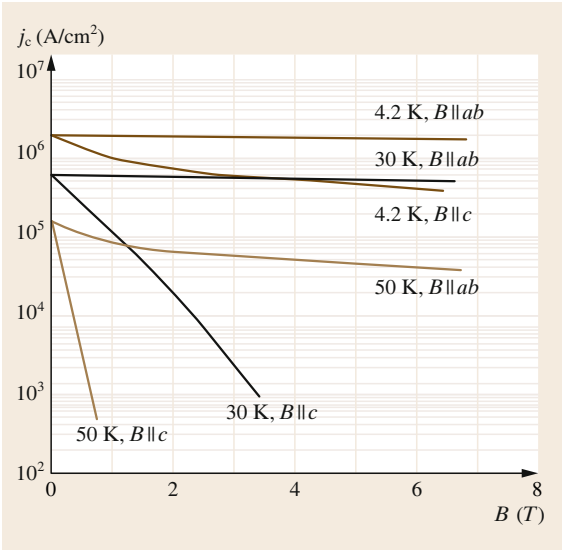


Fig. 50.24 Critical current densities in epitaxial Bi-2212 films as a function of applied magnetic field for B along the ab-planes and parallel to the crystallographic c-direction. (After [50.97])

where θ is the angle between the ab-planes and the field direction, and $j_{c,c}$ is the critical current density for $B||c$. The above expression states that the reduction of the critical current by an applied magnetic field is determined exclusively by the field component parallel to the c-direction.

Grain-Boundary Weak Links

Soon after the discovery of superconductivity above 77 K in Y-123, it was noticed that the transport j_c in sintered samples is typically well below 1000 A/cm^2. In addition, magnetic fields of a few mT are sufficient to reduce the j_c by an order of magnitude. On the other hand, critical current densities well above 10^6 A/cm^2 could be achieved in epitaxial Y-123 films. These experimental results suggest that the grain boundaries present in polycrystalline material can act as barriers for the transport current. Due to extremely short coherence lengths, $\xi_c < 1$ nm, the Cooper pair density can be considerably reduced within the width of a grain boundary. Two grains separated by a grain boundary can therefore be considered as a Josephson junction.

The high j_c values achieved in highly textured films suggest that low-angle grain boundaries are strongly coupled. To clarify this aspect, the transport properties of artificial grain boundaries have been studied [50.98, 99]. In epitaxial growth the crystallography prescribed by the substrate can be transmitted to the Y-123 film. Using $SrTiO_3$ bicrystals, two Y-123 grains connected by an artificial grain boundary with a well-defined mis-

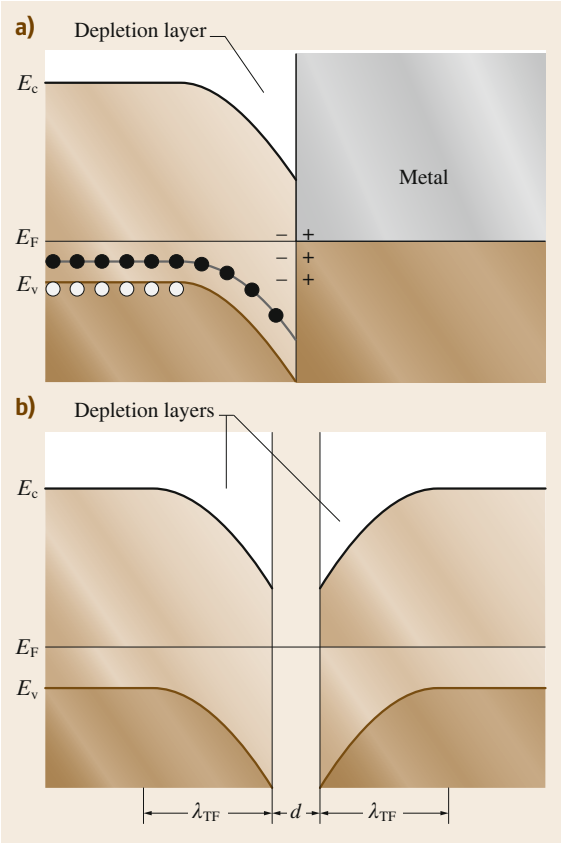


Fig. 50.25a,b Space-charge effects and band bending at a semiconductor–metal interface **(a)** and at a grain boundary in a cuprate superconductor **(b)**. (After [50.102])

orientation angle θ can be prepared. The critical current density across the grain boundaries has been found to decrease rapidly with increasing misalignment angle. For [001]-tilt boundaries (rotation of the second grain around the common c-axis) in Y-123 films, the grain boundary j_c depends exponentially on the misalignment angle θ

$$j_c = j_{c0} \exp\left(-\frac{\theta}{\theta_0}\right), \qquad (50.26)$$

where $\theta_0 \cong 6.3°$ and $j_{c0} \cong 2 \times 10^7$ A/cm^2 at 4.2 K [50.100]. The corresponding values for Bi-2223 [001]-tilt grain boundaries are $j_{c0} = 6 \times 10^6$ A/cm^2 and $\theta_0 = 5°$ at 26 K [50.101].

The weak-link behavior is closely related to the strong reduction of the Cooper pair density in the grain boundaries caused by the extremely short coherence lengths and the low carrier concentrations in cuprate superconductors. The carrier concentration in optimally doped cuprate superconductors is around 5×10^{21} cm^{-3},

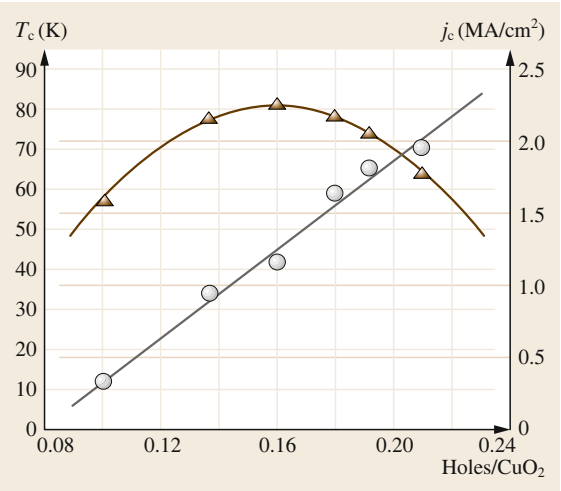


Fig. 50.26 Critical temperature (*triangles*) and current density (*circles*) of a symmetric 24° [001]-tilt boundary as a function of the hole concentration. The charge-carrier concentration in the bicrystalline $Y_{0.7}Ca_{0.3}Ba_2Cu_3O_{7-\delta}$ films has been modified by annealing at different oxygen pressures. (After [50.100])

whereas the carrier concentration in copper is as high as $85\times10^{21}\,\mathrm{cm}^{-3}$. Furthermore, the behavior of grain boundaries in cuprate superconductors is similar to that of semiconductor–metal interfaces [50.102, 103]. In the case of a p-type semiconductor in contact with a metal characterized by a smaller work function, electrons flow from the metal into the semiconductor. As a consequence, both the valence and the conduction band are bent downward (Fig. 50.25a). There an excess of negatively charged acceptor ions and a depletion of holes in the valence band result. Currents from the metal to the semiconductor would further enlarge the space-charge region and are therefore negligible (Schottky contact). Because of low carrier concentrations and large dielectric constants ($\varepsilon\approx10$), band-bending effects can occur in cuprate superconductors. The grain boundaries of cuprate superconductors may be considered as double Schottky barriers (Fig. 50.25b). The resulting Thomas–Fermi screening length is in the range of 0.5−1 nm, which is comparable to the coherence length. The width of the insulating layer at the grain boundary is $t=d+2l_\mathrm{d}$, where d is the structural width of the boundary and l_d the thickness of the depletion layer. The structural width of typical grain boundaries is ≈0.4 nm. On the other hand, the thickness of the depletion layer is 0.16 nm for $\varepsilon=10$, an in-built voltage of 0.1 V, and a carrier concentration of $4.5\times10^{21}\,\mathrm{cm}^3$ [50.102]. Consequently, the total thickness of the insulating layer is considerably enhanced by the space-charge region.

In such a scenario it should be possible to modify the electronic properties of the interface by increasing the carrier concentration. In Y-123 this can be done by Ca doping. Figure 50.26 shows for symmetric 24° [001]-tilt boundaries the critical temperature and current density as a function of the hole concentration. As expected the j_c values at 4.2 K increase with increasing hole concentration even in the over-doped region, where T_c decreases.

Flux Pinning

Transport currents without losses can only flow in a type II superconductor when the flux lines are effectively pinned. The screening currents circulating around each flux line are responsible for a repulsive vortex–vortex interaction, which prevents flux motion as soon as a sufficiently large fraction of the flux lines is pinned. The effectiveness of this collective pinning depends on the stiffness of the flux-line lattice.

The energy required to create a flux line can be estimated from the lower critical field B_{c1} [50.47]. At the lower critical field, flux lines start to enter the superconductor. The formation energy for each flux line is therefore

$$E_\mathrm{fl}=\int_0^{B_{c1}} m\mathrm{d}B \quad\text{with}\quad m=\frac{\phi_0}{\mu_0}l_\mathrm{fl}\,, \tag{50.27}$$

where m is the magnetic moment of a flux line of length l_fl containing a single flux quantum. The required energy per unit length is

$$\frac{E_\mathrm{fl}}{l_\mathrm{fl}}=\frac{\phi_0}{\mu_0}B_{c1}\,. \tag{50.28}$$

Using $\kappa=\lambda_\mathrm{L}/\xi$ the lower critical field can be written as [50.42]

$$B_{c1}=\frac{\phi_0\ln\kappa}{4\pi\lambda_\mathrm{L}^2}\,. \tag{50.29}$$

When the flux line is in a normal region, no energy is required to form the normal vortex core, and the resulting pinning energy per unit length is

$$\frac{E_\mathrm{fl}}{l_\mathrm{fl}}=\frac{1}{4\pi\mu_0}\left(\frac{\phi_0}{\lambda_\mathrm{L}}\right)^2\ln\left(\frac{\lambda_\mathrm{L}}{\xi}\right). \tag{50.30}$$

Generally, regions with reduced Cooper pair densities due to defects can act as pinning centers.

The pinning properties of the cuprate superconductors are highly anisotropic. Relatively strong pinning has been found for magnetic fields transverse to the current direction applied in the *ab* plane. For the resulting

Lorentz force along the *c*-direction, strong intrinsic pinning may be provided by the insulating charge-carrier reservoirs of the layered structure along the *c*-direction. On the other hand, the pinning is weak for magnetic fields transverse to the current direction applied along the *c*-direction, leading to a Lorentz force parallel to the *ab*-planes. Flux lines along the *c*-direction experience strong pinning only within the CuO_2 blocks. As a consequence of the layered structure, the vortices may be cut into vortex segments or even pancake vortices which exist only within the conductive CuO_2 blocks. These pancake vortices could easily move in the direction of the Lorentz force.

The positions of the irreversibility lines in the different cuprate superconductors reflect the anisotropy of the pinning properties. Typically, the pinning efficiency for magnetic fields along the *c*-direction decreases as one progresses from Y-123 via the Hg- and Tl-$12(n-1)n$ compounds to the highly anisotropic Bi- and Tl-$22(n-1)n$ cuprates.

50.4 Superconducting Films

Because of the weakly coupled large-angle grain boundaries present in polycrystalline bulk material, the transport critical current density is much smaller than the j_c of the individual grains. In contrast to polycrystalline bulk material, adjacent grains of epitaxial films are connected by low-angle grain boundaries. As a consequence, the critical current densities measured in epitaxial films are close to the j_c values of individual grains. At low temperatures and zero applied field, the critical current densities in epitaxial Y-123, Bi-2212, Bi-2223 and Hg-1223 films exceed 10^7 A/cm^2 [50.97, 104–106]. Even at 77 K j_c-values well above 10^7 A/cm^2 have been measured in Y-123 films [50.107].

The physical properties of ultrathin films can be considerably different from those of bulk single crystals. First, let us consider the effect of the film thickness on the critical temperature. For the growth of epitaxial ultrathin films, a substrate with good lattice matching is required. $YBa_2Cu_3O_7$ films of thickness equal to the height of a single unit cell in the *c*-direction can be grown onto $SrTiO_3$ single crystals covered by a six-unit-cell-thick nonsuperconducting $PrBa_2Cu_3O_7$ (Pr-123) buffer layer [50.108]. The lattice mismatch of Y-123 and Pr-123 is less than 1.5%. The layer growth can be checked by the observation of the strong intensity oscillations of reflection high-energy electron diffraction. The period of these oscillations has been found to correspond to the height of a single unit cell. Y-123 films with a thickness of 1−10 unit cell heights were grown by coevaporation of the metals under oxygen atmosphere. To provide hole doping the ultrathin Y-123 films are covered by a six-unit-cell-thick Pr-123 layer. The offset critical temperature of these Y-123 films is considerably lower than that of a 100-nm-thick film (Fig. 50.27). On the other hand, the onset critical temperature has been found to be much less affected. The reduction of the offset critical temperature has been attributed to the more pronounced critical fluctuations in the thinner films. Critical fluctuations are variations of the superconducting order parameter $\psi = n_C^{1/2}\exp(i\varphi)$ (n_C is the Cooper pair density) around the thermal equilibrium values.

The critical temperature of ultrathin Y-123 layers in Y-123/Pr-123 multilayers has been found to depend on the thicknesses of the insulating Pr-123 layers [50.109]. The thickness of 1.2 nm of the embedded Y-123 layers corresponds to the height of a single unit cell in the *c*-direction. The critical temperature of the Y-123/Pr-123 multilayer decreases from ≈ 50 K (10% of normal-state resistance) to ≈ 13 K for Pr-123 interlayer thicknesses of 1.2 and 7.2 nm, respectively. Because of the Josephson coupling of adjacent Y-123 layers in the multilayer the critical fluctuations are expected to be less pronounced than in a single Y-123 film

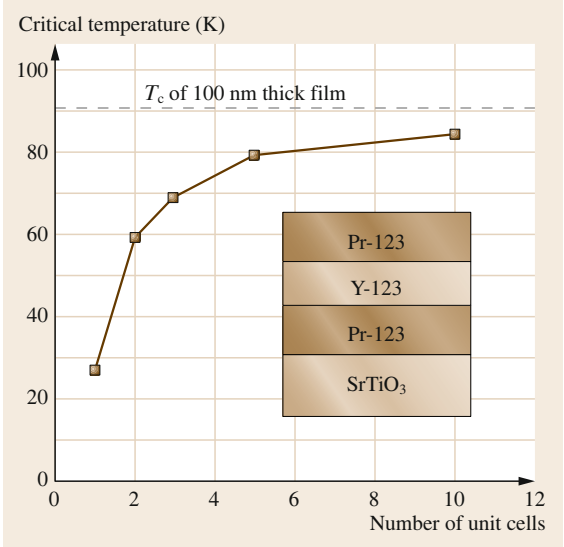


Fig. 50.27 Offset critical temperature of ultrathin Y-123 films as a function of the layer thickness normalized to the *c* lattice parameter. (After [50.108])

of the same thickness. Thus, the T_c of the Y-123 layers in the Y-123/Pr-123 multilayer with a Pr-123 thickness of 1.2 nm is higher than that of a single Y-123 layer of the same thickness (Fig. 50.27).

Finally, let us consider the effects of epitaxial strain on the critical temperature of (La, Sr)$_2$CuO$_4$ films [50.67, 110, 111]. In epitaxial growth, the crystallography prescribed by the substrate is transmitted to the film. For a sufficiently small film thickness, the lattice constant of the film is equal to that of the substrate. As a consequence of the lattice mismatch a considerable epitaxial strain can be induced in the film. Under epitaxial growth conditions, a compressive strain in the CuO_2 planes of (La, Sr)$_2$CuO$_4$ is accompanied by a tensile strain in the *c*-direction. Typically, in-plane compressive and out-of-plane tensile strains enhance the T_c of cuprate superconductors. Thus, the critical temperature of cuprate superconductors should be much more sensitive to epitaxial strain effects than to hydrostatic pressure, where the opposite in-plane and out-of-plane effects partly cancel each other. Because of film deposition at temperatures above 750 °C, differences in the thermal expansion coefficients provide an additional contribution to the strain in the films. In thick films the epitaxial strain is released by the appearance of misfit dislocations. The expected change in the critical temperature of the films is given by

$$T_c = T_c(0) + 2\frac{dT_c}{d\varepsilon_{ab}}\varepsilon_{ab} + \frac{dT_c}{d\varepsilon_c}\varepsilon_c \,, \qquad (50.31)$$

where $\varepsilon = (d_{bulk} - d_{strained})/d_{bulk}$, with d being a lattice parameter.

To study the epitaxial strain effects, *Locquet* et al. [50.67] deposited $La_{1.9}Sr_{0.1}CuO_4$ films ($a = 0.3784$ nm, $\alpha = 8.5\times10^{-6}\,K^{-1}$) of 10–15 nm thickness on $SrTiO_3$ ($a = 0.3905$ nm, $\alpha = 9\times10^{-6}\,K^{-1}$) and $SrLaAlO_4$ ($a = 0.3754$ nm, $\alpha = 10.5\times10^{-6}\,K^{-1}$) substrates. Using x-ray diffraction, strain values of $\varepsilon_{ab} = 0.63\%$ and $\varepsilon_c = -0.76\%$ have been determined for the La-214 film on $SrTiO_3$. The corresponding strain values for the La-214 film deposited onto $SrLaAlO_4$ are $\varepsilon_{ab} = -0.54\%$ and $\varepsilon_c = 0.35\%$. The results of *Locquet* et al. [50.67] are shown in Fig. 50.28. The critical temperatures of the La-214 films of 10 ($SrTiO_3$) and 49 K ($SrLaAlO_4$) are, as expected, considerably different from the bulk value of 25 K. In addition, the temperature dependencies of the normal state resistivities of the two films are remarkably different. The film experiencing a tensile in-plane strain shows an upturn of the normal state resistance at temperatures below 100 K, which is an effect also found in under-doped cuprate superconductors.

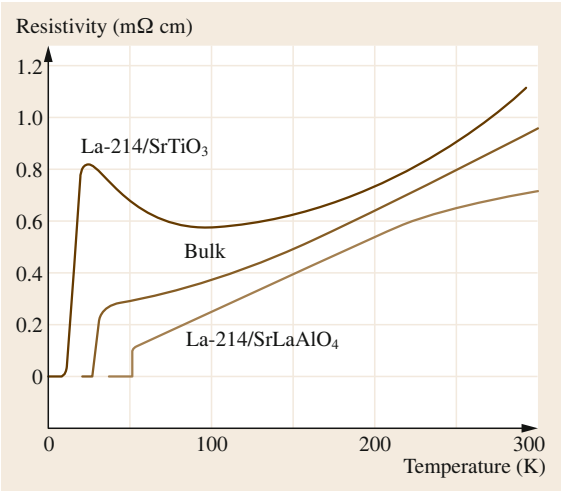


Fig. 50.28 Resistivity versus temperature for La-214 films on $SrTiO_3$ and $SrLaAlO_4$ substrates. The different epitaxial strains lead to considerably different critical temperatures of 10 and 49 K. For comparison the resistivity of La-214 bulk material ($T_c = 25$ K) is also shown. (After [50.67])

50.5 The Special Case of MgB_2

In spite of the fact that MgB_2 has been known since the early 1950s, it was unknown until 2001 that it is a superconductor with a remarkably high transition temperature. Figure 50.29 shows the hexagonal crystal structure of MgB_2, consisting of alternating layers of boron and magnesium. As a consequence of the layered structure, anisotropic physical properties result. The arrangement of the boron atoms is the same as that of the carbon atoms in graphite. Each of the boron atoms is surrounded by three other boron atoms at the same distance.

To study the role of phonons in the superconductivity of MgB_2, the critical temperatures of isotopically pure $Mg^{10}B_2$ and $Mg^{11}B_2$ specimens have been studied [50.117]. The transition temperatures of $Mg^{11}B_2$ and $Mg^{10}B_2$ defined by a 2% onset criterion are 39.2 and 40.2 K, respectively. Defining a partial boron isotope-effect exponent $\beta_B = \Delta \ln T_c / \Delta \ln M_B$, where M_B is the atomic mass of boron, a value of $\beta_B = 0.26 \pm 0.03$ results from the experimental data. On the other hand, the isotope effect for $^{26}MgB_2$ and $^{24}MgB_2$ is much less pronounced [50.118]. The shift of

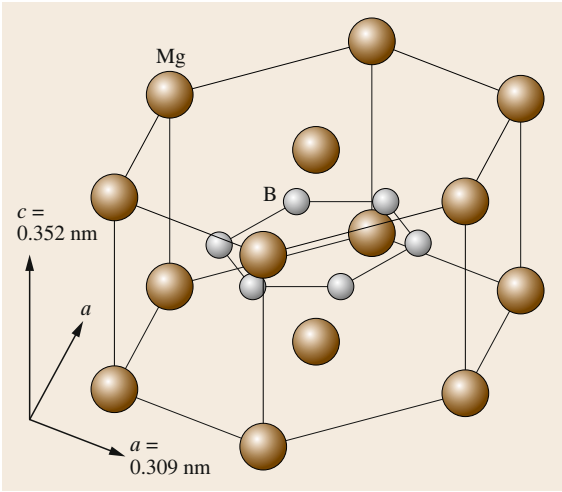


Fig. 50.29 Hexagonal crystal structure of MgB_2 consisting of alternating layers of magnesium and boron. (After [50.15])

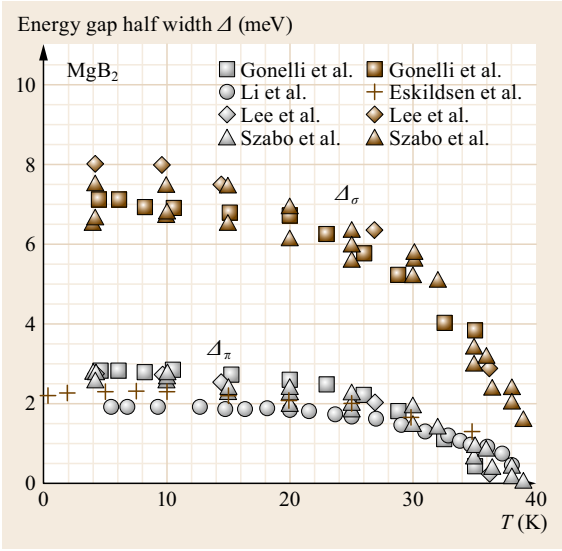


Fig. 50.30 Half-widths of the σ- and π-band gaps of MgB_2. (After [50.112–116])

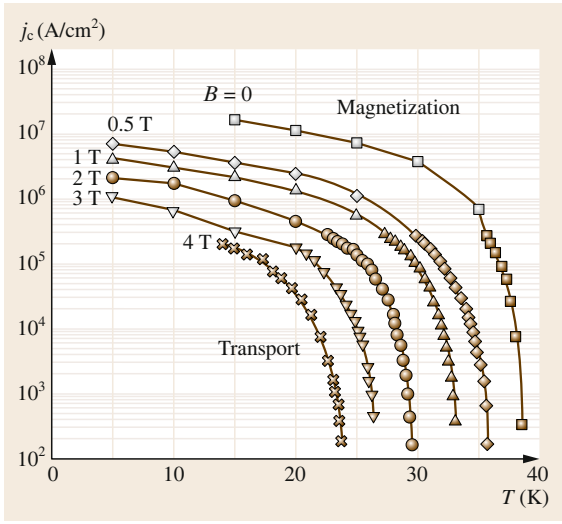


Fig. 50.31 Critical current density of a c-axis-oriented MgB_2 film as a function of temperature for magnetic fields up to 4 T. (After [50.133])

0.1 K in T_c corresponds to a partial Mg isotope-effect exponent β_{Mg} of 0.02. The strong boron isotope effect suggests that the pairing of the electrons is caused by the boron phonons. Thus, MgB_2 seems to be a conventional metallic superconductor with an extremely high transition temperature. The high transition temperature seems to be caused by the small atomic mass of boron, leading to high phonon frequencies.

The currently known physical properties of MgB_2 are summarized in Table 50.8. As in cuprate superconductors, the operation range of MgB_2 is limited by the irreversibility field. In polycrystalline bulk samples, the irreversibility field typically reaches 50% of B_{c2}. In nanocrystalline MgB_2 bulk material the irreversibility field is shifted to 80% of B_{c2}. The enhancement of the irreversibility field in the nanocrystalline material has been attributed to the improved pinning provided by the large number of grain boundaries [50.119]. Maximum irreversibility fields of more than 14 T at 4.2 K have been reported for MgB_2 bulk material annealed in Mg vapor [50.120]. MgB_2 is an extreme type II superconductor with a Ginzburg–Landau parameter κ of 38. The coherence length of $\xi_c \approx 2.5$ nm is considerably larger than the values of ≈ 0.2 nm measured for cuprate superconductors (Table 50.7). The anisotropy of the upper critical fields reported for c-axis-oriented films [50.121] and single crystals [50.122] is typically close to $B_{c2,ab}/B_{c2,c} \approx 3$.

A special feature of MgB_2 is the existence of two energy gaps. In MgB_2, layers of Mg^{2+} ions and metallic boron alternate along the crystallographic c direction (Fig. 50.29). The electronic structure of MgB_2 is similar to that of graphite. The highest occupied electronic states are σ- and π-bonding boron orbitals [50.134–136]. The incompletely filled σ-bands are confined to the boron planes, while the nonbonding π-electrons fill two three-dimensional metallic bands. Band structure calculations indicate that σ-electrons are transferred to the π-bands [50.134]. The superconductivity of MgB_2 is dominated by the σ-band holes, which are strongly coupled to the in-plane vibrations of boron atoms,

Table 50.8 Physical properties of MgB_2

Parameter	Value	Comments	Reference
T_c (K)	40.2	$Mg^{10}B_2$	[50.117]
Lattice parameters		Crystal structure	[50.15]
a (nm)	0.3086	Hexagonal	
c (nm)	0.3524	Space group P6/mmm	
ρ (mΩ cm)	1–18	at 40 K	[50.120]
ρ (300 K)/ρ(40 K)	15–3		[50.120]
B_{c1} (mT)	18–26.5	$T = 0$	[50.123, 124]
$B_{c1,ab}$ (mT)	22	Single crystal, $T = 0$	[50.125]
$B_{c1,c}$ (mT)	63	Single crystal, $T = 0$	[50.125]
$B_{c2,ab}$ (T)	19.6/26.4	Films T_c: 37.5/31.4 K, $T = 0$	[50.126]
$B_{c2,c}$ (T)	13.8/14.6	Films T_c: 37.5/31.4 K, $T = 0$	[50.126]
$B_{c2,ab}$ (T)	14.5	Single crystal, $T = 0$	[50.125]
$B_{c2,c}$ (T)	3.18	Single crystal, $T = 0$	[50.125]
B_{irr} (T)	$> 14/\approx 6$	$T = 4.2/20$ K	[50.120]
B_{irr} (T)	14.5–25.5	Tapes	[50.127, 128]
ξ_{ab} (nm)	6.8–8.1	$T = 0$	[50.124, 129]
ξ_c (nm)	2.3–2.7	$T = 0$	[50.124, 129]
λ (nm)	110–180	$T = 0$	[50.123, 130]
λ_{ab} (nm)	82	Single crystal, $T = 0$	[50.125]
λ_c (nm)	370	Single crystal, $T = 0$	[50.125]
$\kappa = \lambda/\xi$	38	Polycrystalline, $T = 0$	[50.123]
$\kappa_{ab} = (\lambda_{ab}\lambda_c/\xi_{ab}\xi_c)^{0.5}$	37.1	Single crystal, $T = 0$	[50.125]
$\kappa_c = \lambda_{ab}/\xi_{ab}$	8.1	Single crystal, $T = 0$	[50.125]
Δ (meV)	8.5–9.5[a]	$T = 4.2$ K	[50.131]
Δ_σ (meV)	6.5–8		[50.112–116]
Δ_π (meV)	1.9–2.8		[50.113, 115, 116]

[a] MgB_2 is a double gap superconductor (see for example [50.132])

leading to strong pairing and an energy gap with half-width $\Delta_\sigma \approx 7$ meV [50.135]. In addition, weaker pairs ($\Delta_\pi \approx 2$ meV) are formed by the π-electrons [50.135]. The measured values of the half-widths of the two energy gaps are presented in Fig. 50.30.

Immediately after the discovery of superconductivity in MgB_2, the current transport across grain boundaries was studied [50.137, 138]. The results of these investigations indicate that the grain boundaries in MgB_2 do not act as weak links. This behavior reflects the fact that the coherence length of MgB_2 is considerably larger than that of cuprate superconductors.

To give an impression of the current-carrying capacity of MgB_2, the critical current density of c-axis-oriented films as a function of temperature is shown in Fig. 50.31 for selected applied magnetic fields ($B||c$). The critical current density at 15 K and zero applied field is as high as 16 MA/cm^2 [50.133, 139], which is comparable to the j_c values of cuprate superconductors.

Finally, let us briefly consider the synthesis of MgB_2. Polycrystalline MgB_2 can be synthesized by the reaction of boron powder with magnesium vapor. Stoichiometric quantities of boron and magnesium are sealed into a tantalum tube under an argon atmosphere. In addition, the tantalum tube containing the B/Mg mixture is sealed into a quartz ampoule. To form the desired MgB_2 intermetallic compound, the mixture is heat treated at 950 °C for 2 h [50.140]. In a similar way boron precursor films can be converted to MgB_2 in the presence of Mg vapor. The annealing temperature for the formation of MgB_2 films is typically in the 850 °C range. Very recently c-axis-oriented epitaxial MgB_2 films were grown onto Al_2O_3 substrates [50.121]. For applications in electronics, it is desirable to grow the MgB_2 films in situ in a single step on a suitable substrate. The substrate and electronic materials ZrO_2, MgO, TiN, TaN, AlN, and SiC show no reaction with MgB_2 up to an annealing temperature of 800 °C [50.141].

50.6 Iron-Based Superconductors

In 2008, *Hosono* and collaborators discovered that $LaO_{1-x}F_xFeAs$ is a superconductor with a critical temperature of 26 K [50.16]. Very recently a critical temperature, as high as 58.1 K, was reported for $SmO_{0.74}F_{0.26}FeAs$ [50.142]. Because of these high critical temperatures, the iron-based superconductors are considered as a second class of complex high-temperature superconductors. There exist four main families of iron-based superconductors, which, like the cuprates, are characterized by layered crystal structures. The (1111)-family can be represented by the chemical formula REOFeAs, where RE is a rare earth element. Layers of REO and FeAs alternate along the crystallographic c direction. Figure 50.32 shows a sketch of the tetragonal crystal structure of the undoped LaOFeAs (La-1111) at room temperature.

In a second family of iron-pnictide superconductors, layers of earth alkaline ions alternate with Fe_2As_2 layers. These (122) compounds crystallize with a tetragonal crystal structure (space group I4/mmm) [50.145, 146]. Figure 50.33 shows as an example the tetragonal crystal structure of $CaFe_2As_2$ (Ca-122) [50.146].

In the (111)-family of iron-pnictide superconductors, FeAs layers alternate with layers of alkali ions (Li, Na) along the crystallographic c direction [50.147, 148].

In the cuprate superconductors, the current flows mainly along the conductive CuO_2 planes, which are separated by the insulating layers of the charge-carrier reservoirs (Fig. 50.14). The iron-pnictide superconductors are characterized by layered crystal structures similar to the cuprates. In the iron pnictides, the FeAs layers are conductive, while the layers of REO ((1111)-family), alkaline earth ions ((122)-family) and alkali ions ((111)-family) act as charge-carrier reservoirs. A further analogy to the cuprate superconductors is the fact that the nondoped parent compounds are not superconducting. However, the parent compounds of the iron-pnictide superconductors are metals showing an antiferromagnetic order, whereas the parent compounds of the cuprate superconductors are antiferromagnetic Mott insulators. Doping suppresses the antiferromagnetic ordering and the low-temperature

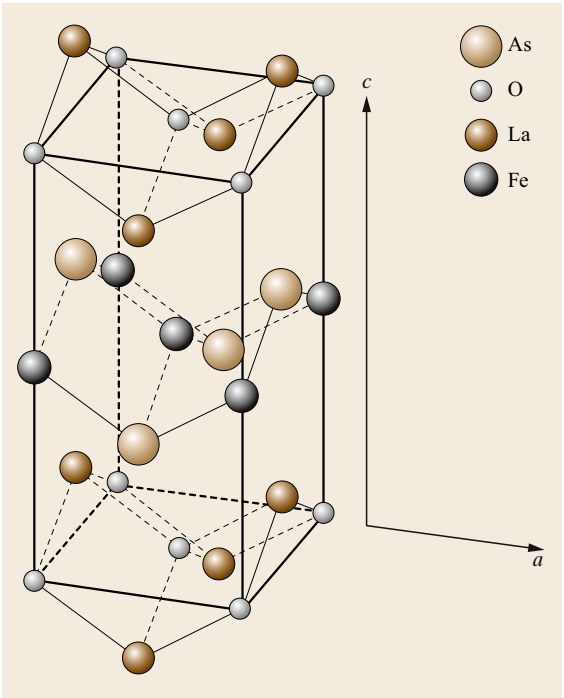


Fig. 50.32 Sketch of the tetragonal crystal structure of LaOFeAs (space group P4/nmm). The lattice parameters of the non-fluorine-doped compound are $a = 0.4033$ nm and $c = 0.8741$ nm [50.143]. The z positions of the atoms are $z = 0$ and $z = c$ for oxygen, $z = 0.141c$ and $z = 0.859c$ for La, $z = 0.5c$ for Fe and $z = 0.348c$ and $z = 0.652c$ for As [50.143]. (After [50.144, 145])

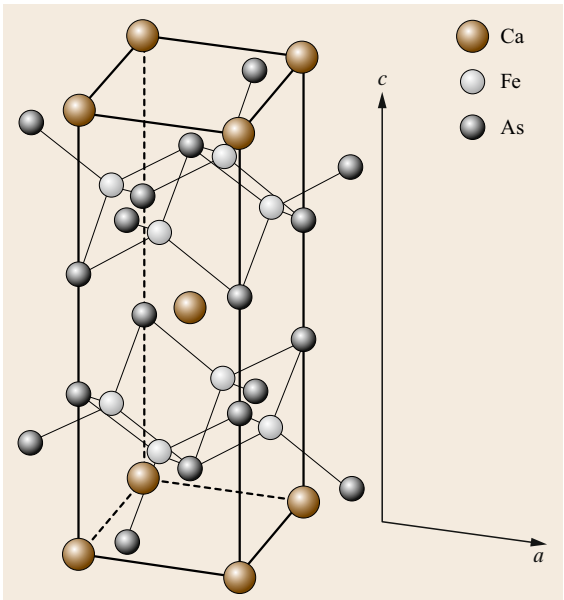


Fig. 50.33 Sketch of the tetragonal crystal structure of $CaFe_2As_2$ (space group I4/mmm). The lattice parameters of the nondoped compound are $a = 0.3872$ nm and $c = 1.1730$ nm at room temperature [50.146]. The z positions of the atoms are $z = 0$, $z = 0.5c$ and $z = c$ for Ca, $z = 0.25c$ and $z = 0.75c$ for Fe, and $z = 0.133c$, $z = 0.367c$, $0.633c$ and $z = 0.867c$ for As. (After [50.145, 146])

transition from a tetragonal to an orthorhombic crystal structure (Fig. 50.34).

The critical temperature of iron-pnictide superconductors depends on the doping level. In LaOFeAs, superconductivity can be established by fluorine doping. The charges per formula unit of LaOFeAs are +1 in the LaO layer and −1 in the FeAs layer. The charges of La, O, and F are +3, −2 and −1 respectively. Substitution of 10% of the oxygen atoms by fluorine (leading to $LaO_{0.9}F_{0.1}FeAs$) increases the charge of the LaOF layer to +1.1. For charge neutrality, the negative charge of the FeAs layer must be increased to −1.1, i. e., the FeAs layer is doped with electrons. In a similar way, replacement of Ba by K leads to hole doping in $Ba_{1-x}K_xFe_2As_2$.

In contrast to cuprate superconductors, in which doping in the CuO_2 planes is detrimental for superconductivity, cobalt doping for iron can establish superconductivity in $Ba(Fe, Co)_2As_2$ [50.148]. Furthermore, phosphorous doping leads to superconductivity in $BaFe_2(As, P)_2$ without the introduction of additional charge carriers to the FeAs planes [50.150]. In the iron-pnictide superconductors, not only the addition of charge carriers but also the effect of doping on the crystal structure are of importance.

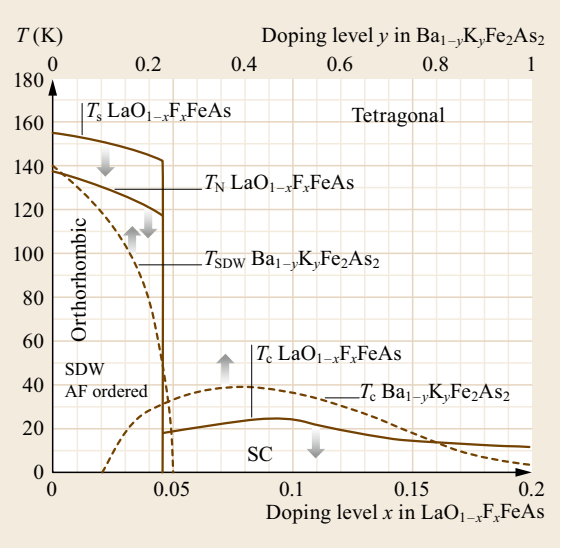


Fig. 50.34 Phase diagrams of La-1111 (*solid lines*) [50.148] and Ba-122 (*dotted lines*) [50.149]. In the case of La-1111, the T_s line indicates the temperature of the structural transformation from a tetragonal to an orthorhombic structure. At a slightly lower temperature T_N, a transition to an antiferromagnetic state occurs. Above 4.5% fluorine doping, superconductivity is observed, while the transition to an orthorhombic crystal structure and antiferromagnetic ordering is suppressed. For Ba-122, the transition to an orthorhombic crystal structure and SDW ordering occurs at the same temperature T_{SDW}. Again, doping suppresses of the structural transition and the antiferromagnetic SDW ordering. A maximum critical temperature of ≈ 38 K is reached for $y = 0.4$

In addition to the iron pnictides, there exists a fourth family of iron-based superconductors without FeAs layers. In this (11)-family with the chemical formula FeCh, where Ch represents Se, Te, or S, the FeAs layers are replaced by iron chalcogenide layers. Moreover, there are no layers to act as charge-carrier reservoirs [50.148].

The next aspect to be considered are the phase diagrams of $LaO_{1-x}F_xFeAs$ (La-1111) [50.148] and $Ba_{1-y}K_yFe_2As_2$ (Ba-122) [50.149] (Fig. 50.34). The La-1111 parent compound without fluorine doping shows a transition from a tetragonal to an orthorhombic structure at a temperature $T_s \approx 155$ K. Slightly below 140 K, antiferromagnetic ordering has been observed. Doping with fluorine suppresses both the structural instability and the antiferromagnetic order. Above 4.5% fluorine doping, superconductivity is observed, and a maximum critical temperature of 26 K is reached at around 10%. In the case of the Ba-122, the establishment of an antiferromagnetic spin density wave (SDW) order is accompanied by a structural change from a tetragonal to an orthorhombic structure. In the nondoped parent compound, the transition to the orthorhombic crystal structure and the SDW order occurs around 140 K. Doping suppresses the structural instability and the antiferromagnetic SDW order. Superconductivity can exist above 10% potassium doping. For doping levels between 10 and 25% superconductivity and SDW order can coexist in Ba-122 in the orthorhombic crystal structure. The maximum critical temperature of around 38 K is reached when around 40% of the barium atoms are replaced by potassium.

The known physical properties of various iron-based superconductors are listed in Table 50.9. The maximum critical temperatures of (1111) compounds exceed 50 K. For Ba-122 and Sr-122, maximum critical temperatures of 38 K have been reported. The (1111) compounds crystallize in a tetragonal crystal structure with the space group P4/nmm. The lattice parameters of the fluorine-doped compounds are slightly shorter than those of the parent compounds. For the compounds considered, the a lattice parameters are in the range of 0.3932 and 0.4039 nm, while the c lattice parameters are between 0.8490 and 0.8742 nm. The crystal structures of the (122) compounds are body-centered tetragonal (space group I4/mmm). The lattice parameters of the nondoped Ba-122 are $a = 0.3964$ nm and $c = 1.3022$ nm. In some of the iron-pnictide superconductors, evidence for the existence of two gaps has been found. The ratio of $2\Delta/T_c$ is between 3 and 4

for the smaller gap, and above 7 for the larger. A characteristic feature of the iron-based superconductors is an extremely high upper critical field of the order of 100 T. Because of the layered crystal structures, the physical properties of iron-based superconductors are anisotropic. The coherence length along the FeAs (*ab*) planes varies between 1.2 and 5 nm, while along the *c* direction it is in the range of 0.35 to 1.2 nm. The penetration depth for screening currents flowing in the *ab* planes is of the order of 200 nm.

The last aspect to be considered is the critical current density in iron-based superconducting films. The critical current densities of Fe(Te,Se) [50.177], $Ba(Fe,Co)_2As_2$ [50.178], $BaFe_2(As,P)_2$ [50.179], and IBAD-MgO/$Ba(Fe,Co)_2As_2$ films [50.180] are presented in Fig. 50.35. The critical current densities at 4.2 K and zero applied field reach values between 0.4 (Fe(Te,Se)) and 3.5 MA/cm^2 ($BaFe_2(As,P)_2$). For the $Ba(Fe,Co)_2As_2$ film, critical current densities for fields applied parallel and perpendicular to the crystallographic *c*-direction are shown. The critical current densities at 4.2 K and an applied magnetic field of 9 T parallel and perpendicular to the *c*-direction are 0.12 and 0.2 MA/cm^2, respectively. This anisotropy with respect to the direction of the applied magnetic field is less pronounced than in conductors coated with $REBa_2Cu_3O_7$ (RE: rare earth element). The j_c data of the $BaFe_2(As,P)_2$ film at temperatures of 4.2, 10, and 20 K indicate that the dependence on the applied magnetic field at 20 K is much more pronounced than at 10 or 4.2 K. The high critical current densities and the weak field dependence at low temperatures suggest that the iron-based superconductors show promise in the generation of extremely high magnetic fields. Because of the very short coherence length along the crystallographic *c*-direction, they may show a weak-link behavior of the grain boundaries similar to the cuprates. Several research groups [50.170, 181–186] have studied the intergrain and the intragrain critical current densities in iron-based superconductors. Evidence for intergrain critical current densities much smaller than the intragrain j_c has been found. However, the large difference between the two is caused partly by the wetting of the grain boundaries with FeAs. The extent to which the observed reduction of the intergrain j_c is caused by an intrinsic weak-link problem needs to be clarified.

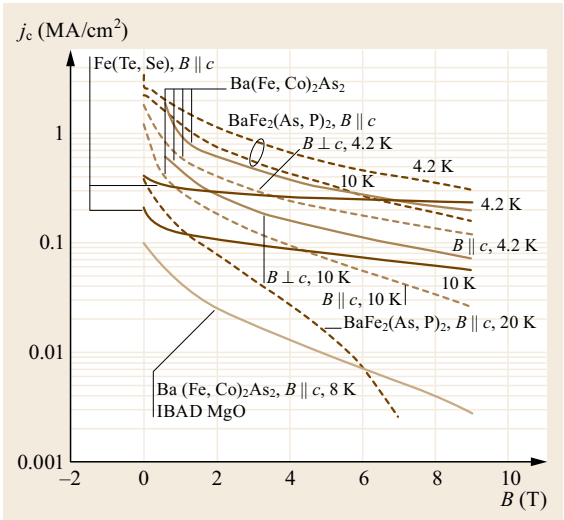


Fig. 50.35 Critical current density versus field in Fe(Te,Se) [50.177], $Ba(Fe,Co)_2As_2$ [50.178], $BaFe_2(As,P)_2$ [50.179], and IBAD-MgO/$Ba(Fe,Co)_2As_2$ films [50.180]

50.7 Summary

The cuprate high-temperature superconductors are characterized by layered crystal structures. Conductive CuO_2 layers in the crystallographic *ab*-planes are a common feature of all cuprate superconductors. Along the *c*-direction, conducting copper oxide blocks and insulating charge-carrier reservoirs alternate. As a consequence, the physical properties of cuprate superconductors are strongly anisotropic.

The critical temperature as well as the other physical properties depends considerably on the charge-carrier concentration. The maximum critical temperature is typically reached for a carrier concentration of 0.18 holes per CuO_2. Furthermore, the highest T_c of optimally doped cuprate superconductors is typically reached in compounds with three or four CuO_2 layers in the CuO_2 blocks. The highest known critical temperature of 135 K at ambient pressure was achieved in Hg-1223.

The cuprate superconductors are close to a metal–insulator transition. The d-wave symmetry of the energy gap and the existence of a pseudogap above T_c, especially in under-doped cuprate superconductors, suggest that the pairing mechanism of the electrons is different from that in metallic superconductors. Furthermore, only a weak isotope effect has been observed in the cuprates, which again suggests that the electron pairing is not phonon mediated.

Extremely short coherence lengths and low carrier concentrations seem to be responsible for the weak-link

Table 50.9 Physical properties of various iron-based superconductors

Compound	Parameter	Value	Comments	Reference
Critical temperatures				
LaOFeP	T_c (K)	≈ 4	P analog to LaOFeAs	[50.30]
$La_{0.89}F_{0.11}FeAs$	T_c (K)	26	O partly replaced by F	[50.16]
$La_{0.87}Sr_{0.13}OFeAs$	T_c (K)	25	La partly replaced by Sr	[50.151]
$SmO_{0.74}F_{0.26}FeAs$	T_c (K)	58.1	O partly replaced by F	[50.34]
$SmO(Fe_{0.9}Co_{0.1})As$	T_c (K)	15.2	Fe partly replaced by Co	[50.152]
$CeO_{0.84}F_{0.16}FeAs$	T_c (K)	41	O partly replaced by F	[50.153]
$NdO_{0.75}F_{0.25}FeAs$	T_c (K)	55	O partly replaced by F	[50.154]
$NdO_{0.85}FeAs$	T_c (K)	46	Oxygen deficiency	[50.155]
$GdO_{0.85}FeAs$	T_c (K)	53.5	Oxygen deficiency	[50.156]
$Gd_{0.8}Th_{0.2}OFeAs$	T_c (K)	56	Gd partly replaced by Th	[50.144]
$DyO_{0.9}F_{0.1}FeAs$	T_c (K)	45	O partly replaced by F	[50.157]
LiFeAs	T_c (K)	≈ 18		[50.147]
NaFeAs	T_c (K)	12–25		[50.147]
$Ba_{0.6}K_{0.4}Fe_2As_2$	T_c (K)	38	Ba partly replaced by K	[50.148]
$Sr_{0.6}K_{0.4}Fe_2As_2$	T_c (K)	38	Sr partly replaced by K	[50.148]
$Ba(Fe_{0.8}Co_{0.2})_2As_2$	T_c (K)	22	Co doping in FeAs layers	[50.148]
$\alpha - FeSe$	T_c (K)	8–27		[50.148]
$Fe(Se_{0.5}Te_{0.5})$	T_c (K)	≈ 15.5	Se partly replaced by Te	[50.148]
$Fe(Se_{0.8}S_{0.2})$	T_c (K)	≈ 15.5	Se partly replaced by S	[50.147]
$Fe(Te_{0.8}S_{0.2})$	T_c (K)	10	Te partly replaced by S	[50.147]
Crystal structures and lattice parameters				
LaOFeAs	a (nm)	0.4039	Tetragonal (P4/nmm)	[50.158]
	c (nm)	0.8742		
$LaO_{0.8}F_{0.2}FeAs$	a (nm)	0.4030	Tetragonal (P4/nmm)	[50.158]
	c (nm)	0.8716		
SmOFeAs	a (nm)	0.3940	Tetragonal (P4/nmm)	[50.159]
	c (nm)	0.8501		
$SmO_{0.85}F_{0.15}FeAs$	a (nm)	0.3932	Tetragonal (P4/nmm)	[50.159]
	c (nm)	0.8490		
$BaFe_2As_2$	a (nm)	0.3964	Tetragonal (I4/mmm)	[50.160]
	c (nm)	1.3022		
$BaFe_{1.87}Co_{0.13}As_2$	a (nm)	0.3963	Tetragonal (I4/mmm)	[50.160]
	c (nm)	1.2991		
LiFeAs	a (nm)	0.3791	Tetragonal (P4/nmm)	[50.161]
	c (nm)	0.6364		
Superconducting energy gaps				
$LaO_{0.9}F_{0.1-\delta}FeAs$	Δ (meV)	3.9 ± 0.7	$T_c = 27\,K$, $2\Delta/T_c = 3.35$	[50.162]
$SmO_{0.8}F_{0.2}FeAs$	Δ (meV)	6.45 ± 0.25	$T_c = 51.2\,K$, $2\Delta/T_c = 2.95$	[50.163]
(2 gaps)	Δ (meV)	16.6 ± 1.6	$T_c = 51.2\,K$, $2\Delta/T_c = 7.65$	
$Ba_{0.6}K_{0.4}Fe_2As_2$	Δ (meV)	6	$T_c = 37\,K$, $2\Delta/T_c = 3.76$	[50.164]
(2 gaps)	Δ (meV)	12	$T_c = 37\,K$, $2\Delta/T_c = 7.53$	
LiFeAs (single gap)	Δ (meV)	3 ± 0.2	$T_c = 17\,K$, $2\Delta/T_c = 4.1$	[50.165]
Upper critical fields				
$LaO_{0.9}F_{0.1-\delta}FeAs$	B_{c2} (T)	56	$T = 0$, $T_c = 28.9\,K$	[50.166]
$SmO_{0.65}F_{0.35}FeAs$	B_{c2} (T)	≈ 120	$T = 0$, $T_c = 52\,K$	[50.167]
$CeO_{0.9}F_{0.1}FeAs$	B_{c2} (T)	94	$T = 0$, $T_c = 38.4\,K$	[50.168]
$Ba_{0.6}K_{0.4}Fe_2As_2$	B_{c2} (T)	75	$B_{c2}^{ab}/B_{c2}^{c} \approx 1-3$	[50.149]
$(Ba, K)Fe_2As_2$	B_{c2} (T)	70	$B_{c2}^{ab}/B_{c2}^{c} \approx 1.4$, $T_c = 28\,K$	[50.169]
Fe(Te, Se)	B_{c2} (T)	87	$T = 0$	[50.148]

Table 50.9 (continued)

Compound	Parameter	Value	Comments	Reference
Characteristic length scales				
$LaO_{0.89}F_{0.11}FeAs$	$\xi_{ab}(0)$ (nm)	5	Coherence length	[50.170]
	$\xi_c(0)$ (nm)	1.2		
$LaO_{0.9}F_{0.1}FeAs$	$\lambda_{ab}(0)$ (nm)	254	Penetration depth	[50.171]
$NdO_{0.82}F_{0.18}FeAs$	$\xi_{ab}(0)$ (nm)	3.1	Coherence length	[50.172]
	$\xi_c(0)$ (nm)	0.78		
$SmO_{0.9}F_{0.1}FeAs$	$\lambda_{ab}(0)$ (nm)	205	Penetration depth	[50.173]
$Ba_{0.6}K_{0.4}Fe_2As_2$	$\xi_{ab}(0)$ (nm)	1.25	Coherence length	[50.172]
	$\xi_c(0)$ (nm)	0.45		
	$\lambda_{ab}(0)$ (nm)	125	Penetration depth	
	$\lambda_c(0)$ (nm)	304		
$SrFe_{1.75}Co_{0.25}As_2$	$\lambda_{ab}(0)$ (nm)	316	Penetration depth	[50.174]
	$\lambda_c(0)$ (nm)	870		
LiFeAs	$\xi_{ab}(0)$ (nm)	2–4	Coherence length	[50.165]
	$\lambda_{ab}(0)$ (nm)	210	Penetration depth	
$FeSe_{0.89}$	$\xi(0)$ (nm)	3.14	Coherence length	[50.175]
	$\lambda(0)$ (nm)	1008	Penetration depth	
Fe(Se, Te)	$\xi_{ab}(0)$ (nm)	1.2	Coherence length	[50.176]
	$\xi_c(0)$ (nm)	0.35		

behavior of large-angle grain boundaries in the cuprate superconductors. To achieve transport critical current densities comparable to the intra-grain j_c, a biaxial texture is required.

Since the discovery of superconductivity above 50 K in the iron pnictides, the cuprates are no longer the only class of high-temperature superconductors. In the iron pnictides, layers of FeAs alternate with layers of REO, alkaline earth or alkali ions, which act as charge-carrier reservoirs resembling those in the cuprate superconductors. As in the cuprates, the superconducting state is close to an antiferromagnetic state. However, the parent compounds of the iron-pnictide superconductors are metals and not Mott insulators as in the case of the cuprates. Because of their very high upper critical fields, the iron-based superconductors are potential high field superconductors.

The metallic intermediate-temperature superconductor MgB_2 may have applications in spite of its moderate critical temperature of 39 K, because of its simpler chemistry and the fact that its grain boundaries do not act as weak links. The observation of a pronounced boron isotope effect suggests that MgB_2 is a conventional BCS superconductor. The anisotropy of the physical properties of MgB_2 is less pronounced than that of the cuprate superconductors.

The maximum operation fields of MgB_2, cuprate and iron-based superconductors are limited by the irreversibility line, which is well below the upper critical field. The existence of an irreversibility line may be caused by thermally activated depinning or flux-line lattice melting.

Acknowledgments. The author wishes to thank P. Bruzzone for his support and encouragement. The careful reviewing and many constructive suggestions on the manuscript by J. F. Crawford are especially appreciated.

References

50.1 R. de Bruyn Ouboter: IEEE Trans. Mag. **23**, 355 (1987)
50.2 W. Meissner, R. Ochsenfeld: Naturwissenschaften **21**, 787 (1933)
50.3 F. London, H. London: Physica **2**, 341 (1935)
50.4 J. Bardeen, L.N. Cooper, J.R. Schrieffer: Phys. Rev. **106**, 162 (1957)
50.5 J. Bardeen, L.N. Cooper, J.R. Schrieffer: Phys. Rev. **108**, 1175 (1957)
50.6 I. Giaever: Phys. Rev. Lett. **5**, 147 (1960)
50.7 I. Giaever: Phys. Rev. Lett. **5**, 464 (1960)
50.8 R. Doll, M. Näbauer: Phys. Rev. Lett. **7**, 51 (1961)
50.9 B.S. Deaver Jr., W.M. Fairbank: Phys. Rev. Lett. **7**, 43 (1961)
50.10 B.D. Josephson: Phys. Lett. **1**, 251 (1962)
50.11 R.C. Jaklevic, J. Lambe, J.E. Mercereau, A.H. Silver: Phys. Rev. A **140**, 1628 (1965)
50.12 J.G. Bednorz, K.A. Müller: Z. Phys. B **64**, 189 (1986)
50.13 A. Schilling, M. Cantoni, J.D. Guo, H.R. Ott: Nature **363**, 56 (1993)

50.14 J.J. Capponi, J.L. Tholence, C. Chaillout, M. Marezio, P. Bordet, J. Chenavas, S.M. Loureiro, E.V. Antipov, E. Kopnine, M.F. Gorius, M. Nunez-Regueiro, B. Souletie, P. Radaelli, F. Gerhards: Physica C **235–240**, 146 (1994)

50.15 J. Nagamatsu, N. Nakagawa, T. Muranaka, Y. Zenitani, J. Akimitsu: Nature **410**, 63 (2001)

50.16 Y. Kamihara, T. Watanabe, M. Hirano, H. Hosono: J. Am. Chem. Soc. **130**, 3296 (2008)

50.17 L.R. Testardi, J.H. Wernick, W.A. Royer: Solid State Commun. **15**, 1 (1974)

50.18 J.R. Gavaler, M.A. Janocko, C.K. Jones: J. Appl. Phys. **45**, 3009 (1974)

50.19 J.F. Schooley, W.R. Hosler, M.L. Cohen: Phys. Rev. Lett. **12**, 474 (1964)

50.20 A.W. Sleight, J.L. Gillson, P.E. Bierstedt: Solid State Commun. **17**, 27 (1975)

50.21 M.K. Wu, J.R. Ashburn, C.J. Torng, P.H. Hor, R.L. Meng, L. Gao, Z.J. Huang, Y.Q. Wang, C.W. Chu: Phys. Rev. Lett. **58**, 908 (1987)

50.22 D. Jérome, A. Mazaud, M. Ribault, K. Bechgaard: J. Phys. Lett. **41**, L95 (1980)

50.23 K. Oshima, H. Urayama, H. Yamochi, G. Saito: Physica C **153–155**, 1148 (1988)

50.24 J.M. Williams, A.J. Schultz, U. Geiser, K.D. Carlson, A.M. Kini, H.H. Wang, W.-K. Kwok, M.-H. Whangbo, J.E. Schirber: Science **252**, 1501 (1991)

50.25 A.F. Hebard, M.J. Rosseinsky, R.C. Haddon, D.W. Murphy, S.H. Glarum, T.T.M. Palstra, A.P. Ramirez, A.R. Kortan: Nature **350**, 600 (1991)

50.26 K. Lüders: Phys. Bl. **50**, 166 (1994)

50.27 T.T.M. Palstra, O. Zhou, Y. Iwasa, P.E. Sulewski, R.M. Fleming, B.R. Zegarski: Solid State Commun. **93**, 327 (1995)

50.28 Y. Kamihara, H. Hiramatsu, M. Hirano, R. Kawamura, H. Yanagi, T. Kamiya, H. Hosono: J. Am. Chem. Soc. **128**, 10012 (2006)

50.29 K.-W. Yeh, T.-W. Huang, Y.-L. Huang, T.K. Chen, F.-C. Hsu, P.M. Wu, Y.-C. Lee, Y.-Y. Chu, C.-L. Chen, J.-Y. Luo, D.-C. Yan, M.K. Wu: Europhys. Lett. **84**, 37002 (2008)

50.30 M. Rotter, M. Tegel, D. Johrendt: Phys. Rev. Lett. **101**, 107006 (2008)

50.31 Z.-A. Ren, G.-C. Che, X.-L. Dong, J. Yang, W. Lu, W. Yi, X.-L. Shen, Z.-C. Li, L.-L. Sun, F. Zhou, Z.-X. Zhao: Europhys. Lett. **83**, 17002 (2008)

50.32 G. Wu, Y.L. Xie, H. Chen, M. Zhong, R.H. Liu, B.C. Shi, Q.J. Li, X.F. Wang, T. Wu, Y.J. Yan, J.J. Ying, X.H. Chen: J. Phys. Condens. Matter **21**, 142203 (2009)

50.33 C. Wang, Z. Gao, L. Wang, Y. Qi, D. Wang, C. Yao, Z. Zhang, Y. Ma: Supercond. Sci. Technol. **23**, 55002 (2010)

50.34 M. Fujioka, S.J. Denholme, T. Ozaki, H. Okazaki, K. Deguchi, S. Demura, H. Hara, T. Watanabe, H. Takeya, T. Yamaguchi, H. Kumakura, Y. Takano: Supercond. Sci. Technol. **26**, 085023 (2013)

50.35 T.R. Strobridge: IEEE Trans. Nucl. Sci. **16**, 1104 (1969)

50.36 W.-J. Yeh, L. Chen, F. Xu, B. Bi, P. Yang: Phys. Rev. B **36**, 2414 (1987)

50.37 F.J. Kedves, S. Mészáros, K. Vad, G. Halász, B. Keszei, L. Mihály: Solid State Commun. **63**, 991 (1987)

50.38 D.J. Quinn, W.B. Ittner: J. Appl. Phys. **33**, 748 (1962)

50.39 R. Wesche: *High-Temperature Superconductors: Materials, Properties, and Applications* (Kluwer Academic, Boston 1998)

50.40 H. Fröhlich: Phys. Rev. **79**, 845 (1950)

50.41 J. Bardeen: Phys. Rev. **80**, 567 (1950)

50.42 C.P. Poole, H.A. Farach, R.J. Creswick: *Superconductivity* (Academic, San Diego 1995)

50.43 C.A. Reynolds, B. Serin, W.H. Wright, L.B. Nesbitt: Phys. Rev. **78**, 487 (1950)

50.44 E. Maxwell: Phys. Rev. **78**, 477 (1950)

50.45 B. Serin, C.A. Reynolds, C. Lohman: Phys. Rev. **86**, 162 (1952)

50.46 J.W. Garland: Phys. Rev. Lett. **11**, 114 (1963)

50.47 W. Buckel: *Supraleitung* (VCH, Weinheim 1993)

50.48 T.P. Orlando, K.A. Delin: *Foundations of Applied Superconductivity* (Addison–Wesley, Reading 1991)

50.49 G. Sparn, J.D. Thompson, R.L. Whetten, S.-M. Huang, R.B. Kaner, F. Diederich, G. Grüner, K. Holczer: Phys. Rev. Lett. **68**, 1228 (1992)

50.50 C. Renner, I. Maggio-Aprile, Ø. Fischer: Phys. Bl. **54**, 427 (1998)

50.51 C. Renner, B. Revaz, J.-Y. Genoud, K. Kadowaki, Ø. Fischer: Phys. Rev. Lett. **80**, 149 (1998)

50.52 Z.A. Xu, N.P. Ong, Y. Wang, T. Takeshita, S. Uchida: Nature **406**, 486 (2000)

50.53 N.E. Hussey: Normal state transport properties. In: *Handbook of High-Temperature Superconductivity: Theory and Experiment*, ed. by J.R. Schrieffer, J.S. Brooks (Springer, New York 2007)

50.54 J.-H. Ma, Z.-H. Pan, F.C. Niestemski, M. Neupane, Y.-M. Xu, P. Richard, K. Nakayama, T. Sato, T. Takahashi, H.-Q. Luo, L. Fang, H.-H. Wen, Z. Wang, H. Ding, V. Madhavan: Phys. Rev. Lett. **101**, 207002 (2008)

50.55 J. Orenstein, A.J. Millis: Science **288**, 468 (2000)

50.56 J.M. Tranquada, B.J. Sternlieb, J.D. Axe, Y. Nakamura, S. Uchida: Nature **375**, 561 (1995)

50.57 R. Beyers, G. Lim, E.M. Engler, R.J. Savoy, T.M. Shaw, T.R. Dinger, W.J. Gallagher, R.L. Sandstrom: Appl. Phys. Lett. **50**, 1918 (1987)

50.58 R.J. Cava, B. Batlogg, R.B. van Dover, D.W. Murphy, S. Sunshine, T. Siegrist, J.P. Remeika, E.A. Rietman, S. Zahurak, G.P. Espinosa: Phys. Rev. Lett. **58**, 1676 (1987)

50.59 A. Fukuoka, A. Tokiwa-Yamamoto, M. Itoh, R. Usami, S. Adachi, H. Yamauchi, K. Tanabe: Physica C **265**, 13 (1996)

50.60 T. Schweizer, R. Müller, L.J. Gauckler: Physica C **225**, 143 (1994)

50.61 R. Hott, H. Rietschel, M. Sander: Phys. Bl. **48**, 355 (1992)

50.62 A. Maignan, C. Martin, M. Huve, J. Provost, M. Hervieu, C. Michel, B. Raveau: Physica C **170**, 350 (1990)

50.63 A. Maignan, C. Martin, V. Hardy, C. Simon, M. Hervieu, B. Raveau: Physica C **219**, 407 (1994)

50.64 J.J. Capponi, E.M. Kopnin, S.M. Loureiro, E.V. Antipov, E. Gautier, C. Chaillout, B. Souletie, M. Brunner, J.L. Tholence, M. Marezio: Physica C

256, 1 (1996)
50.65 L. Gao, Y.Y. Xue, F. Chen, Q. Xiong, R.L. Meng, D. Ramirez, C.W. Chu, J.H. Eggert, H.K. Mao: Phys. Rev. B **50**, 4260 (1994)
50.66 N. Watanabe, K. Fukamachi, Y. Ueda, K. Tsushima, A.M. Balbashov, T. Nakanishi, N. Môri: Physica C **235–240**, 1309 (1994)
50.67 J.-P. Locquet, J. Perret, J. Fompeyrine, E. Mächler, J.W. Seo, G. van Tendeloo: Nature **394**, 453 (1998)
50.68 M. Tinkham: Physica C **235–240**, 3 (1994)
50.69 A.J. Panson, G.R. Wagner, A.I. Braginski, J.R. Gavaler, M.A. Janocko, H.C. Pohl, J. Talvacchio: Appl. Phys. Lett. **50**, 1104 (1987)
50.70 D.W. Capone, D.G. Hinks, J.D. Jorgensen, K. Zhang: Appl. Phys. Lett. **50**, 543 (1987)
50.71 U. Welp, W.K. Kwok, G.W. Crabtree, K.G. Vandervoort, J.Z. Liu: Phys. Rev. Lett. **62**, 1908 (1989)
50.72 T.K. Worthington, W.J. Gallagher, T.R. Dinger: Phys. Rev. Lett. **59**, 1160 (1987)
50.73 C. Heinzel, C. Neumann, P. Ziemann: Europhys. Lett. **13**, 531 (1990)
50.74 D.-H. Wu, S. Sridhar: Phys. Rev. Lett. **65**, 2074 (1990)
50.75 M. Akamatsu, L.X. Chen, H. Ikeda, R. Yoshizaki: Physica C **235–240**, 1619 (1994)
50.76 D.R. Harshman, A.P. Mills: Phys. Rev. B **45**, 10684 (1992)
50.77 J.H. Kang, R.T. Kampwirth, K.E. Gray: Appl. Phys. Lett. **52**, 2080 (1988)
50.78 J.N. Li, K. Kadowaki, M.J.V. Menken, Y.K. Huang, K. Bakker, A.A. Menovsky, J.J.M. Franse: Appl. Phys. A **47**, 209 (1988)
50.79 L. Zhang, J.Z. Liu, R.N. Shelton: Phys. Rev. B **45**, 4978 (1992)
50.80 I. Matsubara, H. Tanigawa, T. Ogura, H. Yamashita, M. Kinoshita, T. Kawai: Phys. Rev. B **45**, 7414 (1992)
50.81 F. Shi, T.S. Rong, S.Z. Zhou, X.F. Wu, J. Du, Z.H. Shi, C.G. Cui, R.Y. Jin, J.L. Zhang, Q.Z. Ran, N.C. Shi: Phys. Rev. B **41**, 6541 (1990)
50.82 I. Matsubara, R. Funahashi, K. Ueno, H. Yamashita, T. Kawai: Physica C **256**, 33 (1996)
50.83 M. Hasegawa, K. Izawa, A. Shibata, Y. Matsuda: Physica C **377**, 459 (2002)
50.84 G. Triscone, A. Junod, R.E. Gladyshevskii: Physica C **264**, 233 (1996)
50.85 J. Hofer, J. Karpinski, M. Willemin, G.I. Meijer, E.M. Kopnin, R. Molinski, H. Schwer, C. Rossel, H. Keller: Physica C **297**, 103 (1998)
50.86 R. Usami, A. Fukuoka, H. Kubota, H. Yamauchi: Physica C **243**, 19 (1995)
50.87 G. Villard, A. Daignere, D. Pelloquin, A. Maignan: Physica C **314**, 196 (1999)
50.88 V. Vulcanescu, L. Fruchter, A. Bertinotti, D. Colson, G. Le Bras, J.-F. Marucco: Physica C **259**, 131 (1996)
50.89 T.T.M. Palstra, B. Batlogg, L.F. Schneemayer, J.V. Waszczak: Phys. Rev. Lett. **61**, 1662 (1988)
50.90 H. Kumakura, K. Togano, N. Tomita, E. Yanagisawa, S. Okayasu, Y. Kazumatu: Physica C **251**, 231 (1995)
50.91 V. Hardy, J. Provost, D. Groult, M. Hervieu, B. Raveau, S. Durčok, E. Pollert, J.C. Frison, J.P. Chaminade, M. Pouchard: Physica C **191**, 85 (1992)
50.92 L. Civale, A.D. Marwick, R. Wheeler, M.A. Kirk, W.L. Carter, G.N. Riley, A.P. Malozemoff: Physica C **208**, 137 (1993)
50.93 A. Schilling, R.A. Fisher, N.E. Phillips, U. Welp, D. Dasgupta, W.K. Kwok, G.W. Crabtree: Nature **382**, 791 (1996)
50.94 A. Junod, M. Roulin, J.-Y. Genoud, B. Revaz, A. Erb, E. Walker: Physica C **275**, 245 (1997)
50.95 M. Rupp, A. Gupta, C.C. Tsuei: Appl. Phys. Lett. **67**, 291 (1995)
50.96 L. Krusin-Elbaum, C.C. Tsuei, A. Gupta: Nature **373**, 679 (1995)
50.97 P. Schmitt, L. Schultz, G. Saemann-Ischenko: Physica C **168**, 475 (1990)
50.98 D. Dimos, P. Chaudhari, J. Mannhart, F.K. LeGoues: Phys. Rev. Lett. **61**, 219 (1988)
50.99 D. Dimos, P. Chaudhari, J. Mannhart: Phys. Rev. B **41**, 4038 (1990)
50.100 A. Schmehl, B. Goetz, R.R. Schulz, C.W. Schneider, H. Bielefeldt, H. Hilgenkamp, J. Mannhart: Europhys. Lett. **47**, 110 (1999)
50.101 A. Attenberger, J. Hänisch, B. Holzapfel, L. Schultz: Physica C **372–376**, 649 (2002)
50.102 J. Mannhart, H. Hilgenkamp: Physica C **317–318**, 383 (1999)
50.103 H. Hilgenkamp, J. Mannhart: Appl. Phys. Lett. **73**, 265 (1998)
50.104 R.A. Rao, Q. Gan, C.B. Eom, Y. Suzuki, A.A. McDaniel, J.W.P. Hsu: Appl. Phys. Lett. **69**, 3911 (1996)
50.105 H. Yamasaki, K. Endo, S. Kosaka, M. Umeda, S. Misawa, S. Yoshida, K. Kajimura: IEEE Trans. Appl. Supercond. **3**, 1536 (1993)
50.106 S.H. Yun, J.Z. Wu: Appl. Phys. Lett. **68**, 862 (1996)
50.107 D. Kumar, M. Sharon, R. Pinto, P.R. Apte, S.P. Pai, S.C. Purandare, L.C. Gupta, R. Vijayaraghavan: Appl. Phys. Lett. **62**, 3522 (1993)
50.108 T. Terashima, K. Shimura, Y. Bando, Y. Matsuda, A. Fujiyama, S. Komiyama: Phys. Rev. Lett. **67**, 1362 (1991)
50.109 J.-M. Triscone, Ø. Fischer, O. Brunner, L. Antognazza, A.D. Kent, M.G. Karkut: Phys. Rev. Lett. **64**, 804 (1990)
50.110 W. Si, H.-C. Li, X.X. Xi: Appl. Phys. Lett. **74**, 2839 (1999)
50.111 T. Sekitani, H. Sato, M. Naito, N. Miura: Physica C **378–381**, 195 (2002)
50.112 M.R. Eskildsen, M. Kugler, G. Levy, S. Tanaka, J. Jun, S.M. Kazakov, J. Karpinski, Ø. Fischer: Physica C **385**, 169 (2003)
50.113 R.S. Gonnelli, D. Daghero, G.A. Ummarino, V.A. Stepanov, J. Jun, S.M. Kazakov, J. Karpinski: Phys. Rev. Lett. **89**, 247004 (2002)
50.114 Z.-Z. Li, H.-J. Tao, Y. Xuan, Z.-A. Ren, G.-C. Che, B.-R. Zhao: Phys. Rev. B **66**, 064513 (2002)
50.115 S. Lee, Z.G. Khim, Y. Chong, S.H. Moon, H.N. Lee, H.G. Kim, B. Oh, E.J. Choi: Physica C **377**, 202 (2002)
50.116 P. Szabó, P. Samuely, J. Kačmarčik, T. Klein, J. Marcus, D. Fruchart, S. Miraglia, C. Marcenat, A.G.M. Jansen: Phys. Rev. Lett. **87**, 137005 (2001)
50.117 S.L. Bud'ko, G. Lapertot, C. Petrovic, C.E. Cunningham, N. Anderson, P.C. Canfield: Phys. Rev. Lett.

86, 1877 (2001)
50.118 D.G. Hinks, H. Claus, J.D. Jorgensen: Nature **411**, 457 (2001)
50.119 A. Gümbel, J. Eckert, G. Fuchs, K. Nenkov, K.-H. Müller, L. Schultz: Appl. Phys. Lett. **80**, 2725 (2002)
50.120 V. Braccini, L.D. Cooley, S. Patnaik, D.C. Larbalestier, P. Manfrinetti, A. Palenzona, A.S. Siri: Appl. Phys. Lett. **81**, 4577 (2002)
50.121 S.D. Bu, D.M. Kim, J.H. Choi, J. Giencke, E.E. Hellstrom, D.C. Larbalestier, S. Patnaik, L. Cooley, C.B. Eom, J. Lettieri, D.G. Schlom, W. Tian, X.Q. Pan: Appl. Phys. Lett. **81**, 1851 (2002)
50.122 T. Masui, S. Lee, A. Yamamoto, S. Tajima: Physica C **378–381**, 216 (2002)
50.123 Y. Wang, T. Plackowski, A. Junod: Physica C **355**, 179 (2001)
50.124 Z.X. Shi, A.K. Pradhan, M. Tokunaga, K. Yamazaki, T. Tamegai, Y. Takano, K. Togano, H. Kito, H. Ihara: Physica C **378–381**, 550 (2002)
50.125 M. Zehetmayer, M. Eisterer, J. Jun, S.M. Kazakov, J. Karpinski, A. Wisniewski, H.W. Weber: Phys. Rev. B **66**, 052505 (2002)
50.126 C. Ferdeghini, V. Ferrando, G. Grassano, W. Ramadan, V. Braccini, M. Putti, P. Manfrinetti, A. Palenzona: Physica C **372–376**, 1270 (2002)
50.127 Y. Ma, X. Zhang, S. Awaji, D. Wang, Z. Gao, G. Nishijima, K. Watanabe: Supercond. Sci. Technol. **20**, L5 (2007)
50.128 A. Matsumoto, H. Kitaguchi, H. Kumakura: Supercond. Sci. Technol. **21**, 065007 (2008)
50.129 Y. Eltsev, S. Lee, K. Nakao, N. Chikumoto, S. Tajima, N. Koshizuka, M. Murakami: Physica C **378–381**, 61 (2002)
50.130 F. Simon, A. Jánossy, T. Fehér, F. Murányi, S. Garaj, L. Forró, C. Petrovic, S.L. Bud'ko, G. Lapertot, V.G. Kogan, P.C. Canfield: Phys. Rev. Lett. **87**, 047002 (2001)
50.131 T. Takasaki, T. Ekino, T. Muranaka, H. Fujii, J. Akimitsu: Physica C **378–381**, 229 (2002)
50.132 F. Bouquet, R.A. Fisher, N.E. Phillips, D.G. Hinks, J.D. Jorgensen: Phys. Rev. Lett. **87**, 047001 (2001)
50.133 H.-J. Kim, W.N. Kang, E.-M. Choi, M.-S. Kim, K.H.P. Kim, S.-I. Lee: Phys. Rev. Lett. **87**, 087002 (2001)
50.134 J.M. An, W.E. Pickett: Phys. Rev. Lett. **86**, 4366 (2001)
50.135 H.J. Choi, D. Roundy, H. Sun, M.L. Cohen, S.G. Louie: Nature **418**, 758 (2002)
50.136 I.I. Mazin, V.P. Antropov: Physica C **385**, 49 (2003)
50.137 D.C. Larbalestier, L.D. Cooley, M.O. Rikel, A.A. Polyanskii, J. Jiang, S. Patnaik, X.Y. Cai, D.M. Feldmann, A. Gurevich, A.A. Squitieri, M.T. Naus, C.B. Eom, E.E. Hellstrom, R.J. Cava, K.A. Regan, N. Rogado, M.A. Hayward, T. He, J.S. Slusky, P. Khalifah, K. Inumaru, M. Haas: Nature **410**, 186 (2001)
50.138 K.H.P. Kim, W.N. Kang, M.-S. Kim, C.U. Jung, H.-J. Kim, E.-M. Choi, M.-S. Park, S.-I. Lee: Physica C **370**, 13 (2002)
50.139 W.N. Kang, H.-J. Kim, E.-M. Choi, K.H.P. Kim, S.-I. Lee: Physica C **378–381**, 1246 (2002)
50.140 S.L. Bud'ko, P.C. Canfield, V.G. Kogan: Physica C **382**, 85 (2002)
50.141 T. He, R.J. Cava, J.M. Rowell: Appl. Phys. Lett. **80**, 291 (2002)
50.142 M. Fujioka, S.J. Denholme, T. Ozaki, H. Okazaki, K. Deguchi, S. Demura, H. Hara, T. Watanabe, H. Takeya, T. Yamaguchi, H. Kumakura, Y. Takano: Supercond. Sci. Technol. **26**, 085023 (2013)
50.143 T. Nomura, S.W. Kim, Y. Kamihara, M. Hirano, P.V. Sushko, K. Kato, M. Takata, A.L. Shluger, H. Hosono: Supercond. Sci. Technol. **21**, 125028 (2008)
50.144 C. Wang, L. Li, S. Chi, Z. Zhu, Z. Ren, Y. Li, Y. Wang, X. Lin, Y. Luo, S. Jiang, X. Xu, G. Cao, Z. Xu: Europhys. Lett. **83**, 67006 (2008)
50.145 T. Yildirim: Physica C **469**, 425 (2009)
50.146 G. Wu, H. Chen, T. Wu, Y.L. Xie, Y.J. Yan, R.H. Liu, X.F. Wang, J.J. Ying, X.H. Chen: J. Phys. Condens. Matter **20**, 422201 (2008)
50.147 C.W. Chu, F. Chen, M. Gooch, A.M. Guloy, B. Lorenz, B. Lv, K. Sasmal, Z.J. Tang, J.H. Tapp, Y.Y. Zue: Physica C **469**, 326 (2009)
50.148 P.M. Aswathy, J.B. Anooja, P.M. Sarun, U. Syamaprasad: Supercond. Sci. Technol. **23**, 073001 (2010)
50.149 D. Johrendt, R. Pöttgen: Physica C **469**, 332 (2009)
50.150 S. Jiang, H. Xing, G. Xuan, C. Wang, Z. Ren, C. Feng, J. Dai, Z. Xu, G. Cao: J. Phys. Condens. Matter **21**, 382203 (2009)
50.151 H.-H. Wen, G. Mu, L. Fang, H. Yang, X. Zhu: Europhys. Lett. **82**, 17009 (2008)
50.152 Y. Qi, Z. Gao, L. Wang, D. Wang, X. Zhang, Y. Ma: Supercond. Sci. Technol. **21**, 115016 (2008)
50.153 G.F. Chen, Z. Li, D. Wu, G. Li, W.Z. Hu, J. Dong, P. Zheng, J.L. Luo, N.L. Wang: Phys. Rev. Lett. **100**, 247002 (2008)
50.154 A. Kursumovic, J.H. Durrell, S.K. Chen, J.L. MacManus-Driscoll: Supercond. Sci. Technol. **23**, 025022 (2010)
50.155 J.D. Moore, L.F. Cohen, Y. Yeshurun, A.D. Caplin, K. Morrison, K.A. Yates, C.M. McGilvery, J.M. Perkins, D.W. McComb, C. Trautmann, Z.A. Ren, J. Yang, W. Lu, X.L. Dong, Z.X. Zhao: Supercond. Sci. Technol. **22**, 125023 (2009)
50.156 J. Yang, Z.-C. Li, W. Lu, W. Yi, X.-L. Shen, Z.-A. Ren, G.-C. Che, X.-L. Dong, L.-L. Sun, F. Zhou, Z.-X. Zhao: Supercond. Sci. Technol. **21**, 082001 (2008)
50.157 M.A. McGuire, R.P. Hermann, A.S. Sefat, B.C. Sales, R. Jin, D. Mandrus, F. Grandjean, G.J. Long: New J. Phys. **11**, 025011 (2009)
50.158 C. Shekhar, S. Singh, P.K. Siwach, H.K. Singh, O.N. Srivastava: Supercond. Sci. Technol. **22**, 015005 (2009)
50.159 X.H. Chen, T. Wu, G. Wu, R.H. Liu, H. Chen, D.F. Fang: Nature **453**, 761 (2008)
50.160 K. Togano, Z. Gao, H. Taira, S. Ishida, K. Kihou, A. Iyo, H. Eisaki, A. Matsumoto, H. Kumakura: Supercond. Sci. Technol. **26**, 065003 (2013)
50.161 J.H. Tapp, Z. Tang, B. Lv, K. Sasmal, B. Lorenz, P.C.W. Chu, A.M. Guloy: Phys. Rev. B **78**, 060505(R) (2008)

50.162 L. Shan, Y. Wang, X. Zhu, G. Mu, L. Fang, C. Ren, H.-H. Wen: Europhys. Lett. **83**, 57004 (2008)

50.163 J. Karpinski, N.D. Zhigadlo, S. Katrych, Z. Bukowski, P. Moll, S. Weyeneth, H. Keller, R. Puzniak, M. Tortello, D. Daghero, R. Gonnelli, I. Maggio-Aprile, Y. Fasano, Ø. Fischer, K. Rogacki, B. Batlogg: Physica C **469**, 370 (2009)

50.164 H. Ding, P. Richard, K. Nakayama, K. Sugawara, T. Arakane, Y. Sekiba, A. Takayama, S. Souma, T. Sato, T. Takahashi, Z. Wang, X. Dai, Z. Fang, G.F. Chen, J.L. Luo, N.L. Wang: Europhys. Lett. **83**, 47001 (2008)

50.165 D.S. Inosov, J.S. White, D.V. Evtushinsky, I.V. Morozov, A. Cameron, U. Stockert, V.B. Zabolotnyy, T.K. Kim, A.A. Kordyuk, S.V. Borisenko, E.M. Forgan, R. Klingeler, J.T. Park, S. Wurmehl, A.N. Vasiliev, G. Behr, C.D. Dewhurst: Phys. Rev. Lett. **104**, 187001 (2010)

50.166 X. Zhu, H. Yang, L. Fang, G. Mu, H.-H. Wen: Supercond. Sci. Technol. **21**, 105001 (2008)

50.167 Z. Gao, L. Wang, Y. Qi, D. Wang, X. Zhang, Y. Ma, H. Yang, H. Wen: Supercond. Sci. Technol. **21**, 112001 (2008)

50.168 J. Prakash, S.J. Singh, S. Patnaik, A.K. Ganguli: Physica C **469**, 82 (2009)

50.169 H.Q. Yuan, J. Singleton, F.F. Balakirev, S.A. Baily, G.F. Chen, J.L. Luo, N.L. Wang: Nature **457**, 565 (2009)

50.170 A. Yamamoto, A.A. Polyanskii, J. Jiang, F. Kametani, C. Tarantini, F. Hunte, J. Jaroszynski, E.E. Hellstrom, P.J. Lee, A. Gurevich, D.C. Larbalestier, Z.A. Ren, J. Yang, X.L. Dong, W. Lu, Z.X. Zhao: Supercond. Sci. Technol. **21**, 095008 (2008)

50.171 H. Luetkens, H.-H. Klauss, R. Khasanov, A. Amato, R. Klingeler, I. Hellmann, N. Leps, A. Kondrat, C. Hess, A. Köhler, G. Behr, J. Werner, B. Büchner: Phys. Rev. Lett. **101**, 097009 (2008)

50.172 U. Welp, G. Mu, R. Xie, A.E. Koshelev, W.K. Kwok, H.Q. Luo, Z.S. Wang, P. Cheng, L. Fang, C. Ren, H.-H. Wen: Physica C **469**, 575 (2009)

50.173 C. Ren, Z.S. Wang, H.Q. Luo, H. Yang, L. Shan, H.H. Wen: Physica C **469**, 599 (2009)

50.174 R. Khasanov, A. Maisuradze, H. Maeter, A. Kwadrin, H. Luetkens, A. Amato, W. Schnelle, H. Rosner, A. Leithe-Jasper, H.-H. Klauss: Phys. Rev. Lett. **103**, 067010 (2009)

50.175 S.B. Zhang, X.D. Zhu, H.C. Lei, G. Li, B.S. Wang, L.J. Li, X.B. Zhu, Z.R. Yang, W.H. Song, J.M. Dai, Y.P. Sun: Supercond. Sci. Technol. **22**, 075016 (2009)

50.176 M. Putti, I. Pallecchi, E. Bellingeri, M.R. Cimberle, M. Tropeano, C. Ferdeghini, A. Palenzona, C. Tarantini, A. Yamamoto, J. Jiang, J. Jaroszynski, F. Kametani, D. Abraimov, A. Polyanskii, J.D. Weiss, E.E. Hellstrom, A. Gurevich, D.C. Larbalestier, R. Jin, B.C. Sales, A.S. Sefat, M.A. McGuire, D. Mandrus, P. Cheng, Y. Jia, H.H. Wen, S. Lee, C.B. Eom: Supercond. Sci. Technol. **23**, 034003 (2010)

50.177 P. Mele, K. Matsumoto, K. Fujita, Y. Yoshida, T. Kiss, A. Ichinose, M. Mukaida: Supercond. Sci. Technol. **25**, 084021 (2012)

50.178 S. Trommler, J. Hänisch, V. Matias, R. Hühne, E. Reich, K. Iida, S. Haindl, L. Schultz, B. Holzapfel: Supercond. Sci. Technol. **25**, 084019 (2012)

50.179 S. Adachi, T. Shimode, M. Miura, N. Chikumoto, A. Takemori, K. Nakao, Y. Oshikubo, K. Tanabe: Supercond. Sci. Technol. **25**, 105015 (2012)

50.180 K. Iida, J. Hänisch, S. Trommler, V. Matias, S. Haindl, F. Kurth, I.L. del Pozo, R. Hühne, M. Kidszun, J. Engelmann, L. Schultz, B. Holzapfel: Appl. Phys. Express **4**, 013103 (2011)

50.181 S.-H. Hong, N.H. Lee, W.N. Kang, S.-G. Lee: Supercond. Sci. Technol. **27**, 055007 (2014)

50.182 E.S. Otabe, M. Kiuchi, S. Kawai, Y. Morita, J. Ge, B. Ni, Z. Gao, L. Wang, Y. Qi, X. Zhang, Y. Ma: Physica C **469**, 1940 (2009)

50.183 F. Kametani, A.A. Polyanskii, A. Yamamoto, J. Jiang, E.E. Hellstrom, A. Gurevich, D.C. Larbalestier, Z.A. Ren, J. Yang, X.L. Dong, W. Lu, Z.X. Zhao: Supercond. Sci. Technol. **22**, 015010 (2009)

50.184 M. Eisterer, M. Zehetmayer, H.W. Weber, J. Jiang, J.D. Weiss, A. Yamamoto, E.E. Hellstrom, D.C. Larbalestier, N.D. Zhigadlo, J. Karpinski: Supercond. Sci. Technol. **23**, 054006 (2010)

50.185 S. Haindl, M. Kidszun, A. Kauffmann, K. Nenkov, N. Kozlova, J. Freudenberger, T. Thersleff, J. Hänisch, J. Werner, E. Reich, L. Schultz, B. Holzapfel: Phys. Rev. Lett. **104**, 077001 (2010)

50.186 S.J. Singh, R. Beck, S. Wurmehl, C. Hess, B. Büchner: Supercond. Sci. Technol. **28**, 025006 (2015)

51. Molecular Electronics

Michael C. Petty, Takashi Nagase, Hitoshi Suzuki, Hiroyoshi Naito

The prospects of using organic materials in electronics and optoelectronics applications have attracted scientists and technologists since the 1970s. This field has become known as *molecular electronics*. Some successes have already been achieved, for example, the liquid-crystal display, organic light-emitting displays, and photoreceptors in electrophotography. Other products such as organic photovoltaic devices, chemical sensors and plastic transistors are developing fast. There is also a keen interest in exploiting technologies at the molecular scale that might eventually replace silicon devices. This chapter provides some of the background physics and chemistry to the interdisciplinary subject of molecular electronics. A review of some of the possible application areas for organic materials is presented and some speculation is provided regarding future directions.

51.1 **Electrically Conductive Organic Compounds** 1258
51.1.1 Orbitals and Chemical Bonding 1258
51.1.2 Band Theory 1259
51.1.3 Electrical Conductivity 1260
51.2 **Materials** 1261
51.3 **Plastic Electronics** 1263
51.3.1 Diodes and Transistors 1263
51.3.2 Organic Light-Emitting Structures 1264
51.3.3 Photovoltaic Devices 1265
51.4 **Molecular-Scale Electronics** 1266
51.4.1 Moore's Laws 1266
51.4.2 Nanoscale Organic Films 1267
51.4.3 Patterning Technologies 1269
51.4.4 Molecular Device Architectures 1271
51.5 **DNA Electronics** 1274
51.6 **Conclusions** 1276
References 1276

Molecular electronics is concerned with the exploitation of organic and biological materials in electronics and optoelectronics [51.1–3]. The subject, as it has matured over the last 40 years, can broadly be divided into two themes. The first area, *molecular materials for electronics*, has its origins in materials science and concerns the development of devices that utilize the unique macroscopic properties of organic compounds. The most successful commercial product to date is the liquid-crystal display (LCD). However, following many years of research, devices such as organic light-emitting displays, organic photovoltaic (PV) devices, organic transistors, pyroelectric detectors for infrared imaging, and chemical and biochemical sensors are beginning to make a technological impact. The Nobel prize in chemistry for 2000 was awarded to three scientists working in this area: Alan Heeger, Alan MacDiarmid, and Hideki Shirakawa, who have made significant contributions to the development of electrically conductive polymers. More challenging is *molecular-scale electronics*. Here, the focus is on the behavior of individual organic molecules or groups of molecules. Topics such as molecular switches [51.3–9], molecular memories [51.10–13], molecular rectifiers [51.14], negative differential-resistance junctions [51.15], deoxyribonucleic acid (DNA) electronics [51.16], and molecular manufacturing [51.17, 18] have all been described in the literature. It is much too early to say which, if any, of these could find their way into the commercial arena.

This chapter provides an introduction to the interdisciplinary world of molecular electronics. In the first instance, the physics background to semiconductive organic compounds is outlined. A review of the available materials is presented and some of the possible device applications are described. There are currently a limited number of ways in which organic molecules can be deposited and manipulated on surfaces to form solid films, which can then be used in device structures. The most popular methods are outlined, and methods to pattern the films are described. The prospects for molecular-scale electronics are contrasted with the progress of the inorganic semiconductor industry. Finally, a selection of the ongoing work on molecular-scale devices is described and some speculation about future developments is given.

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_51

51.1 Electrically Conductive Organic Compounds

Metallic and semiconductive behavior is not restricted to inorganic materials. Figure 51.1 shows that the room-temperature conductivity values for organic polymers can extend over much of the spectrum of electrical conductivity, from insulating to semiconducting, and even metallic, behavior [51.19]. The physical explanation can be found in the nature of the chemical bonds that hold solids together.

51.1.1 Orbitals and Chemical Bonding

Carbon-based materials are unique in many respects. This is due to the many possible configurations of the electronic states of a carbon atom. Carbon has an atomic number of six and a valency of four. Its electron configuration is $1s^2$, $2s^2$, $2p^2$, that is, the inner 1s shell is filled and the four valence electrons available for bonding are distributed two each in the 2s orbital and the 2p orbitals. As the 2s orbital is spherically symmetrical (Fig. 51.2) it can form a bond in any direction. In contrast, the 2p orbitals are directed along mutually orthogonal axes and will tend to form bonds in these directions. When two or more of the valence electrons of carbon are involved in bonding with other atoms, the bonding can be explained by the construction of *hybrid* orbitals by mathematically combining the 2s and 2p orbitals. In the simplest case, the carbon 2s orbital hybridizes with a single p orbital. Taking the sum and difference of the two orbitals gives two sp *hybrids*; two p orbitals then remain. The sp orbitals are constructed from equal amounts of s and p orbitals; they are linear and 180° apart.

Other combinations of orbitals lead to different hybrids. For example, from the 2s orbital and two 2p orbitals (e.g., a $2p_x$ and a $2p_y$), three equivalent sp^2 hybrids may be constructed. Each orbital is 33.3% s and 66.7% p. The three hybrids lie in the *xy* plane (the same plane defined by the two p orbitals), directed 120° from each other, and the remaining p orbital is perpendicular to the sp^2 plane. Four sp^3 hybrids may be derived from an s orbital and three p orbitals. These are directed to the corners of a tetrahedron with an angle of 109.5° between the bonds; each orbital is 25% s and 75% p.

A chemical bond can also be formed from a mixture of the above hybrid orbitals, for example, it is possible to have a hybridized orbital that is 23% s and 77% p. Thus, sp, sp^2 and sp^3 hybrids must be considered as limiting cases. Electrons in s orbitals have a lower energy than electrons in p orbitals. Therefore, bonds with more s character tend to be stronger.

Carbon forms four bonds in most compounds, resulting from its four valence electrons. In ethane, C_2H_6, the C–H bonds are all approximately $C(sp^3)$–H(s), while the C–C bond is approximately $C(sp^3)$–$C(sp^3)$. However, in the ethylene molecule, C_2H_4, each of the two carbons is attached to just three atoms while in acetylene, C_2H_2, each carbon atom is only attached to one carbon. In these compounds, the two carbon atoms are bound together by double ($CH_2{=}CH_2$) or triple ($CH{\equiv}CH$) bonds, which involve sp^2 and sp hybrids, respectively. These bonds can be considered to have two distinct components. For example, in ethylene, two sp^2 hybrids on each carbon bond with the hydrogens. A third sp^2 hybrid on each carbon forms a $C(sp^2)$–$C(sp^2)$ single bond, leaving a p orbital *left-over* on each carbon. This orbital lies perpendicular to the plane of the six atoms. The two p orbitals are parallel to each other and have regions of overlap above and below the

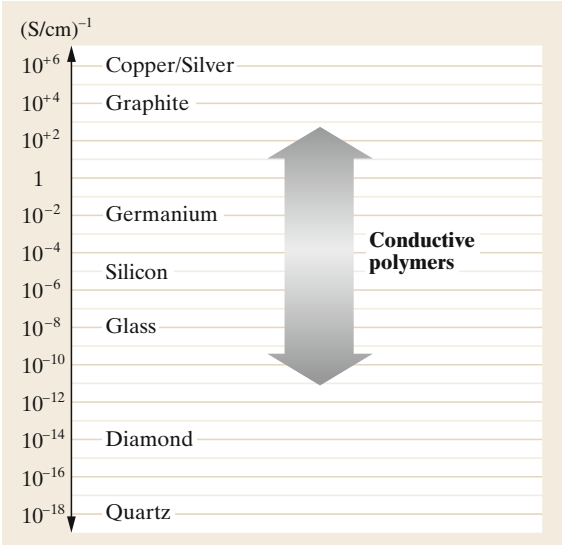


Fig. 51.1 Room-temperature conductivity values of conductive polymers compared to other materials. (After [51.19], with permission from Wiley)

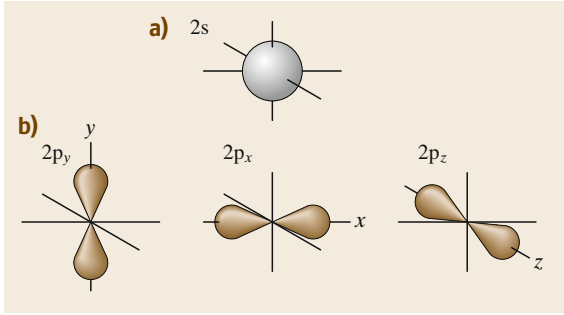


Fig. 51.2 **(a)** 2s orbital. **(b)** $2p_y$, $2p_x$, and $2p_z$ orbitals

molecular plane. This type of bond, consisting of two separate sausage-like electron clouds lying above and below the planes of the carbon nuclei, is called a pi (π) bond. In contrast, the bond formed by the head-on overlap of the two carbon sp^2 orbitals is known as a sigma (σ) bond and the electron cloud is densest at the midpoint between the carbon nuclei. The π bond has little electron density in the region between the carbon atoms and is much weaker than the σ bond. Figure 51.3 illustrates sigma and pi bonding. Carbon compounds that contain π bonds are said to be unsaturated, meaning that the carbon atoms, while having formed the requisite number of bonds, are not fully saturated in terms of their number of potential neighbors. Saturated carbon molecules, on the other hand, contain only single bonds.

51.1.2 Band Theory

The application of quantum mechanics to the electrons in the bonds of inorganic semiconductors led to the development of band theory, one of the great success stories of modern physics. Whenever two identical atoms are brought close together, the electron orbitals overlap and the energy level associated with each electron in the separated atoms is split into two new levels, with one above and one below the original level. A rigorous, quantum-mechanical description of the bonding reveals that the total number of orbitals must be conserved, that is, the total number of molecular orbitals must be the same as the number of atomic orbitals that went into their formation. Consider, for example, the formation of a hydrogen molecule from two separated atoms. When the two atoms approach each other so that their 1s orbitals overlap, two new orbitals (σ bonds) are formed around the atoms, symmetric with respect to the interatomic axis. In one orbital, the bonding orbital, the electron has a lower energy than that in the isolated atomic orbital and in the other, the antibonding orbital, an electron has a higher energy. In the hydrogen molecule, the electron pair normally resides in the bonding orbital. The energy of the electrons in this orbital is lowered relative to that in the atomic orbitals, which is why the atoms remain bound together.

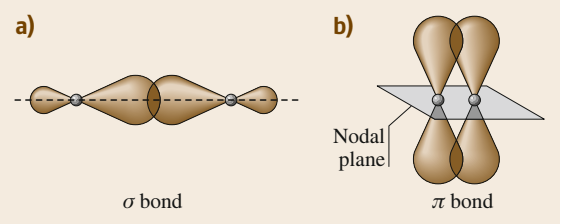


Fig. 51.3 **(a)** Sigma (σ) and **(b)** pi (π) bonding

In an extended solid, many atoms can interact and many similar splittings of energy levels occur. For a solid containing approximately 10^{26} atoms (Avogadro's number) each energy level splits, but the energies between these split levels are very small and continuous ranges or bands of energy are formed. Two such important bands are the valence band and the conduction band, analogous to the bonding and antibonding levels of the two-atom model. The energy gap, or band gap, between them is a forbidden energy range for electrons. Electrical conduction takes place by electrons moving under the influence of an applied electric field in the conduction band and/or holes moving in the valence band. Holes are really vacancies in a band but, for convenience, they may be regarded as positively charged carriers. For an electron, or hole, to gain energy from an applied electric field, and therefore for conductivity to occur, the charge carrier must be able to move into an unoccupied higher energy state. If the carrier cannot be accelerated by the field, then it cannot contribute to the electrical conduction. In a metal, the various energy bands overlap to provide a single energy band that is only partly full of electrons. An insulator has a full valence band and a relatively large energy separation ($> 5\,\text{eV}$) to the higher conduction band. Most semiconductors possess the band structure of an insulator, but a forbidden energy gap of only $0.1-3\,\text{eV}$, so that carriers may be produced in the conduction and/or valence bands by optical or thermal means, or by doping with impurities.

An important feature of the band model is that the electrons are delocalized or spread over the lattice. The strength of the interaction between the overlapping orbitals determines the extent of delocalization that is possible for a given system. For many polymeric organic materials, the molecular orbitals responsible for bonding the carbon atoms of the chain together are the sp^3 hybridized σ bonds, which do not give rise to extensive overlapping. The resulting band gap is large, as the electrons involved in the bonding are strongly localized on the carbon atoms and cannot contribute to the conduction process. This is why a simple saturated polymer such as polyethylene, $(CH_2)_n$, is an electrical insulator.

A significant increase in the degree of electron delocalization may be found in unsaturated polymers, that is, those containing double and triple carbon–carbon bonds. If each carbon atom along the chain has only one other atom, for example, hydrogen, attached to it, the spare electron in a p_z-orbital of the carbon atom overlaps with those of carbon atoms on either side forming delocalized molecular orbitals of π-symmetry. For a simple lattice of length $L = Na$, where N is the total

number of atoms and a is the spacing between them, it can be shown that the total number of electron states in the lowest energy band is equal to N. This result is true for every energy band in the system and applies to three-dimensional (3-D) lattices. Allowing for the two spin orientations of an electron, the Pauli exclusion principle requires that there will be room for two electrons per cell of the lattice in an energy band. If each atom contributes one bonding electron, the valence band will be only half filled.

It might therefore be expected that a linear polymer backbone consisting of many strongly interacting coplanar p_z orbitals, each of which contributes one electron to the resultant continuous π electron system, would behave as a one-dimensional (1-D) metal with a half-filled band. In chemical terms, this is a conjugated chain and may be represented by a system of alternating single and double bonds. It turns out that, for 1-D systems, such a chain can more efficiently lower its energy by introducing bond alternation (alternating short and long bonds). This limits the extent of electronic delocalization that can take place along the backbone. The effect is to open an energy gap in the electronic structure of the polymer. All conjugated polymers are semiconductors, with band gaps more than about 1.5 eV, rather than metals (N.B. the band gap in silicon is 1.1 eV at room temperature). Figure 51.4 shows the chemical π bonding and the electronic band structure of polyacetylene.

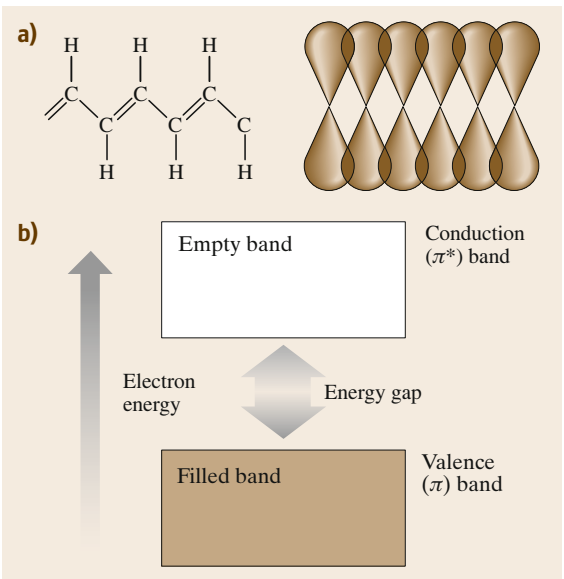


Fig. 51.4 (a) Chemical π bonding in polyacetylene. **(b)** Electronic band structure showing the normally empty π^* band (conduction band) and the normally filled π band (valence band)

51.1.3 Electrical Conductivity

In polyacetylene, the valence band, or π band, is normally filled with electrons while the conduction band, or π^* band is normally empty. Like silicon, the conductivity of polyacetylene can be changed by the addition of impurity atoms. However, the term *doping* is a misnomer as it tends to imply the use of minute quantities, parts per million or less, of impurities introduced into a crystal lattice. In the case of conductive polymers, typically 1–50% by weight of chemically oxidizing (electron withdrawing) or reducing (electron donating) agents are used to alter physically the number of π-electrons on the polymer backbone, leaving oppositely charged counterions alongside the polymer chain. These processes are redox chemistry. For example, the halogen-doping process that transforms polyacetylene to a good conductor is oxidation (or p-doping)

$$[\mathrm{CH}]_n + \frac{3x}{2}\mathrm{I}_2 \rightarrow [\mathrm{CH}]_n^{x+} + x\mathrm{I}_3^- \,.$$

Reductive doping (n-doping) is also possible, for example, using an alkali metal

$$[\mathrm{CH}]_n + x\mathrm{Na} \rightarrow [\mathrm{CH}]_n^{x-} + x\mathrm{Na}^+ \,.$$

In both cases, the doped polymer is a salt. The counterions, I_3^- or Na^+, are fixed in position while the charges on the polymer backbone are mobile and contribute to the conductivity. The doping effect can be achieved because a π electron can be removed (or added) without destroying the σ backbone of the polymer so that the charged polymer remains intact. The increase in conductivity can be as much as 11 orders of magnitude.

The electrical properties of semiconductive organic polymers are not directly comparable to those of silicon. An important material parameter is the mobility of the charge carriers μ. This determines the (additional) velocity that a charge carrier (an electron or hole) acquires because of an applied electric field, and is defined by

$$\mu = \frac{v_\mathrm{d}}{E} \,,$$

where v_d is the drift velocity of the carrier and E is the electric field. The mobility may be further related to the electrical conductivity σ by the expression

$$\sigma = |q| n \mu \,,$$

where n is the density of charge carriers and $|q|$ is the magnitude of their charge (charge on an electron $= 1.6 \times 10^{-19}$ C). The carrier mobility provides an indication of how quickly the carriers react to the field (i. e.,

the frequency response of the material). The greater the degree of electron delocalization, the larger the width of the bands (in energy terms) and the higher the mobility of the carriers within the band. For inorganic semiconductors such as silicon or gallium arsenide, the 3-D crystallographic structure provides for extensive carrier delocalization throughout the solid, resulting in a relatively high mobility μ.

Electrical conduction in polymers not only requires carrier transport along the polymer chains but also some kind of transfer, or *hopping*, between these chains, which tend to lie tangled up like a plate of spaghetti. The charge-carrier mobilities in organic polymers are therefore quite low, making it difficult to produce very high-speed electronic computational devices that are competitive with those based on silicon and gallium arsenide. However, some improvement in the carrier mobility can be achieved by both increasing the degree of order of the polymer chains [51.20, 21] and by improving the purity of the material. Table 51.1 contrasts the room-temperature carrier mobility values for Si, GaAs, and a number of different conductive organic compounds [51.20–26]. The exceptionally high value for graphene ($2\times10^5\,\mathrm{cm^2/Vs}$) should be noted. However, this two-dimensional (2-D) material has a zero band gap, which is an obstacle to its use in certain electronic devices, such as field effect transistors. The values of the other organic materials are low in comparison with those of the inorganic semiconductors, but the field-effect mobilities of $>10\,\mathrm{cm^2/Vs}$ achieved recently make such conjugated small molecules and polymers useful for industrial application. In addition, other features make them more attractive for certain types of electronic devices, as indicated in the later sections.

Table 51.1 Room-temperature field-effect carrier mobilities for field-effect transistors based on organic semiconductors. (After [51.20–26]). The electron mobilities in single-crystal silicon and gallium arsenide are also given

Material	Mobility ($\mathrm{cm^2/Vs}$)
Si single crystal (electrons)	1500
GaAs single crystal (electrons)	8500
Region-regular polythiophene	0.47
Polyacetylene	10^{-4}
Phthalocyanine	10^{-4}–10^{-2}
Thiophene oligomers	10^{-4}–10^{-1}
Organometallic dmit complex	0.2
Pentacene	10^{-3}–1
Bis[1]benzothieno[2,3-d;2′,3′-d′]naphtho[2,3-b;6,7-b′]dithiophene	7
2,7-Dioctylbenzothienobenzothiophene	31
Regioregular poly[4-(4,4-dihexadecyl-4H-cyclopenta[1,2-b:5,4-b']dithiophen-2-yl)-alt-[1,2,5]thiadiazolo[3,4-c]pyridine]	24
C_{60}	0.3
Graphene	2.5×10^5

51.2 Materials

Many conductive polymers have been synthesized to provide certain electronic features (e.g., band gap and electron affinity). The monomer repeat units are often based on five- or six-membered (benzene) carbon ring systems, including polypyrrole, polythiophene (and various other polythiophene derivatives), polyphenelenevinylene, and polyaniline [51.27]. The chemical structure of some of these materials is shown in Fig. 51.5. Such polymers generally show lower electrical conductivity than polyacetylene, however; they can have the advantage of high stability and processability.

Conductive polymers represent only one category of organic electrical conductors. Another important class are the charge-transfer compounds [51.28, 29]. These are formed from a variety of molecules, primarily aromatic compounds (i. e., based on benzene) which can behave as electron donors (d) and electron acceptors (a). Complete transfer of an electron from a donor to an acceptor molecule results in a system that is electrically insulating (e.g., the transfer of a valence electron in a Na atom to a Cl atom, forming the compound NaCl). However, if the ratio of the number of donor molecules to the number of acceptor molecules differs from 1 : 1, for example, the stoichiometry is 1 : 2 or 2 : 3, or if there is incomplete transfer of an electron from a donor to an acceptor (say, six electrons in every ten are transferred), then partially filled electron energy bands can be formed and electron conduction is possible.

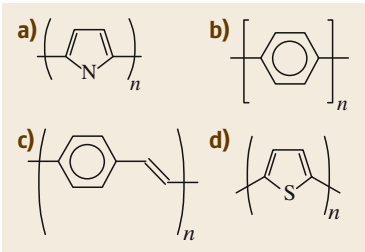


Fig. 51.5a–d Chemical structures of conductive polymers: **(a)** polypyrrole, **(b)** polyparaphenylene, **(c)** polyphenylenevinylene, **(d)** polythiophene

Charge-transfer interactions are strongest between donor molecules of low ionization potential and acceptor molecules of high electron affinity, provided that the donor and acceptor molecules have similar symmetry and are able to approach closely [51.29]. Well-known donor and acceptor molecules are tetrathiafulvalene (TTF) and tetracyanoquinodimethane (TCNQ) (Fig. 51.6a). The latter compound is a very strong acceptor forming first the radical anion and then the dianion (Fig. 51.6b). The stability of the semireduced radical ion, $TCNQ^{-\bullet}$, with respect to the neutral molecule mainly arises from the change from the relatively unstable quinoid structure to the aromatic one, allowing extensive delocalization of the π electrons over the carbon skeleton. As a consequence, TCNQ not only forms typical charge-transfer complexes but is also able to form true radical-ion salts, incurring complete one-electron transfer. Thus, on addition of lithium iodide to a solution of TCNQ, the simple lithium TCNQ salt is formed

$$\mathrm{TCNQ} + \mathrm{LiI} \Leftrightarrow \mathrm{Li^{+}TCNQ^{-\bullet}} + \frac{1}{2}\mathrm{I_2}\,.$$

Following removal of the free iodine precipitate, the TCNQ salt may be crystallized. The crystals show an electronic conductivity of about $10^{-5}\,\mathrm{S/cm}$.

A 1 : 1 TCNQ : TTF salt exhibits a high room-temperature conductivity ($5\times10^{2}\,\mathrm{S/cm}$) and metallic behavior is observed as the temperature is reduced to 54 K. The molecules in such compounds are arranged in segregated stacks, in which the donors and acceptors form separate donor stacks (ddddd...) and acceptor stacks (aaaaa...). The molecules are stacked in such a way that the π bonds on successive molecules can overlap to form bands. This overlap is different from the p-orbital overlap forming the π bands in conjugated polymers. The crystalline packing in these, and other, charge-transfer compounds generally leads to carrier mobility values that are higher than for semiconductive polymers.

Other electroactive compounds that may find application in molecular electronics are based on forms of pure carbon. Graphite consists of vast carbon sheets stacked one on top of another like a sheaf of papers. In pure graphite, these layers are about 0.335 nm apart, but they can be separated further by intercalating various molecules or by using adhesive tape, namely Scotch tape [51.30]. Finally, only one graphite sheet, graphene, can be obtained [51.30]. The bonding between the carbon atoms in the planes is mainly sp^2 hybridizations consisting of a network of single and double bonds. As noted earlier, graphene is a zero band gap material. In case of multilayer graphene, weak interactions between the delocalized electron orbitals hold adjacent sheets together. The delocalized electron system in the planes results in semiconductive electrical behavior.

Under certain conditions, carbon forms regular clusters of 60, 70, 84, etc. atoms [51.31, 32]. A C_{60} cluster, shown in Fig. 51.7, is composed of 20 hexagons and 12 pentagons and resembles a football. The diameter

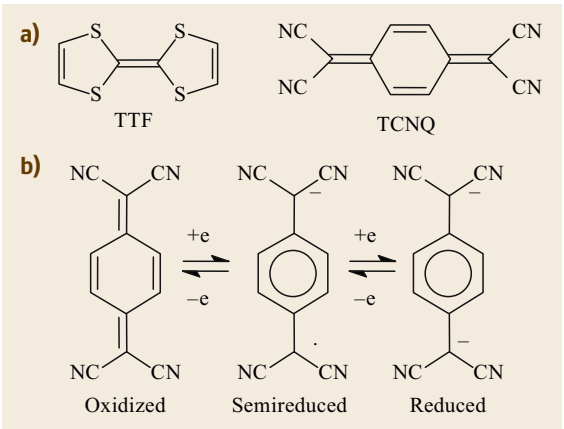


Fig. 51.6 **(a)** Chemical structures of the charge-transfer compounds tetrathiafulvalene (TTF) and tetracyanoquinodimethane (TCNQ). **(b)** Oxidation and reduction of TCNQ

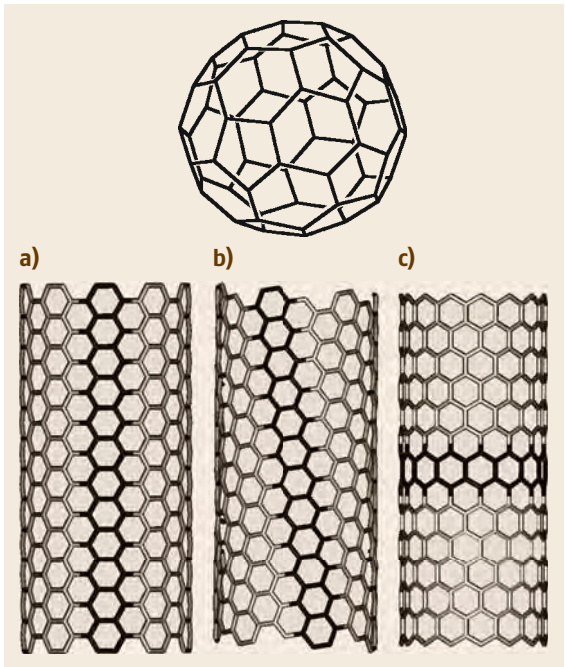


Fig. 51.7a–c Chemical structure of C_{60} (*top*) and three classes of single-wall carbon nanotube (SWNT) **(a)** (10,10) armchair SWNT, **(b)** (12,7) chiral SWNT, and **(c)** (15,0) zigzag SWNT

of the ball is about 1 nm. As with graphite, each carbon atom in C_{60} is bonded to three other carbon atoms. Thus, C_{60} can be considered as a rolled up layer of a single graphite sheet. The term buckministerfullerene was given originally to the C_{60} molecule because of the resemblance to the geodesic domes designed and built by Richard Buckminster Fuller. However, this term (or fullerene or buckyball) is used quite generally to describe C_{60} and related compounds. For example, a molecule with the formula C_{70} can be formed by inserting an extra ring of hexagons around the equator of the sphere, producing an elongated shell more like a rugby ball.

In addition to the spherical-shaped fullerenes, it is possible to synthesize tubular variations – carbon nanotubes [51.32, 33]. Such tubes are comprised of graphite-like sheets curled into a cylinder. Each tube may contain several cylinders nested inside each other. The tubes are capped at the end by cones or faceted hemispheres. Because of their very small diameters (down to around 0.7 nm), carbon nanotubes are prototype 1-D nanostructures. An important feature of a carbon nanotube is the orientation of the six-membered carbon ring in the honeycomb lattice relative to the axis of the nanotube. Three examples of single-wall carbon nanotubes (SWCNs) are shown in Fig. 51.7. The primary classification of a carbon nanotube is as either being chiral or achiral. An achiral nanotube is one whose mirror image has an identical structure to the original. There are only two cases of achiral nanotubes: armchair and zigzag (these names arise from the shape of the cross-sectional ring). Chiral nanotubes exhibit a spiral symmetry whose mirror image cannot be superimposed on the original structure. Carbon nanotubes are characterized by the chiral index (n, m), where the integers n and m specify each carbon nanotube uniquely [51.33]. An armchair nanotube corresponds to the special case $n = m$, while for a zigzag nanotube $m = 0$. All other (n, m) indexes correspond to chiral nanotubes. The electronic structure of a SWCN is either metallic or semiconducting, depending on its diameter and chirality.

At low temperature, a SWNT is a quantum wire in which the electrons in the wire move without being scattered. Resistance measurements for various nanotube samples show that there are metallic and semiconducting nanotubes [51.33]. Carbon nanotubes can also be doped either by electron donors or electron acceptors [51.34]. After reaction with the host materials, the dopants are intercalated in the intershell spaces of the multiwalled nanotubes, and, in the case of single-walled nanotubes, either in between the individual tubes or inside the tubes.

The above confirms carbon's uniqueness as an electronic material. It can be a good conductor in the form of graphite, an insulator in the form of diamond, or a flexible polymer (conductive or insulating) when reacted with hydrogen and other species. Carbon differs from other group IV elements, such as Si and Ge, which exhibit sp^3 hybridization. Carbon does not have any inner atomic orbitals except for the spherical 1s orbital, and the absence of nearby inner orbitals facilitates hybridizations involving only the valence (outer) s and p orbitals. The fact that sp and sp^2 hybridizations do not readily occur in Si and Ge might be related to the absence of *organic materials* made from these elements.

51.3 Plastic Electronics

51.3.1 Diodes and Transistors

Since the discovery of semiconducting behavior in organic materials, there has been a considerable research effort aimed at exploiting these properties in electronic and optoelectronic devices. The term *plastic electronics* refers to electronic devices incorporating polymeric organic compounds (although this term is often used more widely to include devices incorporating other semiconducting organic materials). Organic semiconductors can have significant advantages over their inorganic counterparts. For example, thin layers of conjugated small molecules and polymers can easily be made by low-cost methods such as spin coating. High-temperature deposition from vapour reactants is generally needed for inorganic semiconductors. Synthetic organic chemistry also offers the possibility of designing new materials with different band gaps. As noted in Sect. 51.1.3, the mobilities of the charge carriers in organic field-effect transistors (FETs) have exceeded $10\,cm^2/Vs$. Therefore, such performance and the simple fabrication techniques for conjugated small molecules and polymers have attracted much industrial attention for the fabrication of organic transistor applications such as data storage and thin-film device arrays to address LCDs [51.20–22, 24, 25, 35–37].

Semiconducting organic compounds have been used in a similar fashion to inorganic semiconductors (e.g., Si and GaAs) in metal/semiconductor/metal structures. A diode, or rectifying device, can be made by sandwiching a semiconductor between metals of different work functions. In the ideal case, an n-type

semiconductor should make an Ohmic contact to a low work-function metal and a rectifying Schottky barrier to a high work-function metal [51.38]. One example is that of a semiconductive organic film sandwiched between aluminum and indium-tin-oxide electrodes [51.39]. This device also exhibits PV behavior and is a basic structure of organic PVs.

Organic materials have been used as the semiconducting layer in FET devices [51.22, 40–42]. These are three-terminal structures: a voltage applied to a metallic gate affects the electric current flowing between the source and drain electrodes. For transistor operation, charge must be injected easily from the source electrode into the organic semiconductor and the carrier mobility should be high enough to allow useful quantities of source–drain current to flow. The organic semiconductor and other materials with which it is in contact must also withstand the operating conditions without thermal, electrochemical, or photochemical degradation. Two performance parameters to be optimized in organic FETs are the field-effect mobility and the on/off ratio [51.41].

The operating characteristics of organic transistors and integrated circuits have improved markedly over recent years. This has been brought about by both improvements in the material synthesis and in the thin-film processing techniques [51.20, 21, 24, 25, 43, 44]. State-of-the-art organic FETs possess characteristics exceeding those of devices prepared from hydrogenated amorphous silicon, with mobilities around $1\,\mathrm{cm}^2/\mathrm{Vs}$ and on/off ratios greater than 10^6. As organic FETs are moving closer to applications, their reliability under realistic atmospheric as well as electrical operating conditions has come under more intense scrutiny.

A number of groups have also demonstrated transistor devices incorporating carbon nanotubes [51.45–47]. However, some key issues need to be addressed before nanotubes can be exploited fully in such applications. These include the reproducible fabrication of low-resistance electrical contacts and the accurate control of nanotube growth parameters [51.48]. The nanotube transistor devices realized experimentally have typical dimensions in the micron range. The real promise of carbon nanotube devices, however, lies in the possibility of nanoscale devices. Transistors incorporating graphenes have also been extensively studied because of their high mobility (up to $200\,000\,\mathrm{cm}^2/\mathrm{Vs}$) [51.26, 30].

Thin-film transistors based on organic semiconductors are likely to form key components of plastic circuitry for use as display drivers in portable computers and pagers, and as memory elements in transaction cards and identification tags.

51.3.2 Organic Light-Emitting Structures

Reports of light emission from organic materials upon the application of an electric field (electroluminescence) have been around for many years. However, there has been an upsurge in interest following the initial report of organic light-emitting devices (OLEDs) incorporating the conjugated small molecules (tris(8-hydroxyquinolinato)aluminum, Alq_3 (Fig. 51.9) [51.49] or the conjugated polymer polyphenylenevinylene (PPV (Fig. 51.5)) [51.50]). The simplest OLED is an electroluminescent compound sandwiched between metals of high and low work function, as depicted in Fig. 51.8. The anode electrode is normally indium tin oxide (ITO) as this material is semitransparent, allowing the light out of the device. On application of a voltage, electrons are injected from the low work function electrode into the lowest unoccupied molecular orbital (LUMO) level (conduction or π^* band in the case of an organic compound possessing a delocalized electron system) of the organic compound and holes from the high work-function electrode into the highest occupied molecular orbital (HOMO) level (valence or π band in the case of an organic compound possessing a delocalized electron system). The recombination of these oppositely charged carriers then results in the emission of light. Work is focused on the use of low molecular-weight organic molecules and polymers and there is considerable industrial interest in the application of such materials to various display technologies [51.51–54]. It is estimated that the global

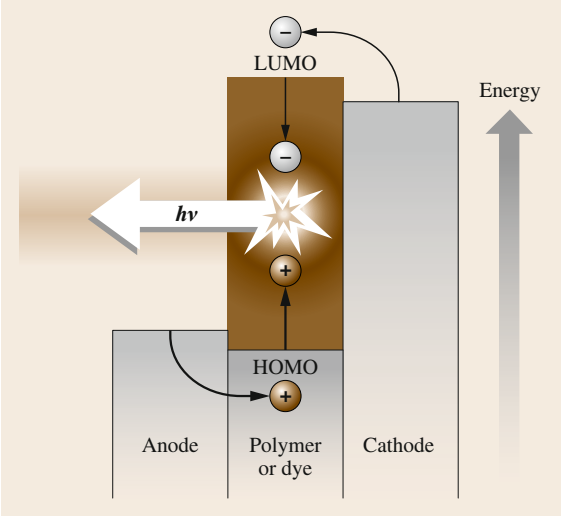


Fig. 51.8 Schematic energy band structure of an organic light-emitting device (OLED). The recombination of electrons and holes results in the emission of light of frequency ν and energy $h\nu$

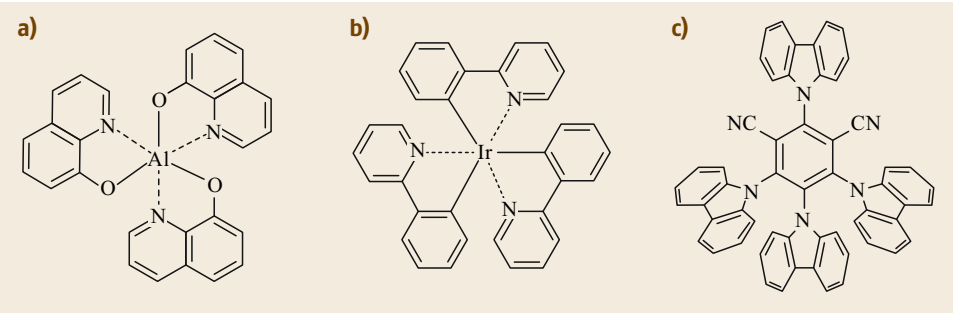


Fig. 51.9a–c Representative chemical structures of green **(a)** fluorescent, **(b)** phosphorescent, and **(c)** TADF emitters

market for OLED displays will increase to $16 billion by 2020 [51.55]. Organic light-emitting devices offer a number of advantages over LCDs and the other technologies that currently dominate the flat-panel display market: OLEDs do not require backlighting and they can be made thinner and lighter. Furthermore, OLED displays provide higher contrast and truer colors, higher brightness, wider viewing angles, better temperature stability and faster response times than LCDs [51.51].

Many techniques have been used in attempts to optimize the performance of OLEDs. For example, a thin inorganic insulating layer such as LiF, or an organic monolayer may be inserted between the cathode and the emissive material [51.56–58]. A thin insertion layer of poly(ethylene imine) (PEI) [51.54, 58] between a metal oxide such as ITO or ZnO as a bottom cathode and an emitting layer greatly improves the performance of OLEDs. Such inverted OLEDs, or iOLEDS, possess the configuration of substrate/cathode/PEI/emitting layer/anode (the configuration of conventional OLEDs shown in Fig. 51.8 is substrate/anode/emitting layer/cathode). In recent years, iOLEDs have attracted much attention because the absence of air-sensitive electron injection materials such as LiF, Ca, and Cs results in a long device lifetime [51.59, 60]. Stringent encapsulation of flexible/stretchable devices on flexible substrates, such as a plastic film, is rather difficult and hence the development of iOLEDs is essential for future flexible/stretchable displays.

Electron- and hole-transporting layers can also be introduced between the cathode and the emitting layer and between the anode and the emitting layer, respectively, to improve and balance the injection of charge carriers [51.61].

When an electron and a hole recombine to form an excited molecular state in an organic material, the spins of the electrons in the excited and ground levels can either point in the same direction (triplet state) or in the opposite directions (singlet state). For quantum-mechanical reasons, 75% of recombination events are associated with triplet states which, in most cases, do not emit photons when they decay to the ground state. Hence, the production of emission from the triplet state of organic materials is a further means to improve the device efficiency [51.62, 63]. Indeed, the internal quantum efficiency, which is defined as the ratio of the number of photons that can be extracted for electroluminescence to the injected current (i. e., the number of injected carriers), is 25% in fluorescent emitters. In contrast, the internal quantum efficiency of OLEDs using phosphorescent materials is 100% [51.64]. However, the design of the phosphorescent materials is greatly limited because the heavy atom effect (spin–orbital interaction) must be induced, for example, by using rare metals, to realize highly efficient radiative transition from the triplet excited state to the ground state. The internal quantum efficiency of 100% has also been achieved using the up-conversion of triplet excitons to singlet excitons by thermal energy, known as thermally activated delayed fluorescence (TADF) [51.65]. TADF emitters can include ubiquitous atoms such as C, H, and N without using rare metals. Recent studies have revealed that highly efficient delayed fluorescence can be achieved by optimizing the molecular design [51.62]. Representative fluorescent, phosphorescent, and TADF emitters are shown in Fig. 51.9.

Table 51.2 shows some important parameters of polymer-based OLEDs with different color outputs [51.54]. There is also a keen interest in developing white-light organic displays. For example, a large-area white-light-emitting OLED could provide a solid-state light source that might compete with conventional lighting technologies. Different methods of making an intrinsically white-emitting OLED by blending emissive species, either in single or multiple layers have been demonstrated [51.66–70].

51.3.3 Photovoltaic Devices

Concerns over global climate change, local air pollution, and resource depletion are making PVs an increasingly attractive method of energy supply. The current technology is based on single-crystal silicon solar cells. These have developed since the 1940s and now possess

Table 51.2 Performance of polymer-based organic light-emitting devices. (After [51.54], Data from Cambridge Display Technology)

Colour	Voltage (V)	Efficiency at 1000 cd/m^2 (cd/A)	Measured T50 lifetime[a] at fixed luminance (h) at 1000 cd/m^2
Red	4.2	31	350 000
Green	4.7	72	350 000
Blue	3.7	12	> 10 000

[a] T50 lifetime is defined as luminance depreciation to 50% of its initial value

conversion efficiencies of around 15% for commercial devices (although figures of around 25% are reported in the laboratory) [51.71]. However, the technology is more expensive than conventional power generation, and there is much research on alternative materials. Photovoltaics using organic compounds, such as polymers or dyes, offer the possibility of large-scale manufacture at low temperature coupled with low cost. Until the end of the twentieth century, little progress had been made and energy conversion efficiencies of up to only about 1% were achieved [51.72]. However, the availability of new conductive organic materials and different PV designs has significantly improved on this figure. To 2015, several laboratories have reported conversion efficiencies of around 10% [51.71, 73–76].

An organic solar cell device is very similar in structure to the OLED described in the previous section. If the incoming photons have energy greater than the band gap of the polymer (or greater than the HOMO–LUMO separation in the case of organic molecular materials) then the light will be absorbed, creating electrons and holes. In an inorganic PV cell, these electrons and holes would be generated within, or close to, a depletion region in the semiconductor and they would be free to migrate to opposite electrodes, where they can do useful work in an external electrical load. However, in the organic material the electrons and holes are bound together in excitons. An immediate problem in organic PV cells is to split these excitons. This can be conveniently done at an interface, the simplest being the junction between the electrodes and the organic material. Under open-circuit conditions, holes are collected at the high work-function electrode (e.g., ITO) and electrons at the low work-function electrode (e.g., Al). Generally, the open-circuit output voltage of the PV device depends on the work function difference between the electrodes. However, in a p-n junction, the maximum available voltage is determined by the difference of the quasi Fermi levels of the two charge carriers, that is, n-doped semiconductor energy level and p-doped semiconductor energy level, respectively. In organic PV, the open-circuit voltage is found to be linearly dependent on the HOMO level of the electron-donating material (p-type semiconductor quasi Fermi level) and the LUMO level of the electron-accepting material (n-type semiconductor quasi Fermi level) [51.77–79]. Improvements in the efficiency of the exciton-splitting process can be achieved using organic compounds incorporating electron-donating and electron-accepting species. By creating interfaces of differing electron affinities, it is possible to enhance the probability of electron transfer between the molecules. The creation of the interfaces is carried out by blending electron-donating and electron-accepting species, to produce a bulk heterojunction [51.80–82]. This exhibits a donor–acceptor phase separation on a 10–20 nm length scale. In such a nanoscale interpenetrating network, each interface is within a distance less than the exciton diffusion length from the absorbing site. The bulk heterojunction concept significantly increases the interfacial area between the donor and acceptor phases and results in improved efficiency solar cells.

An alternative approach to organic PVs exploits a dye-sensitized solar cell, or *Grätzel* cell [51.83]. Here, the incoming photons are absorbed by molecules of a dye on a semiconductor surface with subsequent energy and electron transfer to the semiconductor. An electron is returned to the oxidized dye via an electrolyte. The efficiency of such devices is 11.9% [51.71, 84].

51.4 Molecular-Scale Electronics

51.4.1 Moore's Laws

The second strand to molecular electronics (molecular-scale electronics) recognizes the spectacular size reduction in the individual processing elements in integrated circuits over recent years. The first microprocessor chip manufactured in 1972 by Intel (8008) had a clock speed of 200 kHz and contained 3500 transistors. There are 1.4 billion transistors on the Pentium Core i7 chip (June, 2014), fabricated using 22 nm process technology, and operating at a clock speed of 3–4 GHz.

Moore's law (or Moore's first law) states that the functions per chip double every 1.5 years. This will probably describe developments over at least the next

decade. The semiconductor industries have produced an international technology roadmap for the future of complementary metal oxide semiconductor (CMOS) technology [51.85]. Figure 51.10 shows the anticipated growth in the density of the transistors in both the microprocessor unit (MPU) and the dynamic random-access memory (DRAM) of a CMOS chip over the next decade. The prediction is for a 10 nm minimum feature size (gate length) for the MPU and 10^{11} transistors per cm^2 in the case of the memory by the year 2022. These figures are regularly updated. This device density is remarkable and provides the means to store significant amounts of data (Table 51.3 [51.3]). However, molecular-scale electronics has the potential for further increases in device density. For example, using $\approx 1-3$ nm organic molecules as the processing elements, $10^{13}-10^{14}$ *devices* could be fitted into 1 cm^2 [51.3].

There are, however, a number of technological issues that will need to be overcome for the predictions for the CMOS-based roadmap to be realized [51.86, 87]. Not least are the materials limitations of the silicon/silicon dioxide system. For example, charge leakage becomes a problem when the insulating silicon dioxide layers are thinned to a few nanometer. The scaling of physical dimensions can also be limited by minute variations in size and spatial fluctuations in the doping concentration. Lithographic techniques will face enormous challenges to achieve the high controllability needed to eliminate variations from device to device. Heat dissipation is another key factor. A Pentium Core i7 chip with about 1 billion transistors operating at the current nanosecond rate can emit 60–90 W of heat. Financial issues are also problematic as ever more complex integrated circuits are produced. Intel's Fab 32, a chip-fabrication facility (FAB) that opened in Chandler, AZ, USA, in August 2005, cost $3 billion to construct and equip. The cost of building a fab is expected to rise to over $10 billion by 2025 [51.88]. This significant increase in cost (Moore's second law) is due to the extremely sophisticated tools that will be needed to form the increasingly small features of the devices. Molecular-scale technology will, of course, also need to address such problems – plus many more.

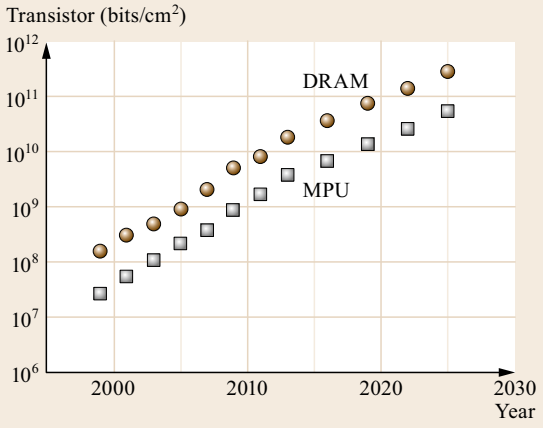


Fig. 51.10 Predicted feature size in CMOS devices. (After [51.85])

Table 51.3 Information content for various *applications*. (After [51.3])

Application	Typical information content (bytes)
Color photograph	10^5
Average book	10^6
Desktop computer	10^8
Genetic code	10^{10}
Human brain	10^{13}
Library of congress	10^{15}

51.4.2 Nanoscale Organic Films

Before organic molecules can be exploited in device architectures (i. e., those familiar from Si-based microelectronics) ways must be found to deposit them onto surfaces. Well-established methods of organic film deposition include electrodeposition, thermal evaporation, and spinning [51.89]. The Langmuir–Blodgett (LB) technique, self-assembly, and layer-by-layer electrostatic deposition are further means for producing layers of organic materials. These allow ultrathin-film assemblies of organic molecules to be engineered at the molecular level and are of particular relevance to molecular-scale electronics [51.90, 91].

Langmuir–Blodgett films are prepared by first depositing a small quantity of an amphiphilic compound (i. e., one containing both polar and nonpolar groups) dissolved in a volatile solvent onto the surface of purified water [51.91, 92]. The classical materials are long-chain fatty acids, such as *n*-octadecanoic acid (stearic acid). When the solvent has evaporated, the organic molecules may be organized into a floating two-dimensional *crystal* by compression on the water surface. As the area available to the organic molecules is reduced, the floating film will undergo several phase transformations. These are, to a first approximation, analogous to 3-D gas, liquid, and solid phases. The phase changes may readily be identified by monitoring the surface pressure as a function of the area occupied by the molecules in the film. This is the two-dimensional equivalent to the pressure-versus-volume isotherm for a gas. In the *gaseous* state the molecules are far enough apart on the water surface that they

exert little force on one another. As the surface area of the monolayer is reduced, the hydrocarbon chains will begin to interact. The *liquid* state that is formed is generally called the expanded monolayer phase. The hydrocarbon chains of the molecules in such a film are in a random, rather than regular orientation, with their polar groups in contact with the subphase. As the molecular area is progressively reduced, condensed phases may appear. There may be more than one of these and the emergence of each condensed phase can be accompanied by constant-pressure regions of the isotherm, as observed in the cases of a gas condensing to a liquid and a liquid solidifying. In the condensed monolayer states, the molecules are closely packed and are oriented with their hydrocarbon chain pointing away from the water surface. The area per molecule in such a state will be similar to the cross-sectional area of the hydrocarbon chain, that is, $\approx 0.19\,\mathrm{nm}^2/\mathrm{molecule}$.

If the surface pressure is held constant in one of the condensed phases, then the film may be transferred from the water surface onto a suitable solid substrate simply by raising and lowering the latter through the monolayer–air interface. Figure 51.11 shows a schematic diagram of the equipment required for LB film deposition. In this technique, introduced by Langmuir and Blodgett [51.91] the floating condensed monolayer is transferred, like a carpet, as the substrate is raised and/or lowered through the air/monolayer interface. A number of different LB deposition *modes* are possible. The most commonly encountered situation is Y-type deposition, which refers to monolayer transfer on both the upward and downward movements of the substrate. Instances in which the floating monolayer is only transferred to the substrate as it is being inserted into the subphase, or only as it is being removed, are also observed. These deposition modes are called X-type and Z-type depositions, respectively.

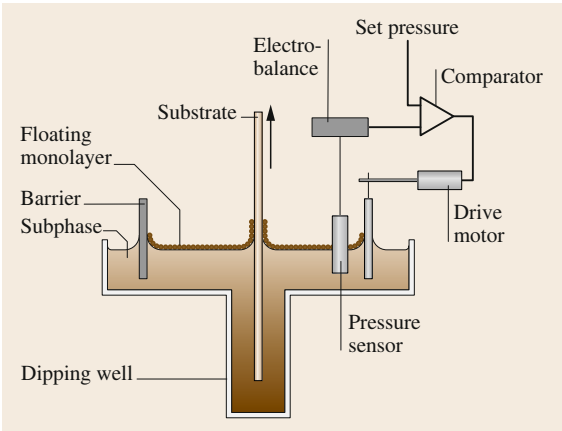


Fig. 51.11 System for the deposition of Langmuir–Blodgett films

Self-assembly is a much simpler process than that of LB deposition. Monomolecular layers are formed by the immersion of an appropriate substrate into a solution of the organic material [51.90]. The best known examples of self-assembled systems are organosilicon on hydroxylated surfaces (SiO_2, Al_2O_3, glass etc.) and alkanethiols on gold, silver, and copper [51.93]. However, other combinations include:

- Dialkyl sulfides on gold
- Dialkyl disulfides on gold
- Alcohols and amines on platinum
- Carboxylic acids on aluminum oxide and silver
- Phosphonic acids on aluminum oxide and ITO.

The self-assembly process is driven by the interactions between the head group of the self-assembling molecule and the substrate, resulting in a strong chemical bond between the head group and a specific surface site, for example, a covalent Si–O bond for alkyltrichlorosilanes on hydroxylated surfaces.

The combination of the self-assembly process with molecular recognition offers a powerful route to the development of nanoscale systems that may have technological applications as chemical sensing or switching devices. For instance, the complexation of a neutral or ionic guest at one site in a molecule may induce a change in the optical or redox properties of the system. Monolayers containing both an electroactive TTF unit and a metal-binding macrocycle assembled onto platinum have been shown to exhibit electrochemical recognition to Ag^+ ions [51.94].

The self-assembly process, as described above, is usually restricted to the deposition of a single molecular layer on a solid substrate. However, chemical means can be exploited to build up multilayer organic films. A method pioneered by *Sagiv* is based on the successive absorption and reaction of appropriate molecules [51.95, 96]. The head groups react with the substrate to give a permanent chemical attachment and each subsequent layer is chemically attached to the one before in a very similar way to that used in systems for supported synthesis of proteins.

Another technique for building up thin films of organic molecules is driven by the ionic attraction between opposite charges in two different polyelectrolytes, the so-called *layer-by-layer* assembly technique [51.97–100]. A solid substrate with a positively charged planar surface is immersed in a solution containing an anionic polyelectrolyte and a monolayer of polyanion is adsorbed (Fig. 51.12). Since the adsorption is carried out at relatively high concentrations of the polyelectrolyte, most of the ionic groups remain ex-

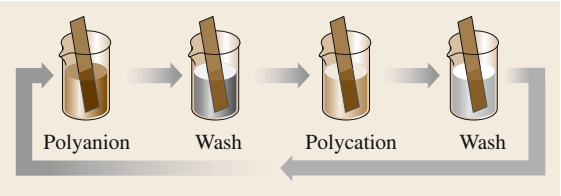


Fig. 51.12 Deposition of layer-by-layer polyelectrolyte films by the sequential immersion of a solid substrate in solutions of the polyanions and polycations

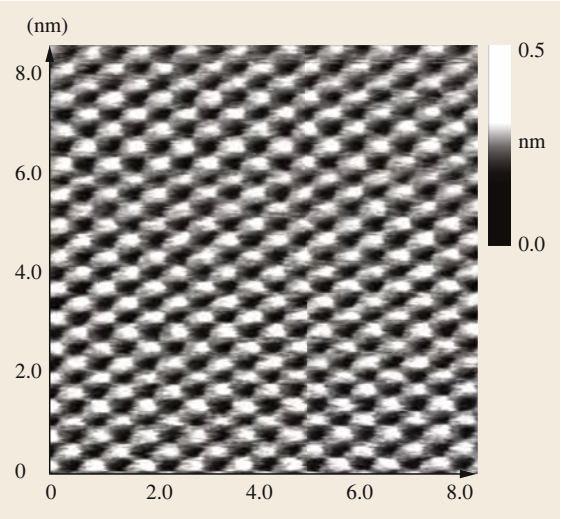


Fig. 51.13 Atomic force micrograph of a 12-layer, *n*-eicosanoic-acid Langmuir–Blodgett film deposited onto silicon. (From [51.105], with permission from ACS)

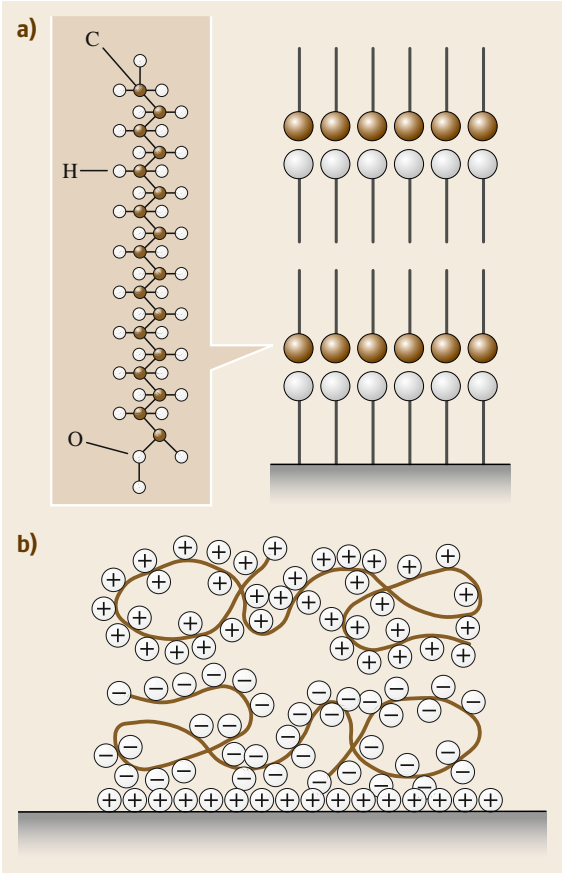


Fig. 51.14a,b Molecular organization in **(a)** a Langmuir–Blodgett multilayer assembly and **(b)** a polyelectrolyte multilayer deposited by the electrostatic method

posed to the interface with the solution and thus the surface charge is reversed. After rinsing in pure water, the substrate is immersed in a solution containing the cationic polyelectrolyte. Again, a monolayer is adsorbed but now the original surface charge is restored, resulting in the formation of a multilayer assembly of both polymers. It is possible to use a sensitive optical technique, such as surface plasmon resonance to monitor, in situ, the growth of such electrostatically assembled films [51.101]. The layer-by-layer method has been used to build up layers of conductive polymers, for example, partially doped polyaniline and a polystyrene polyanion [51.102]. Biocompatible surfaces consisting of alternate layers of charged polysaccharides and oppositely charged synthetic polymers can also be deposited in this way [51.103]. A related, but alternative, approach uses layer-by-layer adsorption driven by hydrogen-bonding interactions [51.104].

The organization of the organic molecules in multilayer assemblies may be investigated by a number of analytical techniques including x-ray and neutron reflection, electron diffraction, and infrared spectroscopy [51.91]. Figure 51.13 shows an example of an atomic force micrograph of a 12-layer *n*-eicosanoic acid LB film deposited onto single-crystal silicon [51.105]. Lines of individual molecules are evident at the magnification shown. Figure 51.14 contrasts the molecular organization expected in an LB multilayer with that in electrostatically deposited layer-by-layer films. For the latter case, the polyelectrolyte chains within each layer will become entangled, and may even penetrate into the layers above and below, leading to a less-ordered film than that produced by LB deposition.

51.4.3 Patterning Technologies

The problem of connecting together the individual processing elements in any future molecular computer is challenging. Each one of the 1.4 billion transistors in the Pentium Core i7 chip is addressable and connected to a power supply. Organic molecules can be difficult

to arrange on a surface or in a 3-D array such that each molecule is addressable. Planar inorganic materials are normally patterned using photolithography. Here, a surface is first covered with a light-sensitive photoresist, which is exposed to ultraviolet light through a contact mask. Either the exposed photoresist (positive resist) or the unexposed regions (negative resist) can then be developed to leave a positive or negative image of the mask on the surface. This approach is routinely used in the fabrication of devices based on inorganic semiconductors. However, difficulties can be encountered when used with organic films, as the photoresists themselves are based on organic compounds.

Brittain et al. [51.106] describe a series of *soft* lithographic methods that may be better suited to the patterning of organic layers. Pouring a liquid polymer, such as polydimethylsiloxane (PDMS) onto a *master* made from silicon forms a pattern-transfer element. The polymer is allowed to cure to form an elastomer, which can then be removed from the master. This replica can subsequently be used as a stamp to transfer chemical ink, such as a solution of an alkanethiol, to a surface.

Scanning microscopy methods offer a powerful means of manipulating molecules. Careful control of an atomic force microscope (AFM) tip can allow patterns to be drawn in an organic film [51.108, 109]. Such techniques can also be used to reposition molecules, such as the fullerene C_{60}, on surfaces and to break up an individual molecule [51.110]. A further approach that has recently been developed is called dip-pen nanolithography (DPN) [51.107] (Fig. 51.15). This is able to deliver organic molecules in a positive printing mode. An AFM tip is used to *write* alkanethiols on a gold thin film in a manner analogous to that of a fountain pen. Molecules flow from the AFM tip to a solid substrate (*paper*) via capillary transport, making DPN a potentially useful tool for assembling nanoscale devices. Recent developments of DPN have included an overwriting capability that allows one nanostructure to be generated and the areas surrounding that nanostructure to be filled with a second type of ink [51.111]. Perhaps the greatest

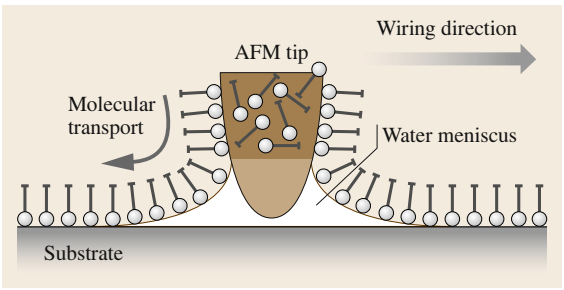


Fig. 51.15 Dip-pen patterning showing the transfer of an *ink* onto *paper* using the tip of an AFM. (After [51.107])

limitation in using scanning probe methodologies for ultrahigh-resolution nanolithography over large areas derives from the serial nature of most techniques. However, multiple-pen nanoplotters capable of performing parallel lithography have been reported [51.112, 113]. In addition, PDMS elastomeric tip arrays of up to 11 million pens that were made with a silicon master have been developed for low-cost, high-throughput patterning [51.114]. The DPN method has also been used to deposit magnetic nanostructures [51.115] and arrays of protein molecules [51.116].

The need to combine large-area coatings with device patterning has resulted in the development of direct-write fabrication methods, such as ink-jet printing [51.117–120]. Although ink-jet printhead droplet ejection can be achieved with thermal (bubble-jet) and piezoelectric modes of operation, the majority of published literature on ink-jet printing as a tool for manufacturing organic devices has been the result of using piezoelectric-actuated printers. Piezoelectric printhead technology is favored primarily because it applies no thermal load to the organic *inks* and is compatible with the printing of digital images. The combination of solution-processable emissive polymers with ink-jet printing offers some promise in the development of low-cost high-resolution displays [51.121]. The technique has also been applied to the manufacturing of all-polymer transistor circuits [51.122, 123]. Ultrafine printing with 1 μm resolution has recently been demonstrated using an ink-jet system with subfemtoliter ejection accuracy [51.124].

Liquid thin films exhibit a number of unique physical processes, such as dewetting, pinning, and capillary action, and these processes have been investigated as a potential tool of patterning organic micro- and nanostructures on large-area substrates by low-cost solution techniques. For example, the control of the dewetting behavior of liquid thin films using substrates having a prepatterned variation of surface energies or structures has been used to guide organic materials to desired surfaces [51.126, 127]. The capillary action of the liquid into narrow channels has also been utilized to manipulate and deposit organic materials [51.128, 129] and biomaterials [51.130, 131], in which PDMS stamps have been adopted to form hydrophobic channels. In several of these approaches, pattern length scale and feature sizes are limited to the lateral dimension of prepatterned templates. However, further investigation of the self-assembly of organic molecules and physical phenomena in liquid thin films enables us to downscale dimensions relative to the size of the templates. An example is shown in Fig. 51.16 [51.125, 132]. This method uses the convex surfaces of an elastomeric stamp as forming positive motifs. As the stamp is gen-

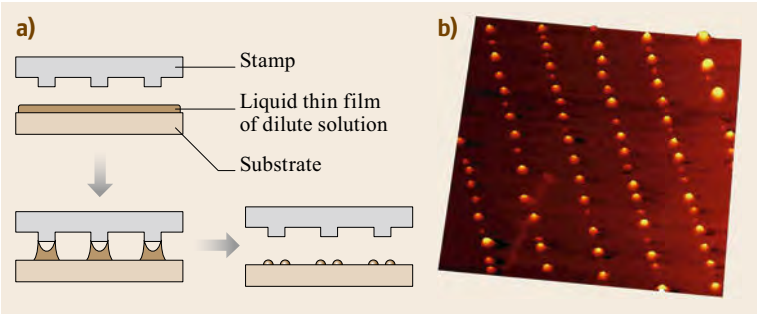


Fig. 51.16 **(a)** Schematics of stamp-assisted deposition process and **(b)** atomic force micrograph of dotted structures printed from a diluted solution of molecular magnets. (After [51.125], with permission from the ACS)

tly placed on a liquid thin film made of a dilute solution, capillary forces drive the solution to form menisci under the stamp instead of filling the grooves, as shown in Fig. 51.16a. Depending on the intermolecular and molecule–substrate interactions, organic molecules are self-assembled on the substrate into patterns of split nanowires or strings of nanodots (Fig. 51.16b). A capability of the method for nanopatterning has been demonstrated using a molecular magnet [51.125] and a molecular semiconductor [51.132].

51.4.4 Molecular Device Architectures

The *bottom-up* approach to molecular electronics offers many intriguing prospects for manipulating materials on the nanometer scale, thereby providing opportunities to build up novel architectures with predetermined and unique physical and/or chemical properties. Two relatively simple examples of organic superlattice structures are the incorporation of an electric polarization into a multilayer array to form thin films exhibiting pyroelectric behavior [51.133] and the use of noncentrosymmetric layers for second-order nonlinear optical response, for example, second-harmonic generation [51.134]. In both instances, the multilayer thin films can be built up using the LB approach.

To realize functional nanoelectronic circuits, a number of workers have investigated the electrical characteristics of structures in which organic molecules are sandwiched between two metallic electrodes. Of particular interest is the possibility of observing molecular rectification using monolayer or multilayer films. This follows the prediction [51.135] that an asymmetric organic molecule containing a donor and an acceptor group separated by a short σ-bonded bridge, allowing quantum-mechanical tunneling, should exhibit diode characteristics. There have been many attempts to demonstrate this effect in the laboratory, particularly in organic thin films [51.136–138]. Asymmetric current-versus-voltage behavior has certainly been recorded for many metal/insulator/metal structures, although these results are often open to several interpretations as a result of the asymmetry of the electrode configuration.

In other cases, *switching* behavior has been reported [51.3–8, 12, 13, 138, 139]. As with the work on molecular rectification, the origin of the switching is not always clear. For example, is this a property of the organic molecules, the metallic electrode or of the presence of any interfacial (e.g., oxide) layer (or a combination of these)? Bistable rotaxane molecules have been used as the basis of some of these studies; an example is shown in Fig. 51.17a. This molecule is amphiphilic and the ring component can move between the polar and nonpolar regions of the main part of the molecule. The molecules can be assembled on an electrode using the LB approach and a top electrode is then deposited to form the crossbar structure shown in Fig. 51.17b [51.5–7]. Alternatively, the molecules can be solution-cast between Pt source and drain electrodes (with gaps of 1–2 nm) in transistor structures [51.9]. In some of the experiments, the absence of switching using *control* compounds suggests that the rotaxane molecule itself is responsible for the bistability [51.7].

Bistable switching intrinsic to a molecular phenomenon has also been reported in single molecular layers formed by the self-assembly approach [51.139, 140]. A switching behavior trigged by an external voltage in a substituted conjugated molecular wire of bipyridy-dinitro oligophenylene-ethynylene dithiol [51.139] has been confirmed by studies both on monolayers and isolated molecules using various two-terminal configurations, such as crossed wire electrodes, a scanning tunneling microscope and a pair of electrodes with a nanometer-sized gap [51.140]. The experimental and theoretical studies indicate that such voltage-triggered switching originates in the presence of the nitro groups and their dipole moment facilitates molecular rotation during voltage sweeps [51.140, 141]. However, the fabrication of monolayer devices sandwiched metallic electrodes often suffers from electrical shorts, due to structural defects and/or the penetration of metallic atoms in the organic layer. In a pioneer-

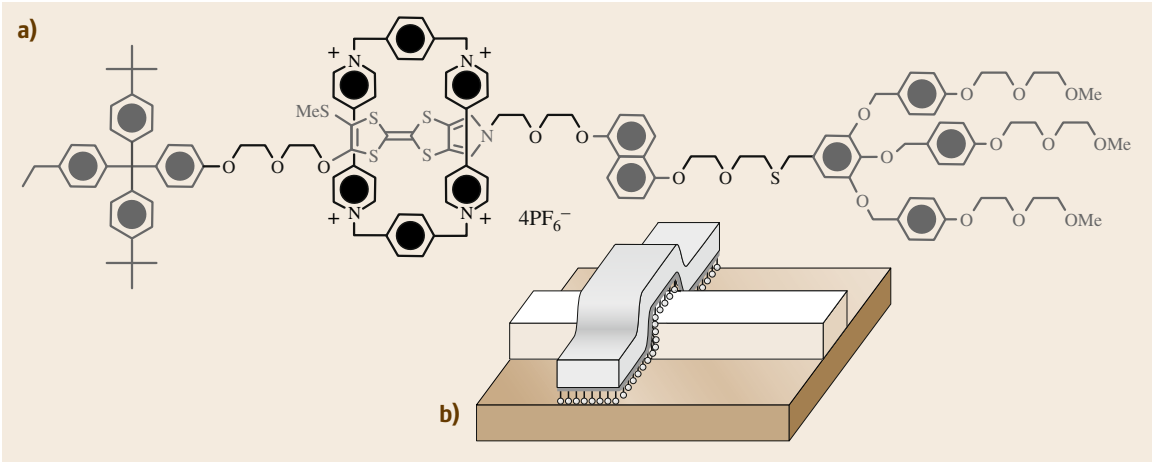


Fig. 51.17 (a) Bistable rotaxane. (b) Schematic diagram of cross-wire structure for switching studies. (After [51.5])

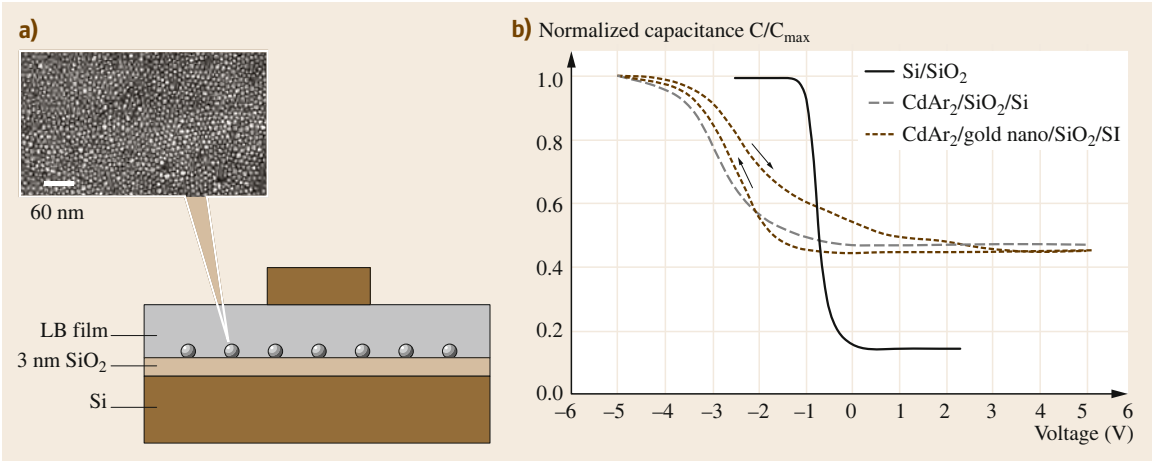


Fig. 51.18 (a) Metal–insulator–semiconductor (MIS) structure incorporating gold nanoparticles; a transmission electron micrograph of the nanoparticles is shown. (b) Normalized capacitance versus voltage characteristics for different MIS devices. (After [51.10], © 2003 ACS)

ing work, monolayer formation in nanopores patterned by lithography has been exploited in device fabrication [51.15]. A simple approach that uses conducting thin films such as poly(3,4-ethylenedioxythiophene)-poly(styrenesulfonate) (PEDOT:PSS) [51.142] and multilayer graphene [51.143] as protecting interlayers for the metal penetration has recently been developed for high-yield, large-area fabrication of monolayer device arrays without electrical shorts. This technique has also been used to the demonstrate switching behavior in monolayer devices comprising photochromic molecules [51.144] and transition metal complex molecules [51.145]. Such results augur well for the development of molecular-scale logic circuitry.

Nanoscale organic devices can also exploit charge storage on nanoparticles or at interfaces. One important metal oxide semiconductor (MOS) device is the flash memory [51.146]. This is similar in structure to a MOS field-effect transistor (MOSFET), except that it has two gate electrodes, one on top of the other. The top electrode forms the control gate, below which a *floating gate* is capacitively coupled to the control gate and the underlying silicon. The memory cell operation involves putting charge on the floating gate or removing it, corresponding to two logic levels. Nanoflash devices utilize single or multiple nanoparticles as the charge-storage elements. These are usually embedded in the gate oxide of a FET and located in close proximity (2–3 nm) to the transistor channel [51.147]. Figure 51.18 shows the characteristics of a memory device incorporating metallic nanoparticles deposited by the LB technique [51.10]. The device structure is shown schematically in the fig-

ure. The gold nanoparticles (Q-Au) were of nominal diameter 10 nm and passivated with an organic capping layer. Cadmium arachidate ($CdAr_2$) LB layers were used to provide an insulating gate layer. Figure 51.18 also shows the normalized capacitance versus voltage (C–V) data, measured at 1 MHz and at a voltage scan rate of 40 mV/s for three different device structures:

- Al/SiO_2/p-Si
- Al/20 LB layers $CdAr_2$/SiO_2/p-Si
- Al/20 LB layers $CdAr_2$/one LB layer Q–Au/p-Si.

The C–V curve for the reference Al/SiO_2/Si sample reveals the usual accumulation/depletion/inversion characteristics associated with metal–insulator–semiconductor (MIS) structures, with a flat-band voltage of approximately −1 V. Negligible hysteresis was evident on reversing the voltage scan. The data for the Si/SiO_2/$CdAr_2$ structure also show clear accumulation, depletion, and inversion regions, again with no hysteresis on reversing the direction of the voltage scan. The most significant difference in the structures with and without the Q-Au nanoparticles is the relatively large hysteresis in the MIS structure containing the Q-Au layer. This was thought to be indicative of charge storage in the gold nanoparticles [51.10]. In a somewhat different approach, the same group has used a self-assembly technique to chemically attach gold nanoparticles to a SiO_2 surface [51.148]. When incorporated into a transistor structure, the resulting device was shown to behave as a nonvolatile electrically erasable programmable read-only memory [51.149]. Alternatively, gold nanoparticles deposited by layer-by-layer assembly using polyions have been applied to the fabrication of nonvolatile memory based on organic transistors [51.150, 151]. Recent progress involves the use of a mixed self-assembled monolayer of aliphatic and C_{60}-functionalized phosphonic acid molecules as a molecular charge storage dielectric layer, which allows low-voltage operation of the organic transistor memory [51.152].

Molecular-scale electronics may also offer increased device densities by fabricating 3-D architectures. In 1989, the principle of a 3-D memory based on LB films was described [51.153]. The device requires a molecule with a central conjugated region of high electron affinity (for an n-type material, the electron affinity is the energy difference between the bottom of the conduction band and the vacuum level) surrounded by aliphatic substituents of low electron affinity, A multilayer structure, as shown in Fig. 51.19, could be used to store one N-bit word, the presence or absence of charge on the n-th layer representing a 0 or 1 of the n-th bit. The LB film could be assembled on the gate of an

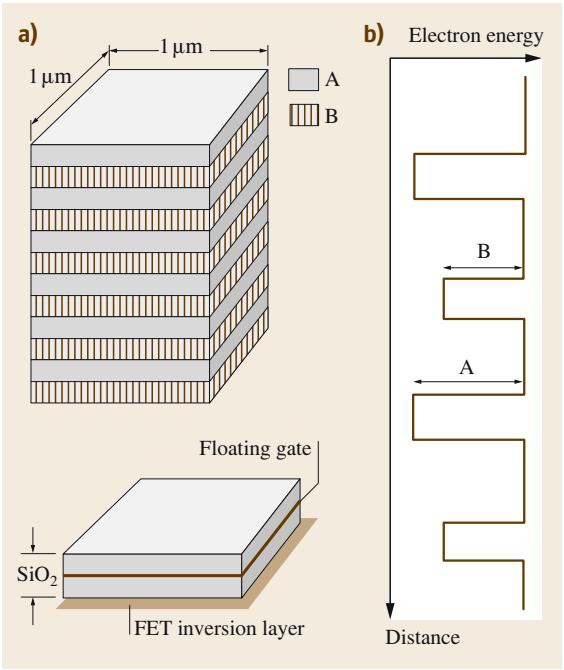


Fig. 51.19 **(a)** Molecular memory holding N bits (*top*) compared to a conventional silicon memory holding 1 bit (*bottom*). **(b)** Electron energy versus perpendicular distance for molecular memory with no applied electric field. (After [51.153], with permission from Elsevier)

FET and, on application of an electric field, transport of bits across the layers may be detected as induced charge on the gate.

Recent technological advancements have allowed the fabrication of structures in which one or few molecules bridge metallic electrodes through chemical bonds, such as SH terminal anchor groups. Such metal–molecule–metal junctions (molecular junctions) [51.154] have been extensively studied as potential building blocks for molecular integrated circuits. One of key technologies is the development of metallic electrodes with a few nanometers gap (nanogap electrode) capable of bridging small molecules. Currently, nanogap electrodes formed by mechanical break and electromigration of a prepatterned notched gold nanowire are most widely used for the fabrication of molecular junctions in planer configurations (these approaches have been referred to as break junction techniques) [51.155, 156]. The gap width of electrodes formed by the mechanical break junction technique can be tuned by bending a flexible substrate under in situ monitoring of tunneling current flowing in the gap, allowing forming a junction of a single molecule with electrodes reproducibly [51.155]. The electromigration break junction technique only provides a statistical for-

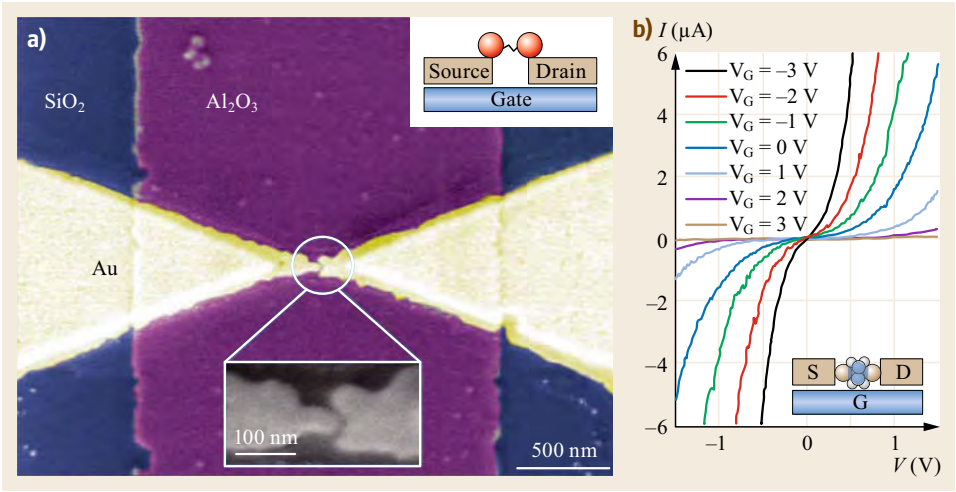


Fig. 51.20 **(a)** Scanning electron micrograph of gold nanogap electrodes formed by the electromigration break junction technique on a gate dielectric layer comprising a 3 nm thick aluminum oxide. (After [51.157], Courtesy of Nature Publishing Group). **(b)** Current versus voltage characteristics of a molecular transistor based on benzene dithiols at different gate voltages. (After [51.158], Courtesy of Nature Publishing Group)

mation of molecule–electrode junctions, but is able to form transistor structures using back-gate electrodes having a thin gate dielectric layer (e.g., Al_2O_3), as shown in Fig. 51.20a [51.157]. A molecular transistor that utilizes the electrostatic gating of molecular orbitals for the modulation of current flow has recently been demonstrated in benzene dithiols covalently bonded to gold electrodes (Fig. 51.20b) [51.158]. The effects of different anchor groups (e.g., NC, NH_3, and COOH), molecular length, molecular conformation and substituents on the electrical properties of molecular junctions have extensively been investigated by the break junction techniques including a scanning tunneling microscope-controlled break junction [51.159] toward the development of molecular devices with various electronic functions.

51.5 DNA Electronics

The study of the electronic behavior of organic compounds has led some scientists to work on the electrical properties of biological materials. DNA is arguably the most significant molecule in nature. It may also be an important material for molecular electronics applications. Reports into the electronic properties of DNA have already generated controversy in the literature [51.16, 160–164]. According to some, DNA is a molecular wire of very small resistance. Others, however, find that DNA behaves as an insulator. These seemingly contradictory findings can probably be explained by the different DNA sequences and experimental conditions used to monitor the conductivity [51.16].

The DNA strand in the double-helix arrangement consists of a long polymer backbone consisting of repeating sugar molecules and phosphate groups (Fig. 51.21). Each sugar group is attached to one of four bases, guanine (G), cytosine (C), adenine (A), and thymine (T). The chemical bonding is such that an A base only ever pairs with a T base, while a G always pairs with a C. Some of the electron orbitals belonging to the bases overlap quite well with each other along the long axis of the DNA. This provides a path for electron transfer along the molecule, in a similar fashion to the 1-D conduction seen in conjugated polymers (Sect. 51.1.2).

Theoretical and experimental work now suggests that a hole (i. e., a positive charge) is more stable on a G–C base pair than on an A–T base pair [51.16]. The energy difference between these two pairs is substantially larger than the thermal energy of the charge carrier. Under these conditions, a hole will localize on a particular G–C base pair. Because the A–T base pairs have a higher energy, they act as a barrier to hole transfer. However, if the distance between two G–C base pairs is small enough, the hole can tunnel quantum-mechanically from one pair to the next. In this way,

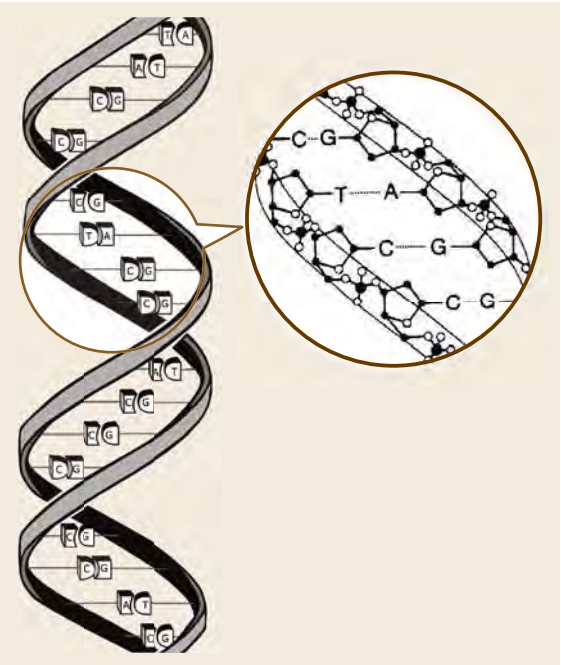


Fig. 51.21 DNA double helix showing the position of the four bases: guanine (G), cytosine (C), adenine (A), and thymine (T)

charge carriers are able to shuttle along a single DNA molecule over a distance of a few nanometers.

Recent experimental and theoretical results suggest that native DNA strands consisting of the specific sequence is a wide band-gap semiconductor [51.165–167]. However, the conductivity of DNA is still controversial, especially in experimental work, because it depends on conditions such as trapped water and counter ions in the DNA strands. Such factors may allow a different channel for conduction and also influence on the geometrical arrangements of base pairs DNA. In addition, modification of DNA with metals and ions can change the electronic behavior. For the realization of DNA electronics, it is imperative to predict the electronic properties theoretically as well as to control the experimental and useable conditions of the DNA.

DNA chips exploit the fact that short strands of DNA will bind to other segments of DNA that have complementary sequences, and can therefore be used to probe whether certain genetic codes are present in a given specimen of DNA. Microfabricated chips with many parallel DNA probes are becoming widespread in analytical and medical applications. Currently, the chips are read out optically, but further miniaturization might require new read-out schemes, possibly involving the electron-transfer properties of DNA.

Computations by chemical or biological reactions overcome the problem of parallelism and interconnections in a classical system. If a string of DNA can be put together in the right sequence, it can be used to solve combinational problems. The calculations are performed in test tubes filled with strands of DNA. Gene sequencing is used to obtain the result. For example, *Adleman* [51.168] calculated the *traveling-salesman problem* to demonstrate the capabilities of DNA computing. DNA computing on parallel problems potentially provides 10^{14} millions of instructions per second (MIPS) and uses less energy and space than conventional supercomputers. While CMOS supercomputers operate 10^9 operations per Joule, a DNA computer could perform about 10^{19} operations per Joule. Data could potentially be stored on DNA in a density of approximately 1 bit/nm^3 while existing storage media such as DRAMs require 10^{12} nm^3 to store 1 bit.

The fact that a DNA strand binds specifically with other strand having a complementary sequence is relevant to the formation of electronic nano-structures. This is so-called DNA origami in a literature. Using DNA strands having various designed sequences, many architectures such as 2-D sheets [51.169], various shapes and patterns [51.170], and 3-D structures [51.171], were formed by self-assembly. These structures are useful as scaffolds for other materials to implement an electronic function [51.172].

The differences in the electronic properties of the nucleotides have been expected to be usable for identifying the sequence of DNA. Since a DNA strand was shown to translocate through nanopores fabricated on solid state films, many efforts have been made to sequence DNA using electronic measurements [51.173–175]. The difference in tunneling current through the nucleotides was measured with small electrodes made by break-junction [51.176], which suggests the possibility of DNA sequencing electronically.

51.6 Conclusions

Organic compounds possess a wide range of fascinating physical and chemical properties that make them attractive candidates for exploitation in electronic and optoelectronic devices. It is not, however, anticipated that these materials will displace silicon in the foreseeable future as the dominant material for fast signal processing. It is much more likely that organic materials will find use in other niche areas of electronics, where silicon and other inorganic semiconductors cannot compete. Examples already exist, such as liquid-crystal displays, photoreceptors in electrophotography, and certain chemical sensors. Organic light-emitting structures are likely to make a major impact in the marketplace as display and lighting devices.

Over the first decades of the twenty-first century, classical CMOS technology will come up against a number of technological barriers. The bottom-up approach to molecular electronics provides an alternative and attractive way forward and, as such, it is currently an area of exciting interdisciplinary activity. However, the challenges in fabricating molecular switches and connecting them together are still formidable. Living systems use a different approach; these assemble themselves naturally from molecules and are extremely energetically efficient when compared with man-made computational devices. More radical approaches to materials fabrication and device design, exploiting self-organization, may be needed to realize fully the potential offered by molecular-scale electronics.

References

51.1 M.C. Petty: *Molecular Electronics: From Principles to Practice* (Wiley, Chichester 2008)
51.2 T.H. Richardson (Ed.): *Functional Organic and Polymeric Materials* (Wiley, Chichester 2000)
51.3 J.M. Tour: *Molecular Electronics* (World Scientific, New Jersey 2003)
51.4 P.E. Kornilovitch, A.M. Bratkovsky, R.S. Williams: Phys. Rev. B **66**, 245413 (2002)
51.5 Y. Chen, G.-Y. Jung, D.A.A. Ohlberg, X. Li, D.R. Stewart, J.O. Jeppesen, K.A. Nielsen, J.F. Stoddart, R.S. Williams: Nanotechnology **14**, 462 (2003)
51.6 Y. Chen, D.A.A. Ohlberg, X. Li, D.R. Stewart, R.S. Williams, J.O. Jeppesen, K.A. Nielsen, J.F. Stoddart, D.L. Olynick, E. Anderson: Appl. Phys. Lett. **82**, 1610 (2003)
51.7 M.R. Diehl, D.W. Steuerman, H.-R. Tseng, S.A. Vignon, A. Star, P.C. Celestre, J.F. Stoddart, J.R. Heath: Chem. Phys. Chem. **4**, 1335 (2003)
51.8 L. Ma, S. Pyo, J. Ouyang, Q. Xu, Y. Yang: Appl. Phys. Lett. **82**, 1419 (2003)
51.9 H. Yu, Y. Luo, K. Beverly, J.F. Stoddart, H.-R. Tseng, J.R. Heath: Angew. Chem. **42**, 5706 (2003)
51.10 S. Paul, C. Pearson, A. Molloy, M.A. Cousins, M. Green, S. Kolliopoulou, P. Dimitrakis, P. Normand, D. Tsoukalas, M.C. Petty: Nano. Lett. **3**, 533 (2003)
51.11 J.M. Tour, L. Cheng, D.P. Nackashi, Y. Yao, A.K. Flatt, S.K. St. Angelo, T.E. Mallouk, P.D. Franzen: J. Am. Chem. Soc. **125**, 13279 (2003)
51.12 L.D. Bozano, B.W. Kean, V.R. Deline, J.R. Salem, J.C. Scott: Appl. Phys. Lett. **84**, 607 (2004)
51.13 J. Ouyang, C.-W. Chu, C.R. Szmanda, L. Ma, Y. Yang: Nature Mater. **3**, 918 (2004)
51.14 R.M. Metzger: J. Solid State Chem. **168**, 696 (2002)
51.15 J. Chen, W. Wang, M.A. Reed, A.M. Rawlett, D.W. Price, J.M. Tour: Appl. Phys. Lett. **77**, 1224 (2000)
51.16 C. Dekker, M.A. Ratner: Phys. World **14**, 29 (2001)
51.17 K.E. Drexler: *Nanosystems: Molecular Machinery, Manufacturing and Computation* (Wiley, New York 1992)
51.18 C. Nicolini (Ed.): *Molecular Manufacturing* (Plenum, New York 1996)
51.19 S. Roth: *One-Dimensional Metals* (VCH, Weinheim 1995)
51.20 M. Ikawa, T. Yamada, H. Matsui, H. Minemawari, J. Tsutsumi, Y. Horii, M. Chikamatsu, R. Azumi, R. Kumai, T. Hasegawa: Nat. Commun. **3**, 1176 (2012)
51.21 H.-R. Tseng, H. Phan, C. Luo, M. Wang, L.A. Perez, S.N. Patel, L. Ying, E.J. Kramer, T.-Q. Nguyen, G.C. Bazan, A.J. Heeger: Adv. Mater. **26**, 2993 (2014)
51.22 C.D. Dimitrakopoulos, D.J. Mascaro: IBM J. Res. Devel. **45**, 11 (2001)
51.23 M.R. Bryce, M.C. Petty: Nature **374**, 771 (1995)
51.24 M. Abe, T. Mori, I. Osaka, K. Sugimoto, K. Takimiya: Chem. Mater. **27**, 5049 (2015)
51.25 H. Minemawari, T. Yamada, H. Matsui, J. Tsutsumi, S. Haas, R. Chiba, R. Kumai, T. Hasegawa: Nature **475**, 364 (2011)
51.26 K.S. Novoselov, V.I. Fal'ko, L. Colombo, P.R. Gellert, M.G. Schwab, K. Kim: Nature **490**, 192 (2012)
51.27 W.J. Feast, J. Tsibouklis, K.L. Pouwer, L. Groenendaal, E.W. Meijer: Polymer **37**, 5017 (1996)
51.28 J.R. Ferraro, J.M. Williams: *Introduction to Synthetic Electrical Conductors* (Academic, Orlando 1987)
51.29 J.D. Wright: *Molecular Crystals* (Cambridge Univ. Press, Cambridge 1995)

51.30 K.S. Novoselov: Nobel Lecture, Graphene: Materials in the Flatland, https://www.nobelprize.org/nobel_prizes/physics/laureates/2010/novoselov_lecture.pdf (2010)
51.31 H.W. Kroto, D.R.M. Walton (Eds.): *The Fullerenes* (Cambridge Univ. Press, Cambridge 1993)
51.32 G. Timp (Ed.): *Nanotechnology* (Springer, Berlin, Heidelberg 1998)
51.33 R. Saito, G. Dresselhaus, M.S. Dresselhaus: *Physical Properties of Carbon Nanotubes* (Imperial College, London 1998)
51.34 L. Duclaux: Carbon **40**, 1751 (2002)
51.35 F. Li, A. Nathan, Y. Wu, B.S. Ong: *Organic Thin Film Transistor Integration: A Hybrid Approach* (Wiley, Weinheim 2011)
51.36 Plastic Logic: A basic display, http://www.plasticlogic.com/ (2016)
51.37 S.R. Forrest: Nature **428**, 911 (2004)
51.38 E.H. Rhoderick: *Metal–Semiconductor Contacts* (Clarendon, Oxford 1978)
51.39 Y.L. Hua, M.C. Petty, G.G. Roberts, M.M. Ahmad, M. Hanack, M. Rein: Thin Solid Films **149**, 161 (1987)
51.40 G. Horowitz: Adv. Mater. **2**, 286 (1990)
51.41 H.E. Katz: J. Mater. Chem. **7**, 369 (1997)
51.42 S. Scheinert, G. Paasch: Phys. Status Solidi (a) **201**, 1263 (2004)
51.43 E.C.P. Smits, S.G.J. Mathijssen, P.A. van Hal, S. Setayesh, T.C.T. Geuns, K.A.H.A. Mutsaers, E. Cantatore, H.J. Wondergem, O. Werzer, R. Resel, M. Kemerink, S. Kirchmeyer, A.M. Muzafarov, S.A. Ponomarenko, B. de Boer, P.W.M. Blom, D.M. de Leeuw: Nature **455**, 956 (2008)
51.44 H. Kawaguchi, T. Someya, T. Sekitani, T. Sakurai: IEEE Solid-State Circuits **40**, 177 (2005)
51.45 M.S. Fuhrer, B.M. Kim, T. Durkop, T. Brintlinger: Nano. Lett. **2**, 755 (2002)
51.46 A. Javey, Q. Wang, A. Ural, Y.M. Li, H.J. Dai: Nano. Lett. **2**, 929 (2002)
51.47 F. Léonard, J. Tersoff: Phys. Rev. Lett. **88**, 258302 (2002)
51.48 K. Tsukagoshi, N. Yoneya, S. Uryu, Y. Aoyagi, A. Kanda, Y. Ootuka, B.W. Alphenaar: Physica B **323**, 107 (2002)
51.49 C.W. Tang, S.A. VanSlyke: Appl. Phys. Lett. **51**, 913 (1987)
51.50 J.H. Burroughes, D.D.C. Bradley, A.R. Brown, R.N. Marks, K. Mackay, R.H. Friend, P.L. Burns, A.B. Holmes: Nature **347**, 359 (1990)
51.51 K. Müllen, U. Scherf (Eds.): *Organic Light Emitting Devices: Synthesis, Properties and Applications* (Wiley, Weinheim 2006)
51.52 T. Tsujimura: *OLED Display Fundamentals and Applications* (Wiley, Hoboken 2012)
51.53 S. Schols: *Device Architecture and Materials for Organic Light-Emitting Devices* (Springer, Dordrecht 2011)
51.54 Cambridge Display Technology: OLED Display Technology, http://www.cdtltd.co.uk/ (2016)
51.55 K. Ghaffarzadeh, N. Bardsley: *OLED Lighting Opportunities 2015–2025: Forecasts, Technologies, Players* (IDTechEx, Cambridge 2015)
51.56 L.S. Hung, C.W. Tang, M.G. Mason: Appl. Phys. Lett. **70**, 152 (1997)
51.57 G.-Y. Jung, C. Pearson, L.E. Horsburgh, I.D.W. Samuel, A.P. Monkman, M.C. Petty: J. Phys. D: Appl. Phys. **33**, 1029 (2000)
51.58 Y. Zhou, C. Fuentes-Hernandez, J. Shim, J. Meyer, A.J. Giordano, H. Li, P. Winget, T. Papadopoulos, H. Cheun, J. Kim, M. Fenoll, A. Dindar, W. Haske, E. Najafabadi, T.M. Khan, H. Sojoudi, S. Barlow, S. Graham, J.-L. Brédas, S.R. Marder, A. Kahn, B. Kippelen: Science **336**, 327 (2012)
51.59 K. Morii, M. Ishida, T. Takashima, T. Shimoda, Q. Wang, M.K. Nazeeruddin, M. Grätzel: Appl. Phys. Lett. **89**, 183510 (2006)
51.60 H. Fukagawa, K. Morii, M. Hasegawa, Y. Arimoto, T. Kamada, T. Shimizu, T. Yamamoto: Appl. Phys. Express **7**, 082104 (2014)
51.61 C. Adachi, S. Tokito, T. Tsutsui, S. Saito: Jpn. J. Appl. Phys. **27**, L269 (1988)
51.62 C. Adachi: Jpn. J. Appl. Phys. **53**, 060101 (2014)
51.63 M. Ikai, S. Tokito, Y. Sakamoto, T. Suzuki, Y. Taga: Appl. Phys. Lett. **79**, 156 (2001)
51.64 C. Adachi, M.A. Baldo, M.E. Thompson, S.R. Forrest: J. Appl. Phys. **90**, 5048 (2001)
51.65 H. Uoyama, K. Goushi, K. Shizu, H. Nomura, C. Adachi: Nature **492**, 234 (2012)
51.66 J. Kido, M. Kimura, K. Nagai: Science **267**, 1332 (1995)
51.67 J. Kido, H. Shionoya, K. Nagai: Appl. Phys. Lett. **67**, 2281 (1995)
51.68 J. Thompson, R.I.R. Blyth, M. Mazzeo, M. Anni, G. Gigli, R. Cingolani: Appl. Phys. Lett. **79**, 560 (2001)
51.69 G.M. Farinola, R. Ragni: Chem. Soc. Rev. **40**, 3467 (2011)
51.70 M.C. Gather, A. Köhnen, K. Meerholz: Adv. Mater. **23**, 233 (2011)
51.71 M.A. Green, K. Emery, Y. Hishikawa, W. Warta, E.D. Dunlop: Prog. Photovolt: Res. Appl. **24**, 3 (2016)
51.72 C.W. Tang: Appl. Phys. Lett. **48**, 183 (1986)
51.73 V. Vohra, K. Kawashima, T. Kakara, T. Koganezawa, I. Osaka, K. Takimiya, H. Murata: Nat. Photonics **9**, 403 (2015)
51.74 M. Saito, I. Osaka, Y. Suzuki, K. Takimiya, T. Okabe, S. Ikeda, T. Asano: Sci. Rep. **5**, 14202 (2015)
51.75 L. Huo, T. Liu, X. Sun, Y. Cai, A.J. Heeger, Y. Sun: Adv. Mat. **27**, 2938 (2015)
51.76 J. Subbiah, B. Purushothaman, M. Chen, T. Qin, M. Gao, D. Vak, F.H. Scholes, X. Chen, S.E. Watkins, G.J. Wilson, A.B. Holmes, W.W.H. Wong, D.J. Jones: Adv. Mat. **27**, 702 (2015)
51.77 S. Günes, H. Neugebauer, N.S. Sariciftci: Chem. Rev. **107**, 1324 (2007)
51.78 C.J. Brabec, A. Cravino, D. Meissner, N.S. Sariciftci, T. Fromherz, M. Minse, L. Sanchez, J.C. Hummelen: Adv. Funct. Mater. **11**, 374 (2001)
51.79 M. Scharber, D. Mühlbacher, M. Koppe, P. Denk, C. Waldauf, A.J. Heeger, C. Brabec: Adv. Mater. **18**, 789 (2006)
51.80 M. Hiramoto, H. Fujiwara, M. Yokoyama: Appl. Phys. Lett. **58**, 1062 (1991)

51.81 S. Morita, A.A. Zakhidov, K. Yoshino: Solid State Commun. **82**, 2084 (1992)
51.82 G. Yu, J. Gao, J.C. Hummelen, F. Wudl, A.J. Heeger: Science **270**, 1789 (1995)
51.83 B. O'Regan, M. Grätzel: Nature **353**, 737 (1991)
51.84 S. Zhang, X. Yang, Y. Numata, L. Han: Environ. Sci. **6**, 1443 (2013)
51.85 Semiconductor Industry Association: International Technology Roadmap for Semiconductors (ITRS), http://www.itrs2.net/ (2011)
51.86 D.J. Frank, R.H. Dennard, E. Nowak, P.M. Solomon, Y. Taur, H.-S.P. Wong: Proc. IEEE **89**, 259 (2001)
51.87 J. Greer, A. Korkin, J. Lanbanowski (Eds.): *Nano and Giga Challenges in Microelectronics* (Elsevier, Amsterdam 2003)
51.88 IC Knowledge LLC: The Economics of the ITRS, http://www.icknowledge.com/ (2011)
51.89 M.C. Petty: Organic thin-film deposition techniques. In: *Functional Organic and Polymeric Materials*, ed. by T.H. Richardson (Wiley, Chichester 2000) p. 7
51.90 A. Ulman: *Ultrathin Organic Films* (Academic, San Diego 1991)
51.91 M.C. Petty: *Langmuir–Blodgett Films* (Cambridge Univ. Press, Cambridge 1996)
51.92 D.R. Talham: Chem. Rev. **104**, 5479 (2004)
51.93 J.C. Love, L.A. Estroff, J.K. Kriebel, R.G. Nuzzo, G.M. Whitesides: Chem. Rev. **105**, 1103 (2005)
51.94 A.J. Moore, L.M. Goldenberg, M.R. Bryce, M.C. Petty, A.P. Monkman, C. Marenco, J. Yarwood, M.J. Joyce, S.N. Port: Adv. Mater. **10**, 395 (1998)
51.95 L. Netzer, J. Sagiv: J. Am. Chem. Soc. **105**, 674 (1983)
51.96 L. Netzer, R. Iscovici, J. Sagiv: Thin Solid Films **99**, 235 (1983)
51.97 G. Decher, J.D. Hong, J. Schmitt: Thin Solid Films **210/211**, 831 (1992)
51.98 G. Decher, Y. Lvov, J. Schmitt: Thin Solid Films **244**, 772 (1994)
51.99 G. Decher, J.B. Schlenoff (Eds.): *Multilayer Thin Films* (Wiley, Weinheim 2003)
51.100 K. Ariga, J.P. Hill, Q. Ji: Phys. Chem. Chem. Phys. **9**, 2319 (2007)
51.101 C. Pearson, J. Nagel, M.C. Petty: J. Phys. D: Appl. Phys. **34**, 285 (2001)
51.102 J.H. Cheung, W.B. Stockton, M.F. Rubner: Macromolecules **30**, 2712 (1997)
51.103 Y. Lvov, M. Onda, K. Ariga, T. Kunitake: J. Biomater. Sci. Polym. Edn. **9**, 345 (1998)
51.104 W.B. Stockton, M.F. Rubner: Macromolecules **30**, 2717 (1997)
51.105 S.A. Evanson, J.P.S. Badyal, C. Pearson, M.C. Petty: J. Phys. Chem. **100**, 11672 (1996)
51.106 S. Brittain, K. Paul, X.-M. Zhao, G. Whitesides: Phys. World **11**, 31 (1998)
51.107 R.D. Piner, J. Zhu, F. Xu, S. Hong, C.A. Mirkin: Science **283**, 661 (1999)
51.108 L.F. Chi, L.M. Eng, K. Graf, H. Fuchs: Langmuir **8**, 2255 (1992)
51.109 A. Krämer, R.R. Fuierer, C.B. Gorman: Chem. Rev. **103**, 4367 (2003)
51.110 J. Gimzewski: Phys. World **11**, 25 (1998)
51.111 S. Hong, J. Zhu, C.A. Mirkin: Science **286**, 523 (1999)
51.112 S. Hong, C.A. Mirkin: Science **288**, 1808 (2000)
51.113 K. Salaita, Y. Wang, J. Fragala, R.A. Vega, C. Liu, C.A. Mirkin: Angew. Chem. Int. Ed. **45**, 7220 (2006)
51.114 F. Huo, Z. Zheng, G. Zheng, L.R. Giam, H. Zhang, C.A. Mirkin: Science **321**, 1658 (2008)
51.115 X.G. Liu, L. Fu, S.H. Hong, V.P. Dravid, C.A. Mirkin: Adv. Mater. **14**, 231 (2002)
51.116 K.B. Lee, S.J. Park, C.A. Mirkin, J.C. Smith, M. Mrksich: Science **295**, 1702 (2002)
51.117 S.P. Speakman, G.G. Rozenberg, K.J. Clay, W.I. Milne, A. Ille, I.A. Gardner, E. Bresler, J.H.G. Steinke: Organic Electron. **2**, 65 (2001)
51.118 G. PerHin, B.T. Khuri-Yakub: Rev. Sci. Instrum. **73**, 2193 (2002)
51.119 H. Sirringhaus, T. Shimoda (Eds.): MRS Bull. **28**, 802 (2003)
51.120 M. Singh, H.M. Haverinen, P. Dhagat, G.E. Jabbour: Adv. Mater. **22**, 673 (2010)
51.121 I.D. Rees, K.L. Robinson, A.B. Holmes, C.R. Towns, R. O'Dell: MRS Bull. **27**, 451 (2002)
51.122 H. Sirringhaus, T. Kawase, R.H. Friend, T. Shimoda, M. Inbasekaran, W. Wu, E.P. Woo: Science **290**, 2123 (2000)
51.123 A.C. Arias, S.E. Ready, R. Lujan, W.S. Wong, K.E. Paul, A. Salleo, M.L. Chabinyc, R. Apte, R.A. Street, Y. Wu, P. Liu, B. Ong: Appl. Phys. Lett. **85**, 3304 (2004)
51.124 T. Sekitani, Y. Noguchi, U. Zschieschang, H. Klauk, T. Someya: Proc. Natl. Acad. Sci. USA **105**, 4976 (2008)
51.125 M. Cavallini, F. Biscarini, J. Gomez-Segura, D. Ruiz, J. Veciana: Nano Lett. **3**, 1527 (2003)
51.126 M. Böltau, S. Walheim, J. Mlynek, G. Krausch, U. Steiner: Nature **391**, 877 (1998)
51.127 R. van Hameren, P. Schön, A.M. van Buul, J. Hoogboom, S.V. Lazarenko, J.W. Gerritsen, H. Engelkamp, P.C.M. Christianen, H.A. Heus, J.C. Maan, T. Rasing, S. Speller, A.E. Rowan, J.A.A.W. Elemans, R.J.M. Nolte: Science **314**, 1433 (2006)
51.128 E. Kim, Y. Xia, G.M. Whitesides: Nature **376**, 581 (1995)
51.129 F. Zhang, T. Nyberg, O. Inganäs: Nano Lett. **2**, 1373 (2002)
51.130 R.S. Kane, S. Takayama, E. Ostuni, D.E. Ingber, G.M. Whitesides: Biomaterials **20**, 2363 (1999)
51.131 E. Delamarche, D. Juncker, H. Schmid: Adv. Mater. **17**, 2911 (2005)
51.132 M. Cavallini, F. Biscarini: Nano Lett. **3**, 1269 (2003)
51.133 C.A. Jones, M.C. Petty, G.G. Roberts: IEEE Trans. Ultrasonics Ferroelec. Freq. Control **35**, 736 (1988)
51.134 P.J. Skabara: Organic second-order non-linear optical materials. In: *Functional Organic and Polymeric Materials*, ed. by T.H. Richardson (Wiley, Chichester 2000) p. 295
51.135 A. Aviram, M.A. Ratner: Chem. Phys. Lett. **29**, 277 (1974)
51.136 A.S. Martin, J.R. Sambles, G.J. Ashwell: Phys. Rev. Lett. **70**, 218 (1993)
51.137 G.J. Ashwell, D.S. Gandolfo: J. Mater. Chem. **11**, 246 (2001)

51.138 J.R. Heath, P.J. Kuekes, G.S. Snider, R.S. Williams: Science **280**, 1716 (1998)
51.139 A.S. Blum, J.G. Kushmerick, D.P. Long, C.H. Patterson, J.C. Yang, J.C. Henderson, Y.X. Yao, J.M. Tour, R. Shashidhar, B.R. Ratna: Nat. Mater. **4**, 167 (2005)
51.140 E. Lörtscher, J.W. Ciszek, J. Tour, H. Riel: Small **2**, 973 (2006)
51.141 V. Meded, A. Bagrets, A. Arnold, F. Evers: Small **5**, 2218 (2009)
51.142 H.B. Akkerman, P.W.M. Blom, D.M. de Leeuw, B. de Boer: Nature **441**, 69 (2006)
51.143 G. Wang, Y. Kim, M. Choe, T.-W. Kim, T. Lee: Adv. Mater. **23**, 755 (2011)
51.144 A.J. Kronemeijer, H.B. Akkerman, T. Kudernac, B.J. van Wees, B.L. Feringa, P.W.M. Blom, B. de Boer: Adv. Mater. **20**, 1467 (2008)
51.145 J. Lee, H. Chang, S. Kim, G.S. Bang, H. Lee: Angew. Chem. Int. Ed. **48**, 8501 (2009)
51.146 B.G. Streetman, S. Banjeree: *Solid State Electronic Devices* (Prentice Hall, New Jersey 2000)
51.147 R. Compañó (Ed.): *Technology Roadmap for Nanoelectronics* (Office for Official Publications of the European Communities, Luxembourg 2001)
51.148 S. Kolliopoulou, P. Dimitrakis, P. Normand, H.-L. Zhang, N. Cant, S.D. Evans, S. Paul, C. Pearson, A. Molloy, M.C. Petty: J. Appl. Phys. **94**, 5234 (2003)
51.149 S. Kolliopoulou, P. Dimitrakis, P. Normand, H.-L. Zhang, N. Cant, S.D. Evans, S. Paul, C. Pearson, A. Molloy, M.C. Petty, D. Tsoukalas: Micro. Eng. **73/74**, 725 (2004)
51.150 Z. Liu, F. Xue, Y. Su, Y.M. Lvov, K. Varahramyan: IEEE Trans. Nanotech. **5**, 379 (2006)
51.151 S.-J. Kim, Y.-S. Park, S.-H. Lyu, J.-S. Lee: Appl. Phys. Lett. **96**, 033302 (2010)
51.152 M. Burkhardt, A. Jedaa, M. Novak, A. Ebel, K. Voitchovsky, F. Stellacci, A. Hirsch, M. Halik: Adv. Mater. **22**, 2525 (2010)
51.153 P.E. Burrows, K.J. Donovan, E.G. Wilson: Thin Solid Films **179**, 129 (1989)
51.154 H. Song, M.A. Reed, T. Lee: Adv. Mater. **23**, 1583 (2011)
51.155 M.A. Reed, C. Zhou, C.J. Muller, T.P. Burgin, J.M. Tour: Science **278**, 252 (1997)
51.156 H. Park, A.K.L. Lim, A.P. Alivisatos, J. Park, P.L. McEuen: Appl. Phys. Lett. **75**, 301 (1999)
51.157 W.J. Liang, M.P. Shores, M. Bockrath, J.R. Long, H. Park: Nature **417**, 725 (2002)
51.158 H. Song, Y. Kim, Y.H. Jang, H. Jeong, M.A. Reed, T. Lee: Nature **462**, 1039 (2009)
51.159 B. Xu, N.J. Tao: Science **301**, 1221 (2003)
51.160 H.-W. Fink, C. Schöneberger: Nature **398**, 407 (1999)
51.161 R.S. Phadke: Appl. Biochem. Biotechnol. **96**, 269 (2001)
51.162 A. Rakitin, P. Aich, C. Papadopoulos, Y. Kobzar, A.S. Vedeneev, J.S. Lee, J.M. Xu: Phys. Rev. Lett. **86**, 3670 (2001)
51.163 C.N.R. Rao, A.K. Cheetham: J. Mater. Chem. **11**, 2887 (2001)
51.164 H. Tabata, L.T. Cai, J.H. Gu, S. Tanaka, Y. Otsuka, Y. Sacho, M. Taniguchi, T. Kawai: Synth. Met. **133/134**, 469 (2003)
51.165 R.G. Endres, D.L. Cox, R.R.P. Singh: Rev. Mod. Phys. **76**, 195 (2004)
51.166 M. Taniguchi, T. Kawai: Physica E **33**, 1 (2006)
51.167 S.S. Mallajosyula, S.K. Pati: J. Phys. Chem. Lett. **1**, 1881 (2010)
51.168 L. Adleman: Science **266**, 1021 (1994)
51.169 E. Winfree, F. Liu, L.A. Wenzler, N.C. Seeman: Nature **394**, 539 (1998)
51.170 P.W.K. Rothemund: Nature **440**, 297 (2006)
51.171 W.M. Shih, J.D. Quispe, G.F. Joyce: Nature **427**, 618 (2004)
51.172 H. Yan, S.H. Park, G. Finkelstein, J.H. Reif, T.H. LaBean: Science **26**, 1882 (2003)
51.173 J. Li, D. Stein, C. McMullan, D. Branton, J.A. Golovchenko: Nature **412**, 166 (2001)
51.174 P. Chen, J. Gu, E. Brandin, Y.-R. Kim, Q. Wang, D. Branton: Nano Lett. **4**, 2293 (2004)
51.175 M. Zwolak, M.D. Ventra: Rev. Mod. Phys. **80**, 141 (2008)
51.176 M. Tsutsui, M. Taniguchi, K. Yokota, T. Kawai: Nat. Nanotech. **5**, 286 (2010)

52. Organic Materials for Chemical Sensing

Asim K. Ray

Organic materials for chemical sensing are broadly classified into three categories: (i) macrocyclic compounds, (ii) conducting polymers, and (iii) cavitand molecules. A short review of current progress in inorganic sensing materials including graphene is given, pointing out their strengths and limitations. Principal wet techniques for depositing organic thin films are described and electrical, optical, and structural properties of all three types of organic materials are analyzed in relation to their importance in chemical sensing. Examples of recent advances in chemical sensing of different analytes and pollutants are presented.

52.1 **Analyte Requirements** 1282
52.2 **Brief Review of Inorganic Materials** 1283
52.2.1 Oxide Semiconductor Sensors 1283
52.2.2 Graphene Based Sensors 1285
52.3 **Macrocylic Compounds for Sensing** 1286
52.3.1 Preparation of Sensing Membranes 1286
52.3.2 Thin-Film Properties 1289
52.4 **Sensing with Phthalocyanine and Porphyrin** 1291
52.4.1 Amperometric Sensor 1291
52.4.2 Optical Sensors 1292
52.4.3 Detection of Volatile Organic Vapor Compounds 1295
52.4.4 Phthalocyanine for Biosensing 1296
52.5 **Polymeric Materials** 1297
52.5.1 Conducting Polymers 1298
52.5.2 Ion Sensing 1299
52.5.3 Examples of Other Polymeric Sensors ... 1300
52.6 **Cavitand Molecules** 1303
52.7 **Concluding Remarks** 1305
References 1305

The development of a new generation of sensors involves the study of fine-tuned sensor-active materials and transducers [52.1, 2]. Metal oxides, polymer/polymer composites, and dyes are regarded as key sensing materials. This chapter is organized into eight sections. The next section presents a summary of the needs for chemical sensing of different pollutants. The second section provides a brief description of different semiconducting oxide materials and their applications in sensor technology. Organized ultrathin organic films of molecular and polymeric semiconductors are currently the focus of considerable research with an emphasis on a molecular understanding of the sensing mechanisms. The low cost of fabrication, large-area processability, and widely diverse and tailorable physical, electronic, and optical properties have stimulated interest in their practical exploitation. Phthalocyanine and related macrocylic compounds have been extensively investigated for chemical sensing primarily because the desired functionality of phthalocyanine (Pc) molecules can be achieved by changing the central metal atom or by introducing substituents in their π-electron aromatic ring [52.3]. The third section is devoted to thin-film formulation of these interesting compounds and their structural, electrical, and optical properties relevant to sensing different types of pollutants. Important transduction methods and their applications in the development of practical sensors are presented in the fourth section, with an emphasis on the interaction at the analyte/film interface of macrocylic compounds. The ability of conducting polymers to change their physical properties during reaction with various redox agents makes them very useful for the development of gas sensors. The most common are polyaniline (PAni) and polypyrrole (PPY). It has been shown that the resistivity of polypyrrole increases in the presence of a reducing gas such as ammonia, but decreases in the presence of an oxidizing gas such as nitrogen dioxide. The

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_52

gases cause a change in the near-surface charge-carrier density by reacting with surface-adsorbed oxygen ions [52.4]. Sensing properties of polymers and their importance in the development of intelligent sensors are described in Sect. 52.5. A class of supramolecules, called cavitands, that exhibit discotic phases depending on the structure, are shown to have potential for sensing applications [52.5]. Section 52.6 deals with the applications of thin films of these molecules such as calixresorcinarene for recognition of organic vapors. The chapter concludes by highlighting future development trends for sensor materials and technology in Sect. 52.7.

52.1 Analyte Requirements

The detection of pollutant gases and volatile organic compounds (VOC) whether in gaseous phase or dissolved in water is of great environmental importance due to the extreme hazards posed by their presence in small amounts in the ambient. Nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), particulate matter (PM), carbon monoxide (CO), and lead (Pb) are six principal pollutants of major concern for maintaining the quality of air in urban areas (Table 52.1). Some of these pollutants, like CO, Pb, NO_2, PM, and SO_2, are emitted directly from a variety of sources including industrial production growth and increased traffic congestion. Ozone is generally formed when emissions of NO_2, SO_2, ammonia (NH_3), VOCs, and other gases react in the atmosphere. The presence of ozone in high concentration is also possible in offices where laser printers and photocopiers are heavily used. Conversely, water is commonly treated with O_3 to destroy microbial contaminants such as *Escherichia coli* and *Salmonella* [52.6, 7] and there is a need for accurate calibration of ozone content. Because of its ability to absorb ultraviolet (UV) radiation from sunlight, stratospheric ozone filters out damaging ultraviolet radiation from the sun. Monitoring of ozone in the stratosphere has thus become a global issue as it forms a protective shield for all forms of life on Earth.

Stringent domestic and international standards for emission control require the monitoring of threshold limit values of these pollutants in order to protect primarily against adverse public health hazards. For example, VOCs and nitric oxide (NO_x) resulting from combustion processes cause respiratory problems. Benzene itself is regarded as a human carcinogen and vehicle exhaust is primarily composed of these toxic gases. Spillage in fuel stations also contribute to the emission of excessive amounts of aromatic hydrocarbons in the atmosphere. VOCs also bind with ground-level ozone and form a photochemical smog, causing breathing difficulty for asthma sufferers. The secondary requirement of pollution monitoring is to reduce disastrous effects on the preservation of ecosystems, plants, and animals. Safeguards should also be in place to protect from decreased visibility due to the presence of PM and damage to crops, vegetation, and buildings due to atmospheric pollution. SO_2 content is known to attack buildings both internally and externally, and in particular the surface of historical glasses. A wide variety of preventive conservation tools and routines have been sought to protect historical and cultural objects against deterioration and degradation [52.8, 9].

The activities of chemical sensing are not solely limited to environmental control. Hydrogen (H_2) fuel cells are clean, quiet, more efficient generators of electricity than any other known technology. Hydrogen is also an important raw material for the aerospace, chemical, and semiconductor industries. As hydrogen is explosive above the lower explosive limit (LEL), H_2 sensors have now become important safety devices in all these applications. The oxygen sensor which detects the air-fuel mixture of a gasoline engine by measur-

Table 52.1 Air quality standards for the UK, USA, and World Health Organization (WHO)

Pollutant	UK		WHO		USA	
	Concentration	Standard measured	Concentration	Standard measured	Concentration	Standard measured
Benzene	5 ppb	Annual mean	Data not available			
1,3-Butadiene	1 ppb	Annual mean	Data not available			
Carbon monoxide (CO)	8.6 ppm	8 h mean	8.6 ppm	8 h mean	9 ppm	8 h mean
Lead (Pb)	0.25 $\mu g/m^3$	Annual mean	0.5 $\mu g/m^3$	Annual mean	1.5 $\mu g/m^3$	Quarterly mean
Nitrogen dioxide (NO_2)	21 ppb	Annual mean	105 ppb	1 h mean	0.05 ppm	Annual mean
Ozone (O_3)	50 ppb	8 h mean	60 ppb	8 h mean	0.08 ppm	8 h mean
Particles (PM_{10})	50 $\mu g/m^3$	24 h mean	70 $\mu g/m^3$	24 h mean	50 $\mu g/m^3$	Annual mean
Sulfur dioxide (SO_2)	100 ppb	15 min mean	188 ppb	10 min mean	0.14 ppm	24 h mean

ing the amount of oxygen in the exhaust gas is an essential component of the car engine management system. The development of reliable oxygen sensors for medical application has received considerable research attention in recent years [52.10–12]. Drinking, industrial, and swimming pool water must be disinfected with suitable oxidizing agents such as chlorine or chlorine compounds. The dosing of the appropriate oxidizing agents must be carefully controlled to suit the application as a higher concentration can result in corrosion effects, impairment of taste, or skin irritation. The ability to detect chlorine at sub-ppm level is therefore essential [52.13]. Ammonia is increasingly being used as a refrigerant alternative to ozone-depleting chlorofluorocarbon (CFC). Large quantities of ammonia can also be found in fertilizer factories, resin-production plants using urea, explosive munitions plants, semiconductor industries, and water utility facilities. The presence of ammonia can cause inhalation problems at a very low concentration (0.6–53 ppm). Airborne ammonia gas dissolves in the moisture on the skin, forming corrosive ammonium hydroxide. Exposure to ammonia at 10 000 ppm is mildly irritating to moist skin but the effects become more pronounced with increasing concentration, producing chemical burns with blistering at 30 000 ppm. Its solubility in water can cause it to cauterize respiratory tracts, resulting in deaths at concentrations of 5000 ppm. Research into the development of reliable ammonia sensors with a fast response to a wide range of concentrations is intensive [52.14–16].

It is obvious from the brief survey above that growing demand exists for the fabrication of low-cost, low-power, robust, and portable chemical sensors for a variety of pollutants, contaminants, and analytes, showing improved sensitivity and selectivity for industrial, health care, and environmental control. Analytical chemical tools provide advanced and reliable methods for detection of different analytes but those methods are in many cases prohibitively expensive, unsuited for large-scale field measurements, or mobile monitoring. For instance, there is currently a good market for detection systems for ozone at low concentration (ppb level) to replace more expensive techniques such as UV absorption measurements [52.17]. Existing devices for the detection of ozone are now mostly based on electrochemical measurements [52.18], or on solid-state semiconductor devices [52.19]. Similarly, the detection of organic vapors involves extracting air samples and subsequently analyzing them using standard laboratory techniques such as gas chromatography [52.20], mass spectroscopy, and Fourier-transform infrared (FTIR) spectroscopy [52.21, 22]. The major disadvantages associated with these techniques are their complexity, the capital cost of the instruments used, and the requirement for trained human resources for both operation and the interpretation of data. Apart from environmental applications, the development of detection instrumentation for these species is also important for both human health monitoring [52.23, 24] and for odor detection [52.25].

52.2 Brief Review of Inorganic Materials

Different types of oxide semiconductors have been used for fabrication of gas sensors with a high degree of sensitivity and specificity. The physically interesting electronic and opto-electronic properties of 2D materials have also been exploited over last two decades for environmental and security control.

52.2.1 Oxide Semiconductor Sensors

Solid-state semiconductor gas sensors are usually made of tin oxide (SnO_2) in the form of thick films, porous pellets, or thin-film coatings to monitor the presence of various gases [52.26]. These sensors are being used in the intelligent automatic control of a large number of processes, ranging from microwave cooking to the efficient combustion of motor engines. SnO_2 sensors operate at high temperatures of 300–400 °C. Gases are adsorbed onto the surface in the presence of air and the semiconductivity of the metal oxide changes due to the formation of a surface depletion layer. The thickness of this layer can be expressed in terms of the Debye length $L_D = \sqrt{\varepsilon_0 \varepsilon k_B T/(q^2 n_0)}$, where ε_0 is the permittivity of free space, ε is the dielectric constant of the semiconducting material, n_0 is the total carrier concentration, q is the electronic charge, k_B is the Boltzmann constant, and T is the absolute temperature [52.27]. The sensitivity of the sensor is generally defined as the ratio of the change in conductance after treatment to the conductance of the sensor in air and can simply be written as $\Delta n L_D/n_0$, where Δn is the change in carrier concentration due to analyte exposure. It is therefore possible to optimize the sensitivity by reducing grain size, lowering the carrier concentration, and increasing the Debye length. The maximum sensitivity is assumed to have been achieved when the Debye length is about half the particle size.

Figure 52.1 shows a photograph of a comb-like interdigitated electrode (IDE) system used for most

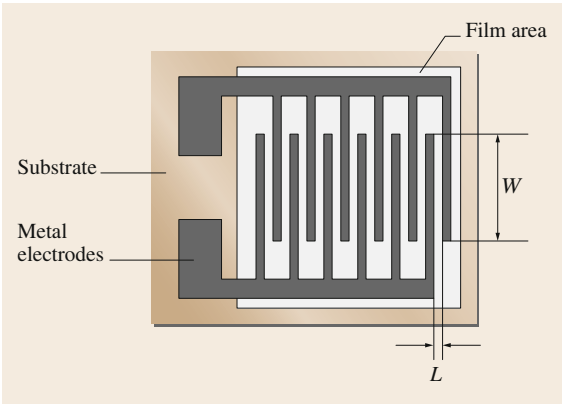


Fig. 52.1 Interdigitated electrode (IDE) system for an SnO_2 sensor

conductivity measurements. The original idea was developed by *Taguchi* [52.28] and the sample resistance in this configuration is smaller than that obtained from the four-contact geometry. The output signal is normally logarithmic to gas concentration. The conductivity change of SnO_2 thick-film sensors was monitored on exposure to compounds forming in smouldering and combustion processes of natural materials, e.g., paper and wood. A strong response was observed for compounds with hydroxyl groups and the highest response was obtained for phenolic structures such as methoxyphenols and benzenediols [52.29]. The response of a SnO_2 sensor element operating at 270 °C produced no sensor response to toluene, furfural, and acetone [52.30]. The impedance output from the SnO_2 sensor array was measured at 300 °C in terms of conductance changes and peak frequency changes. These changes were more sensitive to oxygenated VOCs than other VOCs. The sensor's sensitivity could also be enhanced by using Pd- and Pt-doped sensors [52.31].

Considerable development work is being carried out to integrate the semiconductor sensing membrane with the interrogation and control circuitry on a silicon chip using conventional silicon technology. Using an industrial 0.8 μm complementary metal-oxide-semiconductor (CMOS) process with subsequent micromachining steps, a CO gas-sensor chip was fully integrated with advanced analogue and digital circuitry on a chip for temperature control in the range 170–300 °C with an error of ±2 °C. The whole device was encapsulated using standard microelectronic packaging methods. It was reported that the integrated sensor was capable of measuring CO at the concentration of 5 ppm with a resolution of ±0.1 ppm [52.32].

Solid-state sensor technology relies on the preparation of semiconductor films. SnO_2 sensors suffer from a lack of selectivity and drift. The high temperature required for the surface reactions encourages the growth of grains by coalescence, giving rise to unstable materials. Recent progress in the synthesis of nanomaterials has been employed to overcome some of these problems. Oxide powders were thermally evaporated under controlled conditions without the presence of a catalyst to deposit single-crystalline SnO_2 nanobelts on a platinum IDE structure on an alumina substrate. The surface-to-volume ratio was very high but the size was large enough to contain the depletion layer within the belt. SnO_2 nanobelts were sensitive to CO, NO_2, and ethanol and therefore suitable for applications in breath analysis and food control [52.33].

Advanced methods of information technology are increasingly being used to address the issues regarding selectivity and specificity. A sensor array of two commercial metal-oxide gas sensors was developed to discriminate six organic solvents over a wide concentration range of 2–200 ppm in air. The baseline drift was considerably reduced over a test period of several months by temperature cycling. Further reduction in drift was achieved by suppressing the influence of humidity via additional signal preprocessing. Relatively low computing power was required for the hierarchical pattern classification to evaluate shape features generated from the sensor response curve during temperature cycling. The technique was found to be adaptable to different operating environments, for example, to suppress false alarms from interfering gases [52.34].

Apart from SnO_2, titania (TiO_2), ferric oxide (Fe_2O_3), titania-doped chromium oxide ($Cr_{1.8}Ti_{0.2}O_3$), molybdenum trioxide (MoO_3), tungsten oxide (WO_3), indium oxide (In_2O_3), and perovskite-type thick films ($LaFeO_3$) were studied for sensing a host of analytes such as H_2, CO, NO_2, H_2S, NH_3, and CH_4 in the ppb range. The characteristics of titanium dioxide nanotube hydrogen sensors were found to be completely reversible, highly selective, and free from hysteresis. The sensors showed response times of approximately 150 s at a nominal operating temperature of 180 °C. The sensitivity of the nanotubes was believed to be due to chemisorption of hydrogen onto the titania surface as electron donors. Increased operating temperature improved the sensitivity significantly but reduced the response time. Samples with smaller pore diameter (46 nm) were more sensitive to hydrogen than those with larger pore diameter (76 nm). The sensor showed a response to high concentrations of oxygen but the response time was long and the recovery was not complete [52.35]. Ion-beam-deposited thin films of MoO_3 on alumina substrates with gold IDEs showed selective response to NH_3 at 450 °C. Similarly prepared WO_3 thin-film samples were highly sensitive to NO_2. Both films exhibited orthorhombic structure but the polymor-

phic dissimilarities between two types of materials were responsible for the difference in their sensing behavior [52.36].

52.2.2 Graphene Based Sensors

Successful mechanical exfoliation of a highly oriented pyrolytic graphite sample to produce thin layers of graphene was reported in 2004. A one-atom-thick single layer of graphene consists of sp^2-hybridized carbon atoms arranged in a hexagonal lattice. The overlap between the conduction and valence bands is found to occur in ambilar monocrystalline thin films showing a large carrier concentration of $10^{17}\,cm^{-2}$ and high mobility up to $10\,000\,cm^2V^{-1}s^{-1}$ [52.37]. Mechanical exfoliation, sonication-assisted liquid exfoliation, and chemical vapor deposition (CVD) are commonly used to synthesize high-quality graphene films with a reduced number of defects by annealing. However, these methods are not suitable for deposition over large areas. Printing can be used as an alternative technique to make conductive graphene films. Some of the advantages of graphene printing are the low processing temperature under ambient conditions without using a vacuum or inert atmosphere. The other benefit of graphene printing is that films can be obtained over large areas on flexible substrates, at a reduced cost. Figure 52.2 shows the scanning electron microscopic (SEM) image of a drop-cast film of graphene. The graphene ink was formulated with few-layer graphene particles of lateral dimensions $< 40\,\mu m$, which were obtained by liquid-phase exfoliation. The graphene particles were dispersed in an acrylic polymer system to obtain the ink suitable for thin-film deposition on flexible substrates. The sample consisted of a disordered layered structure with a single isolated graphene flake as shown in the inset.

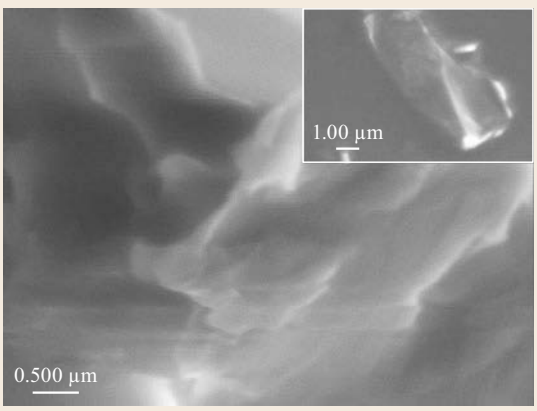


Fig. 52.2 Scanning electron microscopic image of a graphene ink sample drop cast onto aluminum foil

Considerable research efforts have been spent over the last two years to exploit the unique optoelectronic properties of these highly π-conjugated two-dimensional (2-D) materials for the development of a variety of sensors with potential applications in security, environmental and safety control, and disease detection and diagnosis [52.38]. Intrinsic graphene is found to be highly sensitive in detecting industrial toxic gases such as CO, NO_x, and NH_3, but selectivity is poor for practical applications. Considerable improvement in selectivity can be achieved using hybrid structures of graphene blended with metal oxide nanoparticles [52.39]. The electrochemical potential window of 2.5 V for the graphene electrode in 0.1 M phosphate-buffered saline with a pH value of 7.0 is very comparable to those obtained for glassy carbon (GC) and boron-doped diamond. Also, AC impedance measured graphene charge-transfer resistance is smaller than their competitive electrodes. These superior electrochemical properties of graphene have been exploited in recent years to develop the oxidase sensors for detecting the presence of hydrogen peroxide H_2O_2 in food, pharmaceutical, clinical, industrial and environmental analyses with a high degree of reliability and sensitivity [52.40]. The presence of heavy metals such as lead (Pb^{2+}), cadmium (Cd^{2+}), chromium (Cr^{3+}, Cr^{6+}), mercury (Hg^{2+}), copper (Cu^{2+}), and arsenic (As^{3+}) ions in atmosphere and water is recognized by the World Health Organization as a severe threat to human health. Field-effect transistors (FET) using protein-functionalized reduced graphene oxide (RGO) films as an active channel have been employed to record the change in transconductance due to the contamination of target metal atoms. A value of $2.5\times10^{-8}\,mol\,l^{-1}$ for the detection limit and a faster response time of less than 10 s were observed in Hg^{2+} monitoring with thermally reduced GO film decorated with thioglycolic-acid-functionalized gold nanoparticles (AuNps) in an active FET channel. These FET sensors are portable and cost-effective compared to turn-on fluorescence sensors based upon fluorescence resonance energy transfer. It is, however, reported that fluorescence based upon hybrid graphene-gold nanoparticles (AuNp) is able to detect Pb^{2+} ions selectively up to the limit of $10^{-8}\,mol\,L^{-1}$ over the common metal pollutants in an aqueous solution. This effect is attributed to accelerated leaching of AuNp by Pb^{2+} ions. The highly hydrophobic surface with a large specific surface area in the order of $2600\,m^2g^{-1}$ makes graphene very suitable for adsorbing aromatic organic pollutants in water [52.41]. Cofactors such as nicotinamide adenine dinucleotide hydrogen (NADH) are catalytic agents for many important biological activities in the human body. During mitochondrial respiration, the electron transfer takes place

through the NADH producing sufficient adenosine triphosphate (ATP) enzymes for cellular energy supply. The concentration of NADH varies in different parts of the body depending on the energy requirements. An unexpected elevation or lowering of NADH may cause severe diseases like brain ischaemia and Leigh syndrome. The oxidation of the NADH serves as the anodic signal and regenerates the NAD^+ cofactor. The shift of peak potentials of NADH oxidation from 0.70 V on the GC to 0.40 V on the RGO electrode makes graphene a suitable material for highly sensitive electrochemical biosensors of low working potentials, offering low detection limits and long-term stabilities [52.42]. Graphene oxide (GO) and its reduced form (RGO) are studied extensively in the field of biotechnology because of their multivalent functionalization and efficient surface loading with various biomolecules. The negative surface charge originating from a free π-electron cloud can be utilized to condense genes and anticancer drugs. For example, when 6-armed polyethylene glycol (PEG) was covalently linked to the surface of GO sheets to form biocompatible GO-PEG, which is nontoxic to human lung and breast cancer cells. Graphene may act as a potential antimicrobial coating for surgical equipment by preventing the formation of pathogenic and corrosive microorganisms and even killing bacteria. Real-time high-resolution optical imaging as well as electrophysiological recording of neuronal locations are essential for examining the operation of individual neural circuits in neurological disorders. As conventional metallic microelectrodes are opaque, they shadow the investigator's view, making it necessary to record the imaging and electrophysiological data separately. However, transparent, flexible graphene electrodes may now be used for the simultaneous temporal and spatial resolution required for neuroscientific treatments. The nonmagnetic and anti-corrosive properties of graphene offer added advantages [52.43].

52.3 Macrocylic Compounds for Sensing

Phthalocyanine molecules belong to a class of macrocylic compounds. Because of their 18 π-electron aromatic macrocyclic structure these molecules exhibit important optoelectronic, photophysical, conduction, and photoconduction properties [52.44]. As shown in Fig. 52.3, the Pc structure is closely related to the naturally occurring porphyrins, which form a core skeleton in hemoglobin and chlorophyll. Organic molecular nanostructures have huge potential but suffer from unstable device characteristics, low resistance to adverse environments (e.g., temperature, humidity, oxygen, etc.) and lack of reproducibility of material composition, purity, and fabrication conditions. Phthalocyanines (Pcs) are, however, tinctorially strong, and chemically and thermally stable. The materials are abundantly available in a highly purified form.

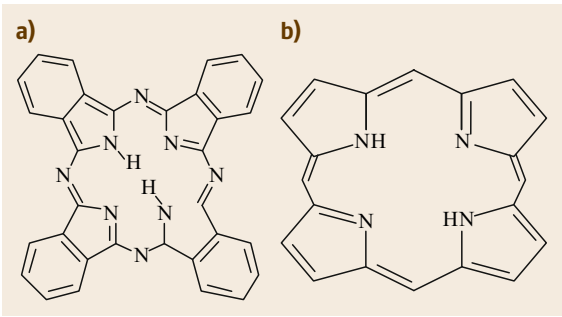


Fig. 52.3a,b Chemical structure of: **(a)** a metal-free phthalocyanine and **(b)** porphyrin molecule

52.3.1 Preparation of Sensing Membranes

Sensing is essentially an interaction mechanism of a solid surface and analytes and the molecules are organized in ultrathin films, offering a large surface-to-volume ratio to improve the probability of these reactions taking place. A variety of methods such as direct-current DC and radio-frequency RF sputtering [52.45], thermal evaporation [52.46], Langmuir–Blodgett (LB) [52.47], spin coating [52.48], and self-assembly (SA) [52.49] are employed to produce organized thin organic films.

The substrates play a key role in determining the film quality and the structure. It is found that the island density and coverage ratio of copper tetra-*tert*-butyl phthalocyanine (CuTTBPc) on hydrophilic glass surfaces were small during the early growth stage of vacuum deposition due to a small nucleation rate arising from the weak interaction of molecules. The film surface was also rough. The high nucleation rate of CuTTBPc on a silanized surface, on the contrary, produced a much smoother morphology consisting of densely distributed fine-grained clusters. The aggregation of the molecules took place in domains during the formation of the LB monolayer, separately distributed on the glass surface. These domains acted as active sites for the nucleation and growth of the later deposition process and high-density clusters were thus found in the early growth stage. X-ray diffraction (XRD) patterns demonstrated

that the LB films grown on the glass substrate were more crystalline than those on other substrates [52.50].

The wet techniques, namely LB, spin coating and SA, are preferred to high-energy deposition because the deposition takes place at room temperature, reducing the possibility of chemical decomposition of the molecules due to thermal damage. These deposition methods are considered below.

Langmuir–Blodgett (LB) Technique

The LB method involves the transfer of individual Langmuir monolayers of amphiphilic molecules which self-assemble in an ultra-clean water subphase onto a suitably prepared solid substrate. An LB film, which may consist of a single or many layers, up to a depth of several visible-light wavelengths, is generally produced with fine control over the thickness and geometry to a precision at the molecular level. With the Y-type construction, these LB films are essentially layered materials with alternating hydrophobic/hydrophillic regions. Noncentrosymmetric LB films can be obtained by X-type construction on immersion of the substrate from air into water only, or Z-type deposition on withdrawal of the substrate from water to air [52.51].

Macrocylic molecules are not generally amphiphilic and are therefore not considered suitable for LB deposition. The synthesis of a series of specially designed Pc molecules for LB deposition has been reported in the literature over the last two decades [52.52–54]. Figure 52.4 shows the chemical structure of asymmetrically substituted 1,4-bis(6-hydroxyhexyl)-8,11,15,18,22,25-hexa-octylphthalocyanine molecules belonging to this family of custom-designed molecules [52.55].

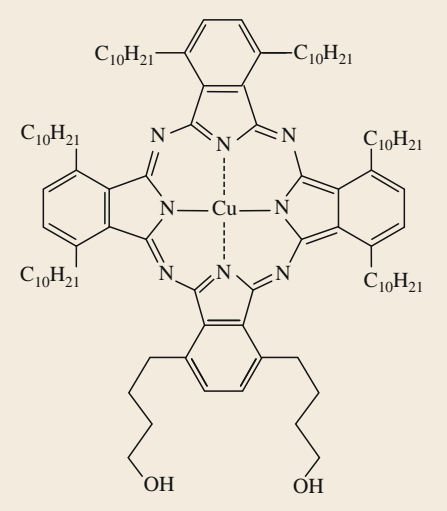


Fig. 52.4 Chemical structure of 1,4-bis(4-hydroxybutyl)-8,11,15,18,22,25-hexakisdecyl phthalocyaninato copper molecules

To start with LB deposition, a Langmuir monolayer is formed from a spreading solution of substituted Pc derivatives in an appropriate volatile solvent (such as chloroform) on the highly purified water subphase (Milli-Q water of resistivity 0.18 MΩm) contained in an LB trough made of a hydrophobic material (Teflon). The solvent is allowed to evaporate fully and the molecules are then slowly compressed to the desired organization by closing a servo-controlled Teflon barrier straddling the air–water interface at a constant speed (say about 100 mm/min). As shown in Fig. 52.5, the phase changes of the molecules during the compression are monitored by recording the surface pressure–area isotherm. The pressure begins to rise rapidly as the solid phase is reached while the area per molecule remains approximately constant. The molecules are transferred from the water–air interface to the substrate by dipping the substrate in at a constant speed of about 10 mm/min when the surface pressure is

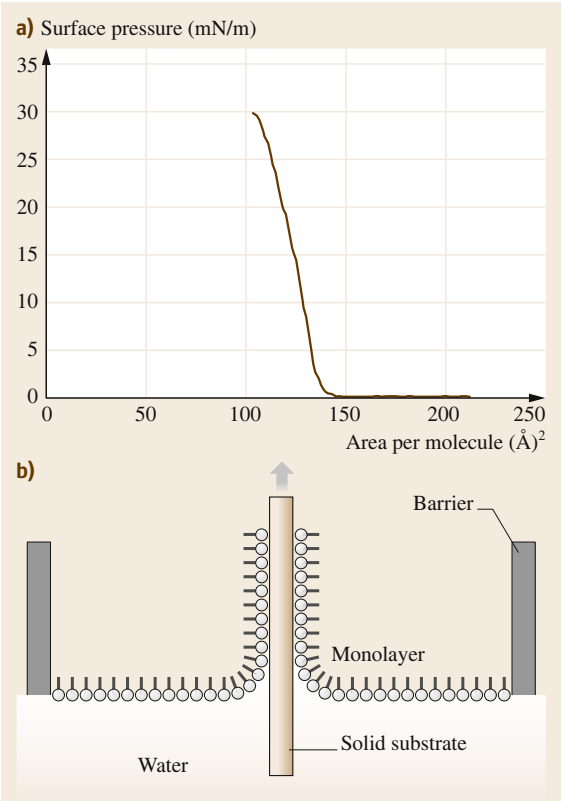


Fig. 52.5a,b Deposition of Langmuir–Blodgett film: **(a)** surface pressure–area isotherms and **(b)** transfer of floating monolayer onto a suitable substrate floating monolayer

in the range 20–40 mNm^{-1}. However, a dipping rate as high as 1000 mm/min was used to produce LB films of 5,10,15,20-tetrakis(3,4-bis[2-ethylhexyloxyphenyl])-21*H*,23*H*-porphine (EHO) molecules. Atomic force microscopy (AFM) showed a morphology of these films which is highly conducive to the host–guest interaction. The presence of isolated micron-sized domains, which were themselves composed of grains of several nm in diameter, was observed [52.56].

The molecular orientation in LB films of these novel substituted Pc derivatives was investigated by spectroscopic methods, including electron spin resonance, reflection adsorption infrared spectroscopy (RAIRS), polarized infrared (IR) spectroscopy, and AFM. Amphiphilic mesogenic Pcs afforded the best films. These all contained a good degree of molecular order. Furthermore, the type of molecular stacking in the films was found to depend upon the length of the side chains and the central metal [52.57].

Spin-Coating Method

The spin-coating method, designed for defining patterns on silicon wafers in integrated circuit (IC) technology, has now been adapted for deposition of thin organic films on suitable substrates [52.58]. Substrates are vacuum held onto the rotating chuck of a photoresist spinner. The spread solution was prepared in the same way as done for the LB deposition and a small volume of the solution is dispensed onto the rotating substrate surface from a microsyringe held at a distance of 5 mm above the spinner platform. A spinning time of 30 s for a solution concentration of 0.5–5.0 mg/ml is found to be sufficient in order to produce a uniform, homogeneous film. The rotation speed is normally varied between 1000–6000 rpm and the speed of spin and the concentration of solution largely determine the thickness of the resulting films. The dependence of the film thickness d on spin speed ω and time t may be written in terms of an empirical expression [52.59]

$$d = \frac{C}{\sqrt{t}}\omega^{-x}\,, \tag{52.1}$$

where C depends upon the evaporation rate of the solvent, and the viscosity and density of the solution. The index x is predominantly determined by the evaporation rate of the solvent. If the evaporation of the solvent is independent of the spin speed, x takes on a value of $2/3$. For most solvents used, the evaporation varies with $\sqrt{\omega}$ and x becomes $1/2$. A value of $x = 1$ is a valid approximation for slow evaporation [52.60, 61].

Spin coating is a simpler and more cost-effective method than LB deposition and also there is no requirement for the molecules to be amphiphilic. The LB technique, however, provides more precise control over the film thickness [52.62]. Both spun and LB thin films of octa-alkyl-substituted phthalocyanine molecules displayed similar characteristic features in their optical absorption spectra. The appearance of Davydov splitting and the XRD pattern indicated the same crystallinity of the films [52.63]. Different physical structure and molecular packing may result from the use of different wet techniques. For instance, asymmetrically substituted phthalocyanine molecules (Fig. 52.4) were deposited by LB and spin-coating techniques. XRD patterns and UV-Vis spectra of these films indicated some degree of anisotropic ordering within the LB films compared to spun films [52.64]. The spin-coated films were, however, exceptionally smooth, even and free from crystallites. A noncompact spin-coated film of phthalocyanine molecules bearing one crown on the ether ring exhibited a faster response and reversal time than the well-ordered and tightly packed LB films [52.65].

Self-Assembled Monolayers (SAMs)

Self-assembled monolayers (SAMs) have attracted considerable interest because of their simplicity of preparation, stability, and versatility. The introduction of due functional groups into SAMs allows changes of state and property on surface or interface. These attributes make SAM films attractive for sensor technology [52.66, 67]. SAM films are formed via reaction of a functional group within a molecule with the surface of a solid substrate. The methoxy or trichlorosilylalkyl derivatives of molecules form weak bonds with hydroxyl groups and are suitable for the formation of a SAM on an oxide surface such as indium tin-oxide (ITO) and glass (SiO_2) [52.68]. A monolayer layer can also be self-assembled on gold or silver substrates if the molecules are attached with thio or disulfide groups at one end [52.69]. In supramolecular aggregation during the self-assembly process, the interactions between the adsorbate molecules are as important as those between adsorbates and the surface. The size and aggregation pattern for the formation of supramolecular structures (J-aggregates) are controlled by changing the number of polar sulfonic groups of meso-tetra (4-sulfonatophenyl) porphine [52.70].

Much fundamental work on SAM films was initially undertaken with simple functionalized alkane derivatives [52.71]. In the case of alkylthiols on gold, it is well established that the monolayer films were densely packed with a high degree of order, with chains typically oriented at about 65° to the substrate surface. In recent years, work has been undertaken to include a single tether chain with thiol or disulfide derivatives for SAM formation on gold-coated glass substrates.

The molecular orientation with respect to the gold surface was found to depend upon the different alkyl-connecting chain lengths and different macrocyclic peripheral side chains [52.72]. The core of substituted Pc molecules bearing mercaptoalkyl groups is not parallel to the surface if the tether chain is increasingly long. For sufficiently long tethers, the core becomes oriented almost normal to the surface [52.73]. SAM films of thiol-derivatized cobalt phthalocyanine complexes on gold electrodes were susceptible to destruction via oxidative and reductive desorption and their potential applications as a sensor were therefore limited [52.74].

Self-assembled ultrathin films have recently received considerable interest because they allow fabrication of supramolecular assemblies with tailored architecture and properties [52.75]. Multilayered films can also be built up by sequential deposition of polycations and polyanions based upon electrostatic attraction [52.76–78]. Alternate deposition of planar functional molecules meso-tetra (4-sulfonyl) porphyrin ($TPPS_4$) or copper phthalocyaninetetrasulfonic acid, tetrasodium salt (CuTsPc) and cationic bis (pyridinium) salt was made to produce a self-assembled multilayered film [52.79]. Multilayer assemblies of porphyrin and phthalocyanine were also prepared by alternating deposition of oppositely charged rigid planar molecules of sodium (phthalocyanine tetra-sulfonate)cobalt $Na_4[(CoTsPc)]$ and tetrakis(N,N,N-trimethyl-4-anilinium) porphyrin cobalt $[(CoTAP)]Br_4$ [52.80]. A further investigation of the C-11 Pc derivative SAM has shown that the monolayer film was stable, exhibiting no oxidation and only minor orientation changes on the gold surface over a period of 24 months as determined by infrared and fluorescence spectroscopy. The long active life of the Pc SAM was thought to be due to the macrocyclic ring possibly preventing oxidation of the thiolate root [52.81].

52.3.2 Thin-Film Properties

Surface/interface composition at the nanometer scale is believed to affect the organization and chemical/physical properties of organic thin films. This long-range organization determines both the optical and electrical properties of thin-film materials, leading to unique, new chemical sensor platforms. The overlap of π-electron wave functions from Pc molecule to Pc molecule is likely to occur between adjacent stacks in a thin-film form, giving rise to the broadening of energy levels [52.82].

Phthalocyanine Films

Phthalocyanines can exist in different polymorphic forms: α and β are among the most common. Both forms possess the herringbone structure for the stacking arrangement. The angle between the stacking axis and the normal to the molecular plane depends upon the method of preparation, types of phthalocyanine (metalo or metal-free), and substituents. For example, this angle is generally 25° and 45° for the α and β phases, respectively [52.83]. A typical absorption spectrum in the UV-visible range is shown in Fig. 52.6 for a substituted metal-free phthalocyanine (Pc) deposited on a glass substrate (similar to one shown in Fig. 52.3). The Q absorption band of the LB film was broader than one for the chloroform solution and appeared at wavelengths between 650–800 nm. The Davydov splitting became apparent when compared with the spectra for the molecules in chloroform. The film-forming properties and the stack order were both found to be influenced by the chain length of the ring substituents [52.84, 85]. An analysis into the in-plane molecular arrangements in the LB films of Pc compounds with short and long alkyl chains was performed by decomposing the absorption spectra into Gaussian–Lorentzian components. Long-chain molecules were found to be organized in a *herringbone* structure leading to the Davydov splitting of the Q-band. For short-chain molecules, the largely broadened blue component of the Davydov doublets was suppressed possibly due to the domination of an anisotropic stack-like molecular arrangement in LB films [52.86]. Characteristic Q-bands observed at about 630 nm in the UV-Vis spectra of the self-assembling monolayers (SAMs) of 2,9,16-tri(*tert*-butyl)-23-(10-mercaptodecyloxy)phthalocyanine and its disulfide on gold substrates were broadened and blue-shifted relative to those observed in solution [52.87]. The formation of high-quality Langmuir–Blodgett films of copper (CuPc) and nickel phthalocyanine (NiPc) derivatives

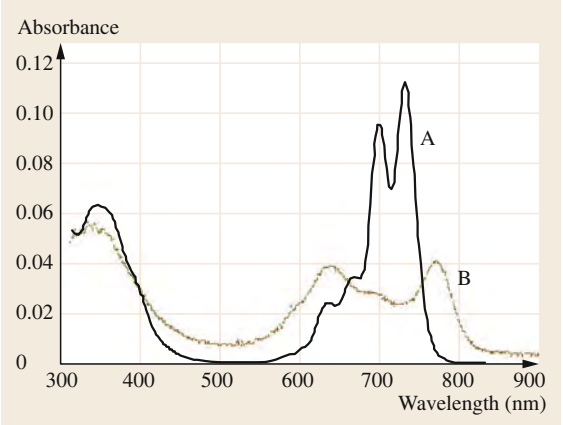


Fig. 52.6 Typical UV-visible spectra for thin films of metal-free phthalocyanine molecules in: (A) chloroform and (B) an LB film

substituted with short branched chains is reported. Ellipsometric and polarized optical absorption measurements suggest that the phthalocyanine molecules have a preferred orientation, with their large faces perpendicular to the dipping direction and to the substrate plane [52.88].

Electrical properties of Pc molecules are very dependent upon the doping and the nature of the substituents in the macrocyclic compounds. Copper(II) tetrasubstituted phthalocyanine complexes, CuPcX (where X = $-NO_2$, $-NH_2$, $-SO_3H$, and $-OH$) exhibited an improvement of conductivity by five orders of magnitude over its unsubstituted counterpart. Further enhancement in conductivity was achieved by iodine doping. This increase in conductivity was attributed to the possible decrease in the metal–metal bond distance [52.89]. The effect of oxygen on charge transport was investigated by measuring the time-dependence behavior of the film conductivity of zinc Pc (ZnPc) layers at a stable pressure of about 0.1 mbar of air and it was found that the conductivity gradually decreased by an order of magnitude with time. The conductivity reached saturation within several minutes for the film thickness smaller than 500 nm whereas similar electrical degradation in a 7 μm-thick film required more than half an hour. This thickness-dependent decrease in conductivity was caused by the velocity of the oxygen out-diffusion processes in the films [52.90]. The effect of doping on conductivity of a thermally deposited composite film of α-nickel phthalocyanine (NiPc) and the strong electron acceptor tetracyanoquinodimethane (TCNQ) was investigated by exposure to air. The doping took place at a much faster rate in the composite than in the NiPc films. Values of 1.3 and 1.9 eV for the activation energy were obtained for devices incorporating TCNQ and the non-TCNQ devices, respectively, and the conductivity of the composite film was therefore found to be 20 times larger than the NiPc film [52.91]. Because of the excellent ability to photogenerate free charge carriers, the conductivity of the metal phthalocyanine (MPc) film is increased considerably as can be seen in Fig. 52.7 showing the dark and illuminated current–voltage characteristics of a 10 nm-thick spun nickel phthalocyanine (NiPc) sandwiched between two metal electrodes.

Phthalocyanines of rare-earth metals primarily occur in the form of bisphthalocyanines with a sandwich-type structure and the cation is eightfold coordinated to the two macrocycles. One of these phthalocyanine rings is virtually planar while the other is significantly distorted [52.92]. The Q-band in the absorption spectra of spun films of substituted lutetium bisphthalocyanine molecules becomes broader and red-shifted compared to the molecules in chloroform solution. Heat treatment of the film at temperatures above 120 °C introduced molecular ordering, possibly due to the edge-to-edge interaction between neighboring Pc moieties [52.93]. All these factors have led to successful formulation of phthalocyanines as chemically sensitive membranes for environmental monitoring [52.94–98].

Porphyrin Films

Inherent stability, unique optical properties, and synthetic versatility of porphyrins and metalloporphyrins are exploited in sensor applications. Synthetic porphyrins and metalloporphyrins incorporated into polymers, glasses, and Langmuir–Blodgett films matrices are used for detection of NO, CO_2, and O_2 because of the ability to bind these gases. Metalloporphyrin sensing arrays have been used for the detection of organic odorants, such as amines, thiols, and phosphines [52.99]. Porphyrins also combine readily with metals, coordinating with them in the central cavity. Many gas sensors take advantage of analyte binding to the center metal atom in the porphyrin ring and a detectable optical change occurs as a result. In particular sensing properties of iron, zinc, copper, nickel, and cobalt containing porphyrins have now been studied [52.100–103]. The porphyrin rings of thiol-derivatized cobalt(II)-5,10,15,20-tetrakis(4-*tert*-butylphenyl)-porphyrin (CoTBPP) molecules were immobilized on the surface of a 4-aminothiophenol (4-ATP) self-assembled monolayer (SAM) on gold (111) by in situ axial legation. The reconstruction of the herringbone structure took place probably due to adsorption/desorption processes of molecules [52.104].

Substituents have considerable effects on the structure of porphyrin monolayers; for example, the interaction between the porphyrins deposited on the ITO

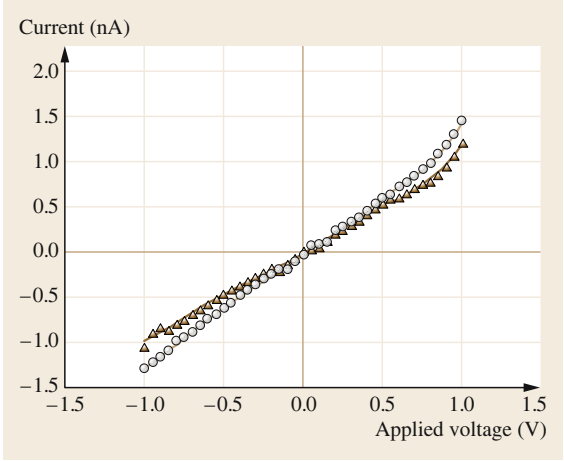


Fig. 52.7 Dark- (*triangles*) and photoconduction (*circles*) through spun NiPc films

substrate without bulky *tert*-butyl groups was much stronger than that of the porphyrins with bulky *tert*-butyl groups [52.105]. Hybrid molecule-silicon capacitors were formed by attaching [5-(4-dihydroxyphosphorylphenyl)-10,15,20-trimesitylporphinatozinc(II)] porphyrin complexes to silicon oxide via a phosphonate linkage and the presence of multiple distinct peaks in electrical characteristics were associated with oxidation and reduction of the molecular monolayer [52.106]. Multilayered thin films of porphyrins bearing bulky alkoxyphenyl substituents at two of the four meso-positions and phenyl phosphonates at the other two were deposited on the ITO substrates using the zirconium phosphonate linkage. Molecular aggregation was prevented by the addition of sterically demanding 2,6-di(*n*-hexoxy)phenyl substituents to the meso-positions of the porphyrin skeleton. The compounds were able to form thin films without the presence of large molecular chromophore interaction relative to sterically unhindered porphyrins [52.107]. Highly conducting tetraruthenated cobaltporphyrin (Co (TRP)) and tetrasulfonated zincporphyrin (Zn (TPPS)) multi-bilayer films were prepared by electrostatic assembly for sensor applications [52.108].

52.4 Sensing with Phthalocyanine and Porphyrin

The interaction of reactive gases such as O_2, NO_2 and NH_3 with the phthalocyanine group of compounds has been shown to result in conductivity changes of these compounds. This effect was associated with charge transfer between the electron-donating phthalocyanine molecule and the electron-accepting gas [52.109].

52.4.1 Amperometric Sensor

The action of the oxidizing gases O_2 and NO_2 and the reducing gases NH_3 and H_2 diluted in N_2 on the electrical conductivity of thin layers of zinc hexadecafluorophthalocyanine ($ZnF_{16}Pc$) and zinc phthalocyanine (ZnPc) was investigated. The conductivity $ZnF_{16}Pc$ increased in the presence of NH_3 whereas ZnPc exhibits no sensitivity to this gas. The sensitivity to H_2 was higher for $ZnF_{16}Pc$ than for ZnPc. This behavior was related to the redox potentials of the two phthalocyanines [52.110, 111].

Response and recovery times for LB CuTTBPc films on platinum IDE are shown in Fig. 52.8 for different gas concentrations of 1–5 ppm. These times were determined from a model based upon the kinematics of adsorption and desorption of NO_2 gas molecules. It was found that recovery times were longer than response times, irrespective of gas concentration. The response time constant varied in the range 105–159 s. Repeated exposures to NO_2-air mixtures resulted in gradual degradation of reproducibility. The response was also believed to be influenced by the past history of the exposure of the sensor to the NO_2 gas, if the recovery time constant was longer than the switch-off time. These factors show that CuTBPc is a material that is suitable for fabricating a disposable *single-shot* NO_2 sensor for practical uses [52.112]. LB films of asymmetrically substituted phthalocyanine, 1,4-bis (4-hydroxybutyl) 8,11,15,18,22,25-hexaoctylphthalocyanine molecules showed selective responses to 1–5 ppm concentrations of NO_2 at the room temperature. Moisture had no effect on the response because of the hydrophobic nature of the alkyl groups. Exposure to Cl_2, NH_3, SO_2, and CO at ppm levels produced no measurable response [52.113]. Further work on the LB films showed that the NO_2 response was larger for the LB film produced by dipping the substrate perpendicular to the electrode finger than those dipped parallel to the fingers. This preferential behavior was attributed to the ease of conduction along the long chains [52.114]. LB films of meso,meso′-buta-1,3-diyne-bridged Cu(II) octaethylporphyrin dimer molecules were used to fabricate amperometric sensors operating at a temperature of 90 °C. LB multilayers were reported to be selectively sensitive to small concentrations of NO in air without being affected by contemporary presence in NO_2, CH_4, C_2H_5OH, and CO [52.115].

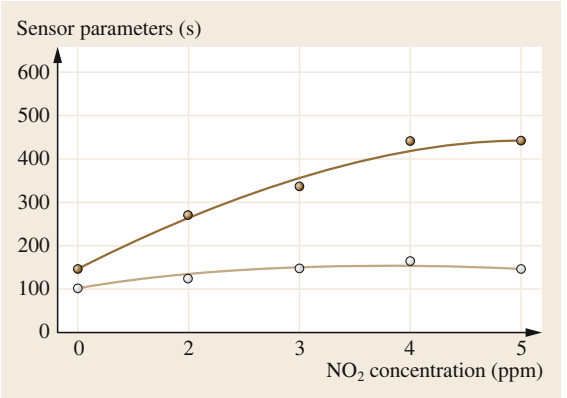


Fig. 52.8 Response (*broken line*) and recovery (*solid line*) of a 15-layer LB film of CuTTBPc when exposed to NO_2 periodically at 2 min intervals at 1, 2, 3, 4, and 5 ppm levels in succession

Thickness, operating temperature, doping and postdeposition treatment of sensing membranes all affect sensor performance. Vacuum-deposited CuPc of thickness 50–400 nm onto gold IDEs on glass substrate were exposed to NO_2 gas of concentration 10–100 ppm. The film thickness had little effect on room-temperature sensitivity, but the recovery ratio decreased with increasing film thickness. Thicker films showed improved recovery at 100 °C [52.116]. Because of the tetra-*tert*-butyl substitution on the periphery of phthalocyanine, CuTTBPc films have a larger lattice spacing and higher film resistance. Sublimed CuTTBPc film showed a higher recovery ratio in the NO_2 sensing experiments [52.117].

NO_2 and O_2 treatment of evaporated CuPc films that were 160–200 nm thick on IDE improved the sensitivity and recovery time when exposed to 0.5–5 ppm of NO_2 gas. A Langmuir behavior was predicted for the adsorption/desorption processes on the weakly binding surface sites of NO_2-treated CuPc films having a single activation energy [52.118]. When the CuPc films were subjected to postdeposition treatment by cooling down to 77 K, the mechanism of NO_2 adsorption, however, followed the Elovich model [52.119]. CuPc and iron Pc (FePc) films (100 nm thick) were heat treated at 240 °C under vacuum and then the responses of the films to the exposure of NO_2 in the 8–12.8 ppm range were monitored. The CuPc films were transformed to a stable β-phase and the adsorption occurred on the ligand rings. Conversely, NO_2 was adsorbed on both the ligand ring and the central ion of the FePc films [52.120]. An amperometric sensor was fabricated using the spun film of tetra-*iso*-propoxyphthalocyaninato copper(II) (*i*-Pro-CuPc), and the response, recovery, and cyclic properties of the sensor were very temperature sensitive. The results showed ideal gas sensitivity at 65 °C and a good linear relationship existed between the resistance of the gas sensor and the concentration of NO_2 [52.121]. NiPc films (25 nm thick) were deposited between the source and the drain of a field-effect transistor (FET) and the change in the drain current exhibited a linear dependence on the ozone concentration in air, in the range 0–150 ppb [52.122, 123].

Water-soluble copper phthalocyanine tetrasulfonate (CuTsPc) molecules were alternatively self-assembled with a bipolar pyridine salt on a 3-mercaptopropionic-acid-modified Au electrode to produce copper ion-selective electrodes (ISE). These ISE were found to have low resistance, short conditioning time, and a fast response. The electrode potentiometric response showed a selective potentiometric response to Cu^{2+} ions in the range from 10^{-5}–10^{-1} M, independent of the pH of the solution between pH 1–5. The response was dependent on the nature of the medium and the detection limit was found to be 7×10^{-6} M in acetate buffer (0.2 M, pH 4.5) [52.124]. Carbohydrates were detected by flow-injection analysis at the water-soluble cobalt phthalocyanine tetrasulfonate (CoTsPc)-modified electrode with high sensitivity. Detection limits obtained in this manner range from 150 pmol injected for glucose and fructose to 600 pmol injected for maltose and sucrose [52.125].

52.4.2 Optical Sensors

The amperometric approach generally suffers from long response times and their use is restricted due to possible fire and explosion hazards caused by electrical power. Organic systems are also prone to slow degradation arising from the chemical reactivity of charge carriers. Conversely, optical transduction techniques offer high speed, high precision, immunity to interference, and remote sensing capabilities for detection of trace gas species.

Optical Absorption and Fluorescence Spectra

Optical transmittance of thermally evaporated lead phthalocyanine (PbPc) and nickel phthalocyanine (NiPc) thin films changed on exposure to NO_2 in the IR region of the optical absorption spectrum [52.126]. The disappearance of Davydov splitting from the UV-Vis absorption spectra of 20-layer-thick LB films of 1,4-bis(4-hydroxybutyl)-8,11,15,18,22,25-hexakisdecyl phthalocyaninato copper molecules was observed when the films were exposed to 100 ppm of NO_2 [52.127]. The local interaction between the central metal atoms and NO_2 distorted the herringbone structure. Gas adsorption and desorption processes were described by exponential functions, both characterized by a single time constant. The dispersion relation in Fig. 52.9 for the refractive index was obtained by using the simple Kronig–Krammer transform of absorption spectra of the LB thin film of another octasubstituted metal-free phthalocyanine molecule. The full recovery of the refractive index to its original value was achieved after flushing with the N_2 gas for wavelengths larger than 782 nm. Ellipsometry measurements on thin films of tetrasulfonated copper phthalocyanine showed increases in the real and the imaginary parts of the dielectric constant on exposure to NO_2 gas possibly due to the rise in the density of dipoles and higher absorption, respectively [52.128].

LB films of 5,15-bis(4-aminophenyl)-10,20-bis[3,4-bis(2-ethylhexyloxy)phenyl]-21H,23H-porphine prepared at a dipping rate as high as 500 mm/min showed an inhomogeneous structure with high porosity. These porphyrin films showed much faster response to 4.6 ppm of NO_2 gas and a sensitivity of 60% relative to absorbance change at 439 nm was achieved.

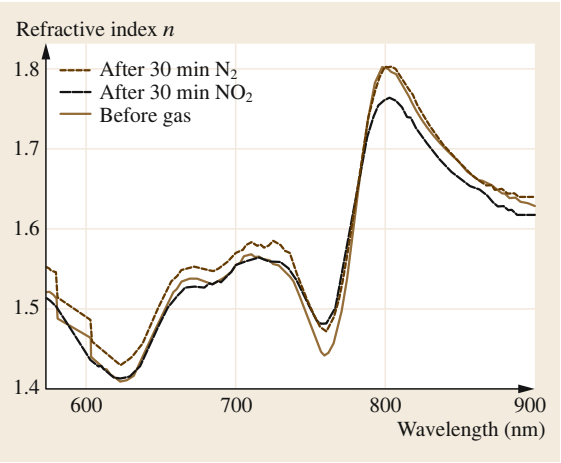


Fig. 52.9 The dependence of the refractive index of octa-substituted phthalocyanine on the wavelength before and after exposure to 90 ppm NO_2 gas for 30 min. The recovery was examined after flushing of the LB films with N_2 for 30 min

Full and accelerated recovery of the original spectrum was also achieved by gentle heating at 353 K. The sensor was reported to have a shelf-life as long as one year [52.129]. Kinetic UV-Vis spectroscopic measurements were made on a porphyrin-entrapped sol–gel film for NO_2 sensing. The method was capable of detecting levels as low as 176 ppb of NO_2 at room temperature and showed a fast and full recovery [52.130]

The blue color of rhodium phthalocyanine LB films becomes transparent upon chlorination. A pronounced quenching of the characteristic triplet centered on the Q-absorption band took place at 662 nm, but the absorption band in the near UV part of the spectrum remained almost unaffected. Optical spectra partially recovered when the exposed film was left in air for several hours. Overall effects of chlorination (quenching and recovery) depended on the period of exposure [52.131]. LB gadolinium phthalocyanine films were exposed to 10 ppm of chlorine and changes in optical absorption spectra were observed as a result of the oxidation of the film by chlorine to form $[Gd(Pc)_2]^+$ species. The kinetics of such changes were described by a polyfunctional process with different decay time constants and the effects were found to be highly reversible [52.132].

SAM films are of interest because the chemically bound monolayer is intrinsically more robust than LB films and the response times are also expected to be faster. Monolayers of diphthalocyanine disulfide Pc molecules were self-assembled on gold-coated glass substrates with the Pc macrocycle oriented parallel to the metal surface. Changes of the reflectivity signal were observed on exposure to NO_2 gas in proportion to the concentration [52.133]. The fluorescence emission spectrum of each of the Pc SAMs was obtained by exciting each monolayer along the longitudinal axis of the optical waveguide via laser-induced evanescent-wave stimulation. The use of the longer mercaptoalkyl connecting chain appeared to inhibit quenching of the electronically excited state through energy transfer to the metal layer. Fluorescence response to 10 ppm NO_2 was found to be selective with no interference from CO and CO_2 [52.134].

Surface Plasmon Resonance Technique

The surface plasmon resonance (SPR) technique is based on the excitation of surface electromagnetic waves of transverse magnetic (TM) modes traveling along the interface between a metal and a dielectric medium. Using a semicylindrical prism of high refractive index n_p in *Kretschmann*'s configuration, as shown in Fig. 52.10, these waves are excited by a p-polarized light of wavelength λ via the evanescent field [52.135]. The evanescent field will be completely attenuated at a particular angle of incidence corresponding to resonance and there will be no reflection. Figure 52.11 shows the SPR results when a spun film of nickel(II) tetrakis(4-cumylphenoxy)phthalocyanine (NiPc) molecules was subjected to a sequence of exposure to ozone (O_3) at 2 ppm. A period of 20 mins was allowed for the film to recover to its original state. The shift of the resonance angle $\Delta\theta_a$ was believed to be caused by a change in the refractive index of the film due to the interaction of O_3. $\Delta\theta_a$ is usually written in the form [52.136]

$$\Delta\theta_a = \frac{(2\pi/\lambda)(|\varepsilon_m|\varepsilon_i)^{3/2}d}{n_p \cos\theta(|\varepsilon_m| - \varepsilon_i)^2\varepsilon}(\varepsilon - \varepsilon_i)\,, \qquad (52.2)$$

where $|\varepsilon_m|$ is the modulus of the real part of the dielectric constant of the gold film; ε_i is the dielectric constant of the medium (air in this case) in contact with the organic thin layer of thickness d and the di-

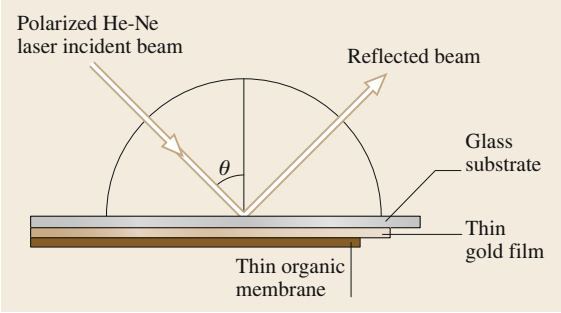


Fig. 52.10 A Kretschmann-type configuration for optical sensing interrogation

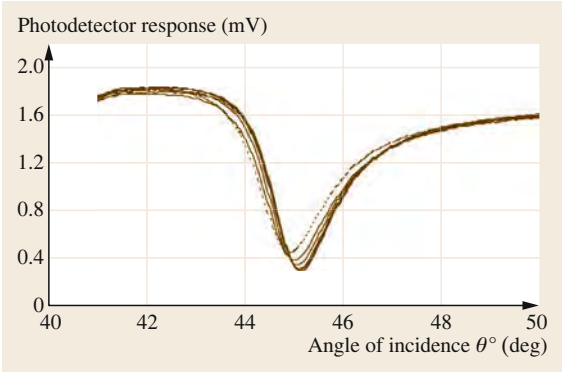


Fig. 52.11 The effect of a sequence of exposure at 20 min intervals of 2 ppm O_3 on a NiPc spun film. The *broken line* represents the system prior to exposure

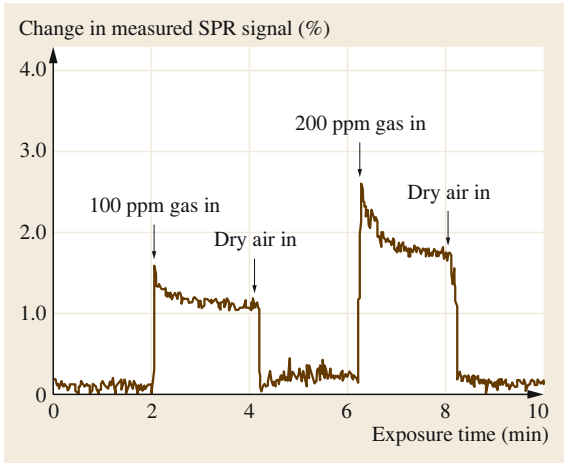


Fig. 52.12 Kinetics of the SPR response of copper hexadecafluorophthalocyanine ($CuPcF_{16}$) film to NH_3 gas at 100 ppm and 200 ppm. Measurements were at a fixed angle of $\theta^* = 46°$

electric constant ε. The resonance minimum moved to an increasingly larger angle after each exposure to ozone. The half-width became narrower at the same time and this was believed to be the consequence of oxidation of the phthalocyanine ring by ozone, which resulted in bleaching of the dye. SPR responses are in general slower and less readily reversible than semiconductivity changes but the technique is capable of providing additional information on the interaction between the gas and films as the shifts in both the angle and depth of the resonance can be monitored. Experimental data from the SPR response of CuPc films to nitrogen dioxide were input to a design process using an evolutionary algorithm. Gold layers with a thickness of 25 nm were found to exhibit 42.9% greater contrast than layers with a thickness of 50 nm. Using evolutionary design predictions, further modifications could be tested for available materials, and redundant layers may be eliminated. By inclusion of the external optics, a design could be selected to accommodate poor precision $(+/-0.5°)$ in the incident angle and a possible multilayer solution was shown using Teflon AF 1600, with a refractive index similar to 1.3. The predicted NO_2 response showed an improvement compared with the classical SPR configuration, and the incident angle chosen by the genetic algorithm for the interrogation of these layers was close to a stationary point in the absolute response curve [52.137].

Figure 52.12 shows the transient response when thick films of copper hexadecafluorophthalocyanine ($CuPcF_{16}$) were exposed to 100 ppm of NH_3 for 2 min followed by the injection of dry air for a further 2 min period. The film was exposed to 200 ppm of NH_3 for the next cycle. The SPR response was very rapid and the recovery was complete. Unsubstituted CuPc was practically insensitive to this gas. Strong electron-withdrawing fluorine substituents made this substituted compound sensitive to reducing gases such as NH_3 [52.138].

SPR experiments were performed on films of different phthalocyanine derivatives in order to investigate the interaction mechanism with NO_2. SPR signal concentration as low as 1 ppm was measured using spun films of crown-ether-substituted phthalocyanines and octa-3,7,11-trimethyldodecyloxy phthalocyanine [52.139]. Adsorption of NO_2 took place both on the surface of the film and into the bulk. The surface adsorption increased the film thickness while the change in the real part of the film refractive index occurred due to the diffusion of NO_2 into the LB film. These processes were reversible and had no effect on the optical absorbance of the LB film. The SPR response of the bare Ag film to the exposure of NO_x (the equilibrium mixture of nitrogen dioxide and dinitrogen tetroxide) was slightly slower than the LB films of the tetra-4-*tert*-butyl-phthalocyanitosilicon dichloride (ttb-$PcSiCl_2$/Ag) system. This effect was associated with the high affinity of the phthalocyanine to NO_x gas. The NO_2 gas was also able to reach the Ag/Pc interface film and the reaction with the metal film resulted in the growth of a surface layer. This irreversible effect was found to be predominant during the recovery cycle and considerably reduced by the presence of a buffer LB layer of ω-tricosenoic (ω-TA) acid [52.140]. SPR studies showed that exposure of thin films of 18-crown-6 metal-free phthalocyanine molecules to NO_2 decreased the film absorbance. The optical permittivity of the NO_2-treated Pc over-layer was regarded as being an average of the optical permittivities of both components since the optical fields associated with surface plasmon po-

laritons sampled both the Pc and the NO_2 molecules. The disruption in the charge conjugation of the phthalocyanine molecules was observed because of the generation of organic radical cations as end products of the charge transfer during the physisorption process [52.141].

Using the excitation at wavelengths of 488 and 632.8 nm, SPR measurements were made on LB films of porphyrin molecules of EHO type deposited on an Ag-coated glass substrate. The resonance shift observed was larger at 488 nm than at 632.8 nm when the EHO films were exposed to NO_2 gas. The thickness and the complex dielectric constants at 632.8 nm were attributed to the island structure of the EHO LB films but the derived properties at 488 nm were believed to be caused by dispersion due to optical absorption [52.142]. Similar results were further obtained from UV-Vis absorption in these films [52.56].

52.4.3 Detection of Volatile Organic Vapor Compounds

Phthalocyanine molecules have also been extensively investigated for the recognition of volatile organic compounds (VOC). Langmuir–Blodgett (LB) and evaporated films of lutetium bisphthalocyanine ($LuPc_2$) molecules were deposited on indium tin-oxide (ITO) interdigitated electrodes in order to study the changes in their conductivity at room temperature due to the presence of a variety of organic vapors with different chemical functionalities. The results demonstrated the viability of the phthalocyanine thin films as the active species for systems specifically designed for the monitoring of aromatic components in food. The majority of the work on vapor sensing has, however, been based upon quartz crystal microbalance (QCM) transduction techniques. The frequency shift Δf from the nominal resonance frequency f_0 of the crystal due to the change in the mass Δm due to the adsorption of the vapors is [52.143]

$$\Delta m(\mathrm{mg/cm^2}) = \frac{\Delta f(\mathrm{Hz})}{2.26\times10^{-6} f_0^2} \,. \qquad (52.3)$$

It was demonstrated that monomeric soluble transition-metal phthalocyanines R_4PcM ($R =$ *tert*-butyl or 2,2-dimethyl-3-phenyl-propoxy) used as sensitive coatings for quartz microbalance transducers show reversible interaction and high sensitivity for organic solvents with high boiling point [52.144]. Quartz microbalance devices coated with LB films of copper(II) tetra-(*tert*-butyl)-5,10,15,20-tetraazaporphyrin exhibit good sensing response to the vapor of benzene and toluene at room temperature [52.145]. Soluble tetrakis-hexyl- and dodecylthiophthalocyaninato nickel(II) and the corresponding Pd(II) and Ag(I) complexes were investigated as sensitive materials for the detection of organic solvent vapors using quartz microbalance and interdigitated capacitance transducers. Sensor responses were found to be reversible at room temperature with response times on the order of several seconds depending on the partition coefficients of the organic solvents in the phthalocyanine film [52.146]. LB films of lanthanide diphthalocyanines such as praseodymium, ytterbium as well as octa-*tert*-butyl praseodymium diphthalocyanines were known to display spectroscopic changes when exposed to tobacco smoke [52.147]. The sensitivity and partition coefficient for ethanol, dichloromethane, acetone, and *n*-hexane were found to be larger for the QCM sensor using thermally annealed ordered membranes of octa(13,17-dioxanonacosane-15-sulfanyl)-substituted mesomorphic nickel(II) phthalocyanine molecules than as-coated untreated films. It was found from Raman spectroscopy that molecules with saturated C–C bonds such as ethanol interact with phthalocyanine films predominantly by the formation of hydrogen bonds and the sensor response to π-bond-containing compounds such as acetone is the result of their π–π interaction with the conjugated phthalocyanine ring [52.148].

Changes in refractive index of lutetium bisphthalocyanine ($LuPc_2$) in the presence of ethanol, hexanal, *n*-butyl acetate, and acetic acid were successfully exploited to produce an optical-fiber sensor operating at the wavelength of 1310 nm, capable of monitoring changes up to 10 dB in the reflected optical power for the detection of an 88 mmol/l concentration of acetic acid [52.149]. UV-Vis spectra of spin-coated layers of 29H,31H-(2,4-di-*t*-amylphenoxy)phthalocyanine, Zn(II) tetra-4-(2,4-di-*t*-amylphenoxy)phthalocyanine, and Zn(II) tris-(2,4-di-*t*-amylphenoxy)-[4-(4-mercapto-phenylimino-methyl)-phenoxy] phthalocyanine were found to be sensitive to *tert*-butylamine, diethylamine, dibutylamine, 2-butanone, and acetic acid. The selectivity was determined by both the metal and the peripheral substituents [52.150]. The synthesis of specially designed bispthalocyanine derivatives for VOC detection has been reported and their UV spectra are selectively sensitive to vapors depending on both the metal and the peripheral substituents. This observation has led to the fabrication of an *electronic optical nose* using an array of as-manufactured sensors for the analysis of some volatile organic compounds (VOC) that are of interest in food analysis [52.151].

The electrical conductivity of sol–gel-derived spun hybrid cobalt porphyrin-SnO_2 thin films showed a fast and reversible response to methanol vapors; the highest responses were observed at 250 °C. The porphyrin

was thermally stable up to 300 °C within the SnO_2 matrix and enhanced the methanol detection sensitivity at lower working temperatures. The incorporation of porphyrin into SnO_2 film had no effect on the detection of CO [52.152]. Hybrid thin films were produced using porphyrin and phthalocyanine and the response of absorption bands of the blend systems to the presence of VOCs was different from that obtained with a single compound [52.153].

Toluene vapor sensing has been successfully demonstrated using LB films prepared from copper tetrakis-(3,3-dimethyl-l-butoxycarbonyl) phthalocyanine (CuPcBC), copper tetrakis-(3,3-dimethyl-1-neopentoxycarbonyl) phthalocyanine (CuPcNC), and nickel tetrakis-(3,3-dimethyl-1-butoxycarbonyl) phthalocyanine molecules. Exposure to toluene resulted in a partially reversible shift in the resonance depth and position of the SPR curves and toluene could be detected down to at least 50 ppm [52.154].

52.4.4 Phthalocyanine for Biosensing

Immobilization of proteins on different surfaces offers a range of applications in biomedical engineering and the development of biosensors. The quantity and mode of adsorption of proteins are important characteristics of biomedical materials used for artificial blood vessels, heart, kidneys, and other organs after their implantation in the human body. Different surface properties of materials like roughness, charge, coverage, hydrophilicity, hydrophobicity, steric hindrance as well as structure, sign and charge of proteins influence protein adsorptions on implanted materials. Serum albumin, the most abundant protein in serum, has several functions for example, regulation of colloidal osmotic pressure of blood, transportation of fatty acids etc. It is often used as a stabilizer of dilute enzyme solutions, protein standard, cell culture media additive, blocking agent of nonspecific protein binding in enzyme-linked immunosorbent assay (ELISA) and Western blot [52.155].

The binding between bovine serum albumin (BSA) and phthalocyanine derivatives has already been studied by the measurements of fluorescence emission [52.156]. The interaction BSA and octahexylthio-substituted Cu(II) phthalocyanine derivative ($(C_6S)_8PcCu$) molecules has recently been reported by us and it was found that the mechanism involved a range of noncovalent interactions including hydrogen bonding especially with the nitrogen in the phthalocyanine ring, ionic interactions between negatively charged residues and the core metal ion, as well as Van der Waal's interactions. Heat treatment produced changes in the orientation of phthalocyanine stacks, making a favorable contribution to the protein adsorption and surface coverage profile [52.157]. The kinetics were measured for adsorption of lysozyme, bovine serum albumin, immunogamma globulin (IgG) on $(C_6S)_8PcCu$ films using a quartz crystal microbalance with dissipation technique. Lysozyme was found to reach to saturation value faster than BSA and IgG. The adsorption of IgGs on the phthalocyanine surface was much slower than that of BSA [52.158].

Using a CCD camera, the surface plasmon resonance imaging (SPRI) method has been employed to investigate BSA protein adsorption on spin coated films of liquid crystalline copper octakishexylthio phthalocyanines ($(C_6S)_8PcM$). Figure 52.13b,c shows the variation of reflected light intensity of these films with wavelength. Phosphate-buffered saline (PBS) washing is found to increase the resonance intensity. As shown in 52.13a, the blue shift in the resonance wavelength from 635 to 628 nm occurred due to PBS washing. As

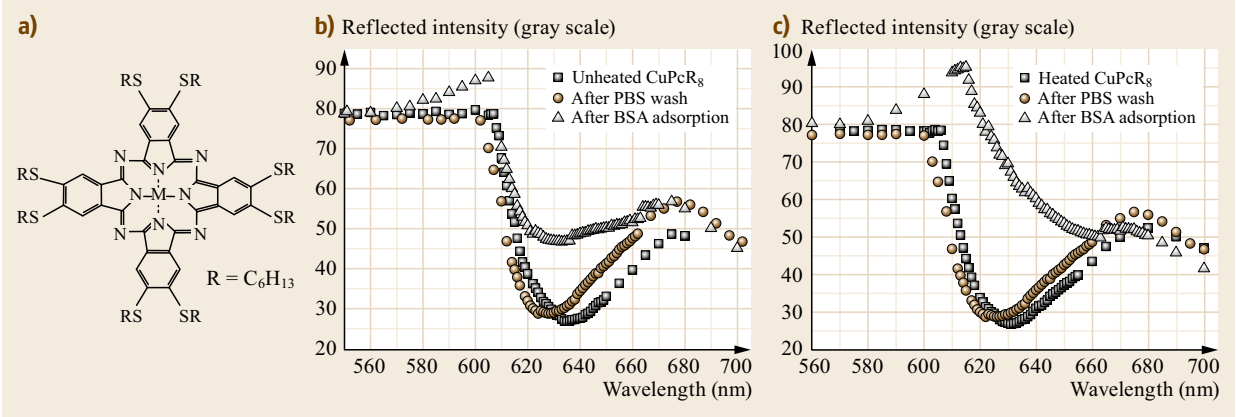


Fig. 52.13 (a) Molecular structure of octakisalkylthiophthalocyanine ($(C_6S)_8PcCu$). Reflected light intensity versus wavelength from SPRI measurements for BSA conjugation with $(C_6S)_8PcCu$ films as-deposited (b) and heated (c)

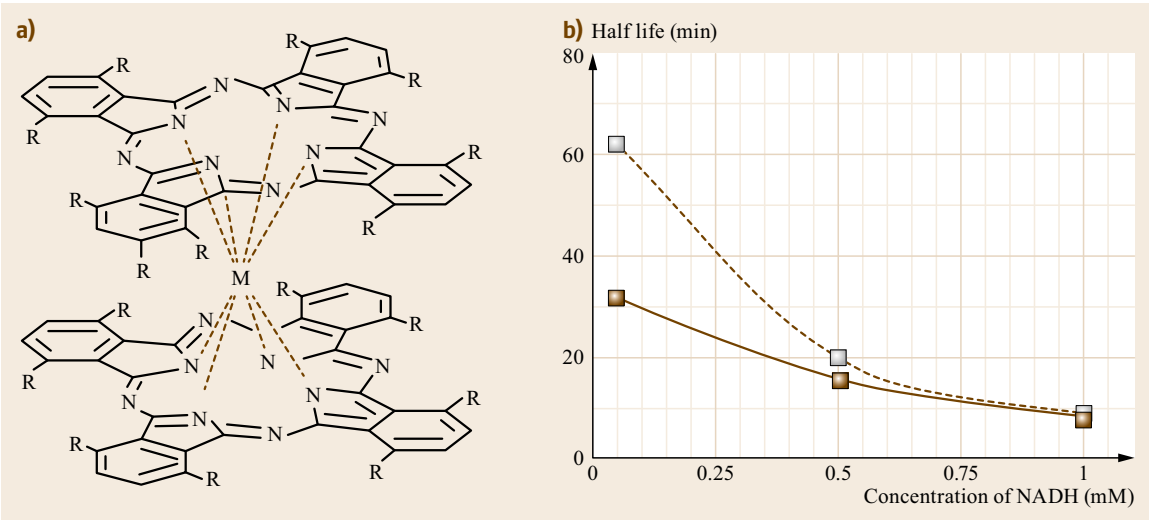


Fig. 52.14 **(a)** Chemical structure of C_8LuPc_2, M = Lu(III) and R = C_8H_{17}. **(b)** Dependence of oxidation half-life on NADH concentration for freshly prepared (*solid line*) and 3-month-old films (*dashed line*)

a result, salts of the PBS buffer are adsorbed on the film.

The reflected light color from the uneven film surface was visually yellowish green. The resonance was observed at 633 nm for the film after BSA adsorption. The mean intensity value for BSA adsorbed film increases up to 73% more than the phthalocyanine film. The spectrum was broader than that of the $(C_6S)_8PcCu$ overlayer because of the presence of damping. The shift in resonance wavelength from 622 to 663 nm was observed after BSA adsorption on the PBS-washed heated phthalocyanine film (Fig. 52.13c). The reflected color from the heat-treated film was visually more yellowish green than that from the as-deposited film. The surface coverage of the protein film and the amount of bounded BSA were higher than that obtained for the as-deposited $(C_6S)_8PcCu$ film. The calculated reflected intensity from the SPR image of the BSA film was increased by 170% at 631 nm compared to the phthalocyanine film. It is clear that the resonance curve is heavily damped for this film thickness. The BSA adsorbed onto the phthalocyanine film both before and after heat treatment was nonuniform as observed by SPRI as the substrate area was sufficiently large [52.159].

Spin-coated films, approximately 100 nm thick, of a newly synthesized bis[octakis(octyl)phthalocyaninato] lutetium(III) complex (chemical structure shown in Fig. 52.14) on ultrasonically cleaned glass substrates exhibit pronounced chemichromic behavior with potential application in health care. Oxidizing of a phthalocyanine film using Br_2 vapor or by applying a certain potential was shown to be effective by redshifts in optical absorption spectra. Using UV-Vis spectroscopy, a simple method was developed to detect NADH and vitamin C in a solution wetting the surface of an ITO-coated glass substrate modified with spin-coated electrochromic films [52.160]. The study showed satisfactory response times, and linear concentration ranges up to two orders of magnitude: 0.05−3 mM for NADH and 0.03−3.48 mM for vitamin C. The method of sensing NADH and vitamin C works well up to a concentration of 10^{-5} M and the complete recovery of the neutral bisphthalocyanine from the oxidized film via NADH reduction leads to the realization of developing reusable membranes. The reduction rate was greatly increased when redox biomolecules were added to the aqueous medium of water and $LiClO_4$. The reduction of the oxidized film in the presence of NADH and vitamin C was described in terms of first-order kinetics. The detection of NADH and vitamin C was found to be reproducible. As shown in Fig. 52.14b, the C_8LuPc_2 film remains sensitive for at least three months even when left in an open space.

52.5 Polymeric Materials

A polymer comprises of repeating molecules with the same chemical structure. The electrical conductivities of the intrinsically conducting polymer systems now range from 10^{-10}−10^{-5} S/cm. The common electronic feature of pristine conducting polymers is the π-conjugated system, formed by the overlap of carbon

p_z orbitals and alternating carbon–carbon bond lengths. Doping of the polymers can increase the conductivity to as high as 10^{+4} S/cm, a value comparable to that of a metal.

52.5.1 Conducting Polymers

Figure 52.15 shows the chemical structure of a few commonly used conducting polymers such as polyaniline, polypyrrole, and polythiophene for chemical sensors. In addition to the deposition techniques described earlier, conducting sensing filaments were produced using the scribbling and platinum-wire techniques [52.161]. Sensing membranes were also prepared by laser-induced chemical vapor polymerization [52.162].

LB films of stearic acid, a polyaniline oligomer (16-mer), and polypyrrole were deposited on the IDEs. Composite films of polyaniline and polypyrrole mixed with stearic acid in equal ratio by weight were also deposited on similar electrodes. Different brands of mineral water, tea, and coffee were sensed by impedance measurements on the films over the frequency range of

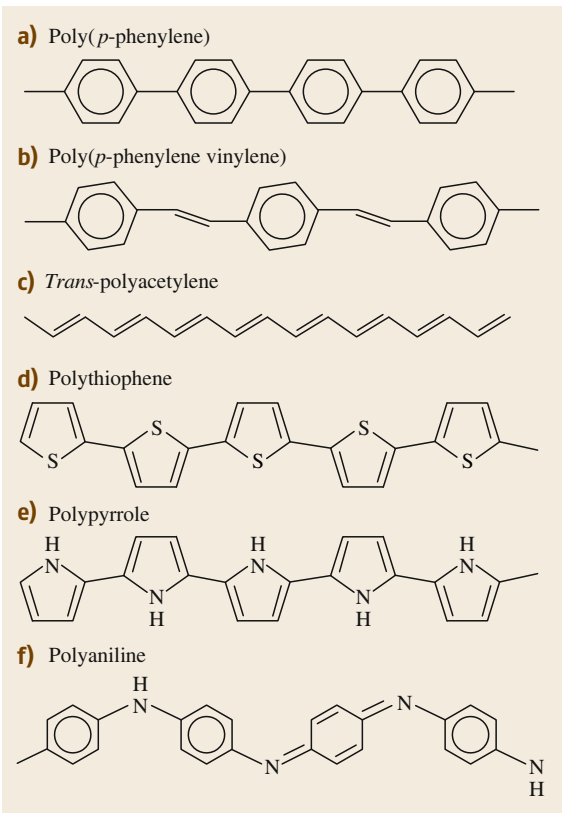


Fig. 52.15a–f Chemical structures of typical conducting polymers for sensors

20–10^5 Hz. The sensor arrays composed of these films were able to differentiate tastants below the human detection threshold [52.163]. The Brunauer, Emmett and Teller (BET) surface area for an amperometric polyaniline/Au/Nafion sensor prepared by the cyclic voltammetry (CV) method was found to be higher than structures prepared by the constant current (CC) method; the maximum sensitivity of CV-fabricated NO_2 gas sensors was 3.04 μA/ppm [52.164].

The type of doping is known to influence conducting-polymer-based sensors. Composite polyaniline emeraldine salts doped with an inorganic metal complex dopant, chromium(III) trioxalate (CTO) were produced by the chemical oxidative polymerization method. Polyanilines were also electrochemically doped with naphthalene-1,5-disulfonic acid (NSA). The conductivities of these materials were monitored by using the four-probe method when pellets of these materials were exposed to several saturated vapors of organic solvents for 30 min; the results are summarized in Fig. 52.16. The conductivity of the CTO-doped pellets was reduced by 18% on exposure to CH_2Cl_2 whereas the exposure had no effect on NSA-doped material. The response of undoped composite polyaniline salt to CH_2Cl_2 was also very small. Conversely, NSA-doped pellets showed a significant change in conductivity on exposure to CCl_4, while only very small changes were observed for the two other polymers. The responses to dimethylsulfoxide (DMSO) and *N*-methylpyrrolidone (NMP) were consistently higher for all three materials than the responses to the solvents

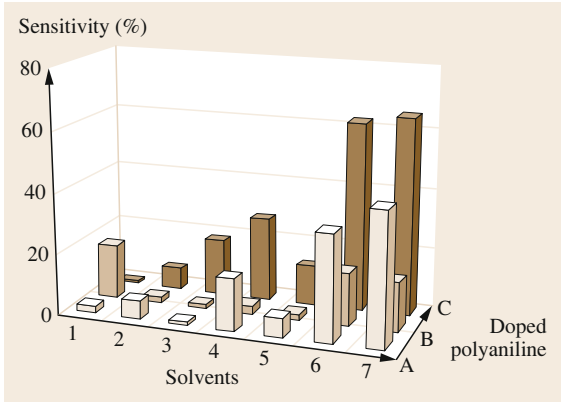


Fig. 52.16 Effects of vapor of: (1) dichloromethane (CH_2Cl_2), (2) chloroform ($CHCl_3$), (3) carbon tetrachloride (CCl_4), (4) acetonitrile (CH_3CN), (5) methanol (MeOH), (6) dimethylsulfoxide (DMSO), and (7) *N*-methylpyrrolidone (NMP) on the conductivity of: (A) pristine polyaniline emeraldine salt, and the salt doped with (B) chromium(III) trioxalate (CTO), and (C) naphthalene-1,5-disulfonic acid (NSA)

CH_3CN and methanol (MeOH). It became apparent that the interaction of solvents with the polyaniline salt became stronger with increasing polarity of the solvents. The response was reversible for all three materials. The nature of the dopant in the material and the consequent change in polarity, especially at the dopant site, and the resultant effect on π-electron delocalization and conductivity are primarily responsible for the characteristic sensing ability of the materials [52.165]. Extruded polymer blends containing polyaniline doped with dodecyl benzene sulfonic acid were dispersed in polystyrene (PSt) matrix filaments and the electrical conductivity of these filaments increased by a few orders of magnitude when exposed to a homologous series of alcohols, including methanol, ethanol, and 1-propanol. The effect was reproducible and recovery was satisfactory. The sorption of analyte molecules was believed to facilitate conduction through the insulating PSt moieties by enhancing charge-carrier mobility through hopping processes between adjacent polyaniline particles [52.166].

Multilayered thin films of polypyrrole were prepared by modifying a preformed Langmuir–Blodgett (LB) film of stearic acid by exposure to gaseous reactants such as hydrochloric acid and pyrrole vapors. A change in the electrical resistance of the film was observed when the film was exposed to polar molecules such as methanol, ethanol, and acetone, while the effect of nonpolar molecules on the resistivity of the film was found to be smaller. The response of LB films to exposure to ethanol was faster than that of electrochemically deposited polypyrrole, possibly due to the high ratio of surface area-to-bulk volume in the LB film [52.167]. Changes in mass and conductivity of polypyrrole thin films were measured to sense the primary alcohols [52.168].

Attempts have been made to integrate polyaniline optical sensor elements with optical communication technology for remote sensing by using 1300 nm as the operating wavelength. A 1% increase in transmission was observed at 1300 nm for electrochemically prepared polyaniline films when exposed to gaseous ammonia levels as low as 6 ppm at 50% relative humidity (RH). Different concentrations of ammonia were read by calibrating the sensors at intervals of 15 s. The response time was found to be insensitive to relative humidity variations in the range 30–70% [52.169]. Certain specially prepared polyaniline films showed partially reversible absorbance changes in the wavelength range 620–728 nm when exposed to 50–100 ppm of ozone at room temperature. The maximum sensitivity was attained at 620 nm. The cost-effective commercial exploitation of this effect was possible since the wavelength range was compatible with implementation on low-cost plastic optical fibers and small light sources [52.170].

Efforts have been invested in the use of information technology techniques to produce smart sensors. The sensor system, consisting of an acoustic two-port resonator operating at 433.92 and 380.8 MHz was configured as a frequency oscillator, including an integrated electronic module. The polymeric membranes were tested at room temperature for response to NO_2, NO, NH_3, CO, CH_4, SO_2, and H_2S in N_2. Responses to relative humidity (RH), and organic vapors (ethanol, acetone, and ethyl acetate) were also monitored. The sensitivity was found to be high. Principal component analysis (PCA) was performed to distinguish between different vapors of low concentrations [52.171]. Fluorescence emission from polymer-immobilized dye molecules on the multifiber tips was studied on exposure to organic vapors. Temporal responses were found to depend upon chemical nature (for example, polarity, shape, and size) of both the vapor and the polymer; these were used as input signals to train a neural network for vapor recognition. The system was able to identify individual vapors at different concentrations accurately [52.172]. Similar work was reported using the solvatochromic dye, Nile red. The dye was immobilized within various polymers. The substrate played a role in determining the sensitivity of the sensor. The sensitivity attained with microstructure glass (MSG) substrates was seven times greater than that with SU-8 photoresist-coated glass substrate and a 50% faster recovery was also achieved with the MSG substrates. The MSG sensor array was able to fingerprint the response for separate analytes with a high degree of repeatability. Using pattern-recognition techniques, sensor arrays were adaptable for gas identification and discrimination [52.173].

52.5.2 Ion Sensing

Water purification requires calibration of metal-ion contents. Polymeric ion-selective electrodes (ISEs) for Pb^{2+} incorporating *N*,*N*′-bis(salicylidene)-2,6-pyridinediamine with 2-nitrophenyl octyl ether and 50 mol% lipophilic additive were reported to have rapid response and excellent selectivity towards lead ions over other interfering metal ions [52.174]. *N*,*N*′-bis(5-methyl salicylidene)-*p*-diphenylene methane diamine formed a complex with Pb^{2+} and its selectivity was high, possibly due to the fact that the distance between the two nitrogen atoms and the position of the two hydroxyl groups is matched with the size of Pb^{2+} ions. ISEs based upon a dispersion of this Schiff base compound into the polyvinylchloride (PVC) membrane were characterized by a fast response time, a wide

linear dynamic range, and a fair selectivity coefficient with a slope of 29.4 mV per decade. The sensor had a low detection limit and an active life of three months without displaying considerable divergence in potentials [52.175]. A cleaned gold microelectrode was soaked in the freshly prepared solution of *o*-amino thiophenol (*o*-AT) and the metal nitrate of Cu, Hg, and Pb in *N*,*N*-dimethylformamide (DMF) for 4 h at room temperature. Simultaneous interaction took place between the thiophenol group of *o*-AT and the gold surface and its thiophenol and amino groups and M^{2+}. The ratio of 4 : 1 was found to be appropriate for *o*-AT:M^{2+} for the self-assembly of the *o*-AT on the gold surface and the formation of recognition cavities. The prepared ISEs showed specific selectivity to the template metal ions in mixed solutions containing the three heavy-metal ions. The limits of detection for ISEs were found to be 1.46×10^{-8} M, 3.73×10^{-8} M, and 4.34×10^{-8} M for Hg^{2+}, Cu^{2+} and Pb^{2+} ions, respectively [52.176]. Membranes containing five-layer LB films of tetracarboxylic perylene derivative and polypyrrole molecules were able to detect trace levels of Cu^{2+} ions in water [52.177]. The thiol-Cu-SAM particles were doped by the adsorption of thioxylenol or decanethiol into the porous polypyrrole film surface, resulting in an improvement in the sensitivity of the film to NH_3 [52.178]. Water-soluble regioregular polythiophenes containing acid C side chains were found to be suitable for the development of new Zn^{2+}, Mn^{2+}, and Cd^{2+} sensors [52.179].

52.5.3 Examples of Other Polymeric Sensors

A coordination polymer poly(CuMBSH) was formed via reaction of the bifunctional amphiphilic ligand, 5,5′-methylenebis (*N*-hexadecylsalicylideneamine) and copper ions at the water subphase. The response of LB films of poly(CuMBSH) to exposure of benzene, toluene, ethanol, and water was monitored by SPR measurements. The film became swollen, possibly due to the diffusion of solvent vapor into the polymer film. Poly(CuMBSH) is regarded as a relatively nonpolar material and its interactions were expected to be greater with vapors having a smaller dipole moment; the sensitivity increased with smaller dipole moment and higher refractive index of the vapor. The refractive index of the film increased because of an increase in the film density with the vapor filling the free volume inside the film. Similar results were obtained from admittance spectroscopic measurements on the same organic systems [52.180, 181].

The electrical properties of poly(methyl methacrylate) polymer (PMMA) films have been exploited for sensing applications (Fig. 52.17a shows the chemical formula). An increase in resistance of up to three orders of magnitude was observed for composite thin films of PMMA with carbon nanotubes when exposed to dichloromethane, chloroform, and acetone. The sensing mechanism is explained on the basis of volume expansion and polar interaction of various vapors on the nanotube surface [52.182, 183]. SPR measurements were made to investigate the room-temperature response of spun PMMA films to benzene, toluene, ethyl benzene, and *m*-xylene (BTEX) vapors under dynamic conditions. The response of PMMA to benzene was fast and reversible. The dissolution of benzene in the bulk of the polymer film formed a heterogeneous layer, primarily giving rise to swelling of the film and the value of the partition coefficient was estimated to be 112. Figure 52.18 shows that the increase in angular shift per ppm was linear with decreasing dipole moment of the solvent vapors. Benzene is a nonpolar solvent and toluene, ethyl benzene, and *m*-xylene are expected to show stronger solvation than benzene. Dissociated solvent ions may have formed a shell of bound ions and, as a result, the effective radius for dissolution in the polymer matrix would have increased. This dissociation might not easily occur for benzene, producing

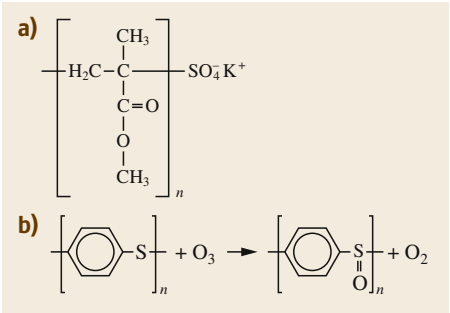


Fig. 52.17a,b Chemical structures of **(a)** poly(methyl methacrylate) derivatives and **(b)** reaction of polyphenylsulfide molecules with ozone

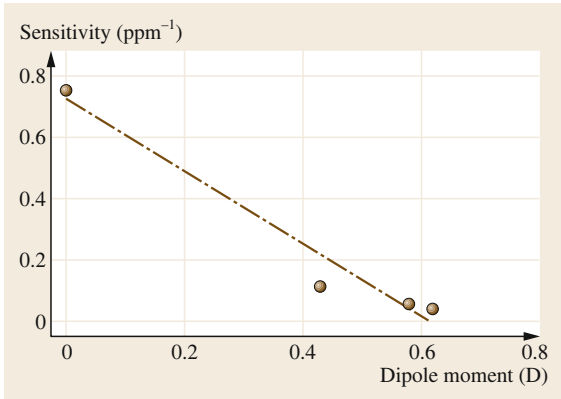


Fig. 52.18 Response of spun PMMA films on exposure to benzene (0 D), toluene (0.43 D), ethyl benzene (0.58 D), and *m*-xylene (0.62 D)

the highest SPR response [52.184]. A sensor has been made from a permselective poly(dimethylsiloxane) (PDMS) hollow-fiber membrane, showing a fast response time, very high oxygen permeability, and optimized oxygen/nitrogen selectivity. This sensor was suitable for monitoring the oxygen content in the outlet stream generated by medical oxygen concentration units [52.185].

Single-polyaniline films generally suffer from poor mechanical stability and polyaniline-blend films were prepared for practical applications using PMMA and PSt. The presence of water vapor, functioning as the interference gas, led to the lowering of the NH_3 response of the sensor in a humid atmosphere. The humidity effect was larger for PMMA than the PSt blend film. The difference was, however, not as large as expected from the water-sorption ability of PMMA [52.186]. The detection limits of ISE membranes with solid contacts are generally lower than a traditional liquid inner contact because of the diminished ion fluxes. Poly(methyl methacrylate)/poly(decyl methacrylate) (PMMA/PDMA) copolymer was employed as a solvent-cast membrane matrix on a layer of poly(*n*-octyl)thiophene (POT) deposited on Au as the internal contact. Ca^{2+}- and Pb^{2+}-selective electrodes were not sensitive to O_2 and the presence of a water film between the membrane and the internal contact was not found. The limits of detection were better by an order of magnitude and the response at low concentrations was much faster than for the liquid-contact electrode. For example, the drift was found to be smaller than 0.4 mV/min after 2 min when the concentration was increased from 10^{-9} to $10^{-7.7}$ M [52.187].

An amperometric hydrogen sensor was proposed using a proton-conducting solid polymer electrolyte membrane (PEM), a blend of palladium and platinum as the anode, and platinum as the cathode. The sensor operated as a fuel cell (H_2/Pd-Pt//PEM//Pt/O_2) and the short-circuit current was found to be linearly related to the hydrogen concentration [52.188].

It has recently been reported that thin films of polyphenylsulfide (PPS) (commercially termed noXon) act as efficient scrubbers of ozone from atmospheric ambient through the oxidation of the PPS molecules [52.189]. Thin films of this compound are found to be a useful material for the fabrication of sensing membranes suitable for the detection of ground-level ozone at concentrations of a few ppm or even lower [52.190, 191]. A 1 h exposure to 2 ppm of O_3 was shown to cause at least a 1.1% change in the refractive index of the film [52.192]. As shown in Fig. 52.17b, oxidation of the PPS polymer by ozone exposure results in an irreversible increase in the index of refraction of the polymer [52.193].

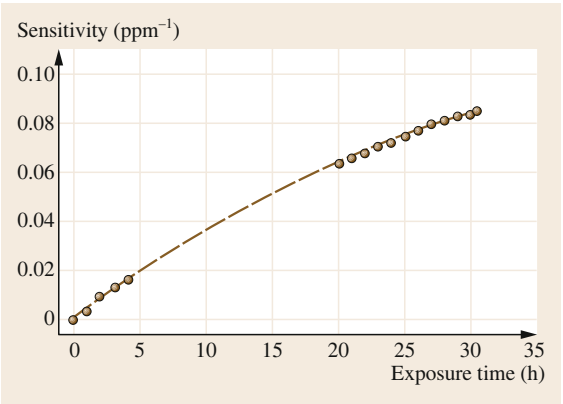


Fig. 52.19 Time-dependence behavior of ozone sensitivity of PPS films. Two sets of hourly readings, separated by 16 h of continuous exposure without SPR measurements (*dashed line*)

Figure 52.19 shows the results of SPR measurements on PPS films when exposed to 1.5 ppm ozone. The increase in the refractive index Δn was found to be linear over the first few hours of exposure with an estimated slope of 10^{-4} refractive index units (RIU)/min. The sensitivity was defined as the ratio of Δn to the concentration of ozone. The interaction of O_3 with PPS films is believed to be an accumulative effect and such mechanisms can therefore serve as both a scavenger as well as an indicator of ozone in a certain ambient. The density of the binding of oxygen to a PPS film after 5 h of exposure to 2 ppm ozone was estimated from QCM measurements to be about 5×10^{18} cm^{-3}.

Electrospun nanofibers with diameters of 100–400 nm were deposited on the surface of a QCM by electrospinning homogenous blend solutions of crosslinkable poly(acrylic acid) (PAA) and poly(vinyl alcohol) (PVA). The NH_3 sensing properties were mainly affected by the content of the PAA component in the nanofibrous membranes, the concentration of NH_3, and relative humidity. The sensitivity of nanofibrous-membrane-coated QCM sensors was also much higher than that of continuous-film-coated QCM sensors [52.194].

Molecularly imprinted polymers (MIPs) are now well recognized as robust, cost-effective, and highly durable synthetic receptors for specific targets having variable sizes and shapes. MIPs which provide interaction points and a coordination sphere around the template molecule have applications ranging from liquid chromatography to assay and sensor technologies and drug delivery [52.195].

The synthesis method of MIPs includes the selection of the required target molecule used as the molecular template, and embedded in the presence of

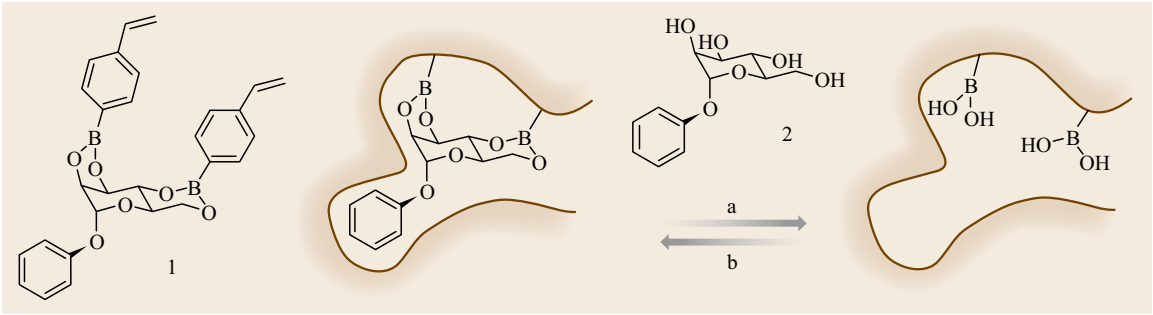


Fig. 52.20 Schematic representation of a cavity obtained by polymerization (1) starting molecule (2) formation of cavity (a) and (b) route to successive chemical formation

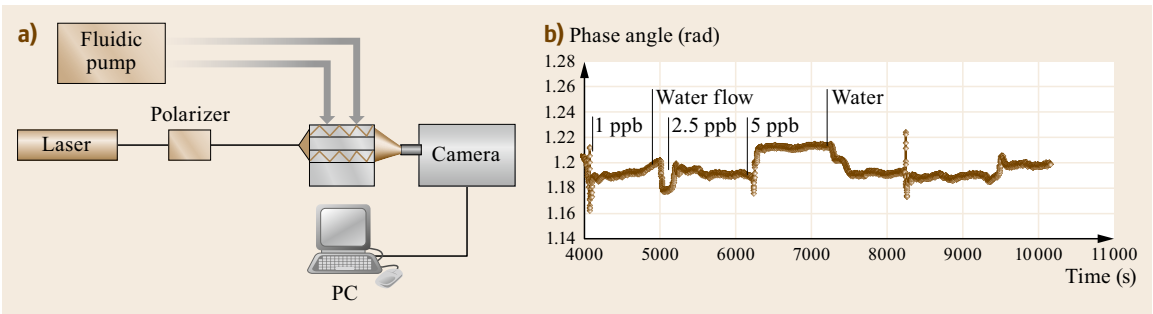


Fig. 52.21 **(a)** Schematic diagram of dual polarization interferometer (DPI) **(b)** DPI time trace transverse electric phase angle in radians upon adding 1, 2.5 and 5 ppb atrazine to atrazine MIP immobilized on a DPI chip

functional monomers, whereby the monomers interact with and engross into sites of the template forming a complex followed by further polymerization. After the completion of the polymerization the functional sites of the molecule template are held in position and surrounded by the highly crosslinked polymeric structure, followed by the removal process of the template molecule which subsequently results in the formation of cavities, specific in size, shape, and chemical functionality of the target molecule (Fig. 52.20). These imprinted functional cavities are then capable of identifying, rebinding, and quantifying the target molecules with excellent degrees of selectivity and sensitivity and are therefore an ideal immobilization method of sensor layers [52.196].

For the specific detection and quantification of the herbicide atrazine, a triazine molecule is used as the template molecule and reacted with the monomer methacrylic acid (MAA) in addition to the crosslink agent ethylene glycol dimethacrylate (EGDMA) initiated by azobisisobytyronitrile (AIBN) during the synthesis process. The polymerization is driven and influenced by heat, whereby the triazine molecule is removed by Sox extraction using acetic acid and chloroform. The optimization process of the final MIP influences the guest capacity factor, as some of the functional sites present remain inaccessible due to the presence of residual template molecules, excessive crosslinking restricting the diffusion or deformation of the network during the extraction process by swelling or relaxation caused by the solvent. Under these influences the guest capacity for MIPs is limited with typical values of 0.1 mg per 1 mg of MIP, however advanced polymeric systems with trifunctional crosslinker agents or higher offer significant increases in capacity which in turn enhances the recognition and decreases possible retention times. The selection of the optimal solvent also plays an essential role in the type of interaction involved and consequently determines its polarity and dissociating energy, capable of influencing the morphology of the material during synthesis, including the size, shape, and quantitative distribution of the cavities. It is therefore crucial during analysis stages to incorporate a solvent similar to the synthesis solvent involved in the production of the original network in order to exclude the possibility of restructuring the effective morphology and to obtain the selectivity to the initial degree [52.197].

Atrazine is a man-made chemical which is very widely used as a herbicide. It can be harmful to wildlife, is long-lasting in the environment, and a potential groundwater contaminant. Commercially available

molecularly imprinted polymer (MIP) is dispersed into the chloroform solution of poly(methyl methacrylate) (PMMA) to produce the sensing membrane for the herbicide atrazine. The interaction of MIP/PMMA matrices with atrazine was investigated using the dual polarization interferometric (DPI) technique (as shown schematically in Fig. 52.21a). The interferometer is able to interrogate the sensor/sample interface by coupling transverse electric (TE) and transverse magnetic (TM) waves representing light parallel and perpendicular to the plane of the chip [52.198]. A thin film of MIP dispersed in PMMA solution was spin-coated onto the silicon oxynitride surface of the DPI chip. Distilled water was used as the running buffer and this was switched to 400 µl injections of 1 ppb of atrazine in solution. The running buffer was switched back to distilled water following completion of each injection loop. As shown in Fig. 52.21b), it was possible to measure the change in phase due to increasing exposure levels of atrazine. The atrazine effect was effectively removed by switching the running buffer back to distilled water. This suggested that the detection of atrazine by the MIP/PMMA thin film integrated to the DPI chip was reversible. However, given that the bound atrazine could be removed by a simple water wash suggests that the herbicide was not strongly bound to the MIP trapped in the PMMA film. Normally a more stringent elution stage is required to remove the atrazine from the MIP.

52.6 Cavitand Molecules

Thin films of some novel cavitand compounds such as crown ethers [52.199] and calixarenes [52.200, 201] can form inclusion complexes with some organic guest molecules. This effect is being extensively exploited for the development of sensors for organic vapors [52.202, 203]. Thin films can be prepared using LB film deposition, spin coating, and self-assembly techniques [52.204–207]. The nanoporous flexible structure of thin LB films of amphiphilic calixresorcinarene derivatives were found to provide a suitable matrix for the incorporation of aromatic molecules [52.208]. This adsorption process was very fast, and full recovery of the film was observed after flushing with clean air. It has to be pointed out, however, that the detected vapors were of a high concentration (a few percent by volume) and the adsorption was not selective since all vapors studied yielded a similar response. However, this may lead to the development of a sensor for explosives, where specificity is not considered relevant. These effects are attributed to weak and nonspecific interactions between the guest molecules and the calixarene LB film. It was also shown that the adsorption of organic vapors occurs in the whole bulk of the LB films, and that the number of adsorbed molecules is much higher than the number of calixarene molecules [52.209]. The adsorption mechanism involved the swelling of the film and even condensation of the adsorbate within the film. The swelling of the film was confirmed by independent measurements using ellipsometric and surface plasmon resonance (SPR) techniques [52.210]; the adsorption mechanism still remained unclear. For all the studied analytes, the number of adsorbed molecules was found to be much larger than that expected from the geometrical dimensions of the intrinsic calixarene cavity and the empty space between the molecules. This discrepancy was explained by either film swelling or the condensation of vapor molecules inside the film, or both [52.208]. Spun films of calixresorcinarenes, however, were found to be homogeneous, and the porosity of the film was similar to that of LB-deposited films. Both LB and spun films of comparable thickness exhibited similar SPR response on exposure to toluene vapor [52.62]. Prolonged irradiation of the CA films with a focused laser beam caused an initial increase in film sensitivity to various organic vapors [52.211].

Figure 52.22 shows the chemical structure of an amphiphilic calixresorcinarene derivative with several different types of substituents. Exposures of layers of CAs to vapors of benzene, ethylbenzene, toluene, and *m*-xylene in low concentrations produced changes in the optical properties, i. e., the film thickness and refractive index of the layers. Fast response and recovery processes have been observed, with response times as short as a few seconds. The normalized SPR responses are shown in the three-dimensional (3-D) diagram in Fig. 52.23. It is apparent that the interaction of the analyte was, in general terms, specific to the thin film of a particular CA derivative. The adsorption mechanism was interpreted in terms of the accumulation of the vapor molecules in the liquid phase inside the film matrix, probably caused by capillary condensation in the porous structure of the CA films [52.212]. These LB films had a characteristic pore size of about 1 nm. Condensation of vapor occurred inside capillaries similar to the porous structure of calixarene films at pressures lower than their saturated pressure at a certain temperature. The saturated vapor pressures of ethylbenzene and *m*-xylene are similar and therefore their exposure to the CA derivatives produced similar effects. Conversely, benzene, with the highest saturated

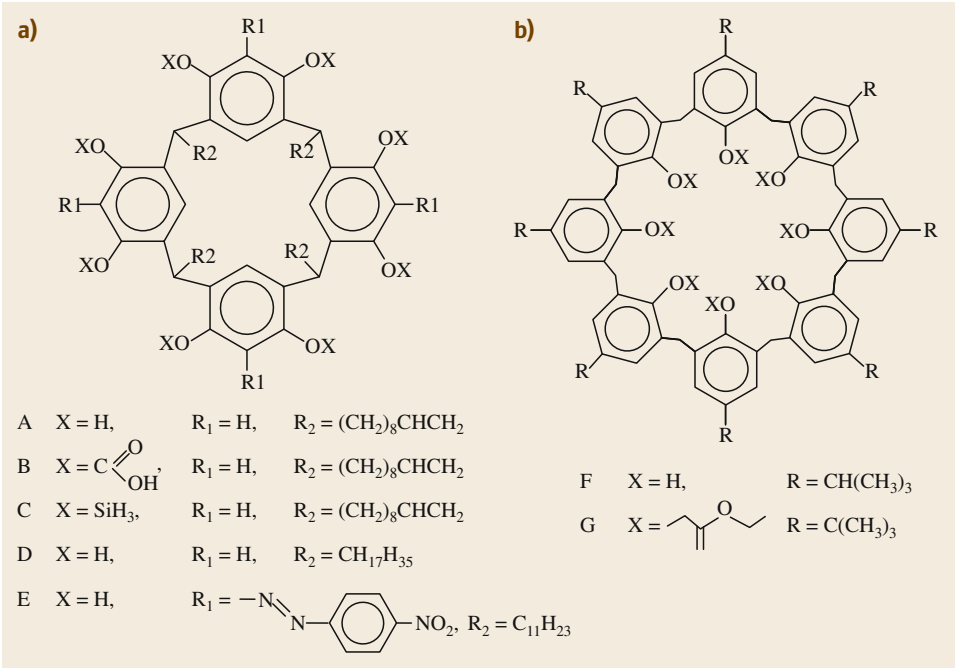


Fig. 52.22a,b Chemical formulae of calixresorcinarene derivatives

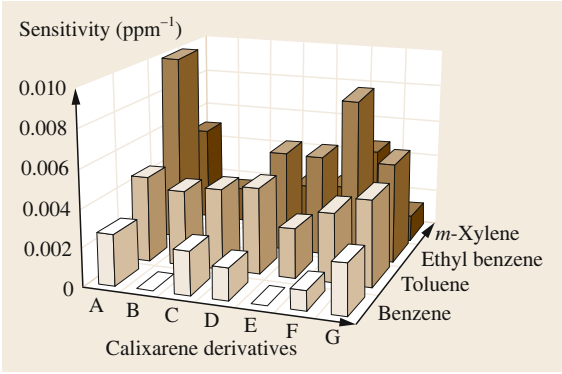


Fig. 52.23 A 3-D representation showing the response of different calixarene derivatives to different organic vapors (benzene (B), *m*-xylene (X), toluene (T), and ethylbenzene (E)) at a concentration of 375 ppm

vapor pressure, yielded the smallest detectable SPR signal.

Composite LB films of amphiphilic CA and PPS were exposed to 2 ppm of O_3 [52.190]. The competition between the permanent oxidation process of the PPS polymer by the adsorbed O_3 molecules and the release of O_3 from the CA matrix was found to depend on the molar ratios of the CA and PPS compounds. Recovery became slower with lower CA proportions in the film matrix.

The conductivity of the CA derivative is intrinsically low. CA films were deposited on the gate between the source and the drain of a charge-flow transistor and the turn-on response for the transistor upon exposure to organic solvent was recorded. The increase in the membrane conductivity is partially attributed to condensation of the vapors in the highly microporous membrane even below the saturation vapor pressure and partially to the effect of the polar analyte molecules complexing inside and between the OH groups of the cavities. The technique offers several advantages over existing methods:

1. Immunity to water vapor, because of the hydrophobic nature of the membrane
2. No catalytic poisoning of the membrane, as is commonly observed for doped SnO_2
3. No accumulative effects, which are responsible for baseline drift in quartz-crystal-type sensors, are observed
4. No porous metal layer, the adhesion of which may be prone to degradation, is needed on top of the membrane [52.213].

52.7 Concluding Remarks

The scope of organic materials for chemical sensing is huge and it is impossible to provide a complete description of these materials. Electronic artificial noses are being developed as systems for the automated detection and classification of odors, vapors, and gases. These instruments consist of three main components:

1. An array of chemical membranes
2. Electronic circuitry for data acquisition, signal processing, and display
3. A pattern-recognition algorithm.

Up-to-date information on applications of materials for electronic noses and tongues is available in a recent review [52.214]. Problems associated with human senses are many: individual variability, impossibility of online monitoring, subjectivity, adaptation, infections, harmful exposure to hazardous compounds, and mental state. These limitations can be overcome by developing intelligent sensor systems. Gas sensors tend to have very broad selectivity, responding to many different analytes of varied concentrations. The electronic nose/tongue, therefore, offers a definite advantage in many applications in the food and utility industries, the health care and medical care sectors, and security services. The demand for advanced electronic noses will continue to stimulate the search for novel sensing materials.

Mathematical and computational tools are expected increasingly to play a part in the design of advanced electronic-nose systems. An electronic nose has been developed to monitor breathing air in human habitats. A molecular modeling study was undertaken to investigate the interactions between resistive sensors of a polymer–carbon black (CB) composite and analytes. Poly(4-vinylphenol), polyethylene oxide, and ethyl cellulose were considered for modeling, based on their stereoisomerism and sequence isomerism. The CB was modeled as uncharged naphthalene rings with no hydrogen. Molecular mechanical and molecular dynamics techniques were employed and the equilibrium composite structure was constructed by inserting naphthalene rings into the polymer matrix. The radial distribution profiles produced information on the composite microstructure. The sensor response was predicted in terms of the interaction energies of the analytes with the composites. Studies included both inorganic and organic analytes [52.215].

Research efforts are also growing in the mineralization of VOC pollutants to innocuous compounds [52.216]. Visible-light-assisted removal of organic pollutants by photocatalytic action of dye-sensitized TiO_2 surfaces offers several attractive advantages. Firstly, the nanoscale device architecture is versatile with several methods of solar energy conversion whilst producing a chemical fuel in the form of pure H_2. Secondly, pollutant species and concentrations of mixtures in industrial waste streams can be identified [52.217]. Results so far reported are encouraging. For example, 40–75% degradation of aromatic pollutants such as phenol, chlorophenol, trichloroethylene, and surfactants was achieved with TiO_2 surfaces modified by methylene blue and rhodamine B dyes after less than 5 h of irradiation with a 150 W xenon lamp [52.218]. Individual VOCs including methylene chloride, ethanol, benzene, acetone, xylene, and isopropanol produced unique signatures as they were oxidized on the sensor surface [52.219]. Interest in hybrid hetero-supramolecular structures will therefore remain alive for the foreseeable future.

Acknowledgments. The author is grateful to the Engineering & Physical Research Council (UK) and the European Union for financial support of their work over the last 20 years. Gratitude is also due to Dr. S. C. Thrope of the Health Safety Executive (UK), Dr. M. Hofton of TQ Environmental Plc., Prof. K. C. Thompson of ALControl, and Dr. F. A. Grunfeld of NIMA Technology for their collaboration and support. Prof. M. J. Cook of the University of East Anglia has remained a generous provider of novel phthalocyanine compounds for sensing experiments over the years. The contributions from Drs. R. Capan, O. Omar, and A. V. Nabok are acknowledged with gratitude. Above all, the author is enormously indebted to his parents Chittarajan and Sarashi of Belgram (India) for stimulating his interest in academic career in science.

References

52.1 W. Gopel: Sensors Actuat. B **18/19**, 1 (1994)
52.2 M.C. Petty: Biosens. Bioelectron. **10**, 129 (1995)
52.3 M.J. Cook: Chem. Rec. **2**(4), 225 (2002)
52.4 L. Alcacer: *Conducting Polymers Special Applications* (Reidel, Dordrecht 1987)
52.5 E.B. Feresenbet, E. Dalcanale, C. Dulcey, D.K. Shenoy: Mol. Cryst. Liq. Cryst. **397**, 585 (2003)
52.6 N.K. Hunt, B.J. Marinas: Water Res. **31**, 1355 (1997)

52.7 T. Oshima, K. Sato, H. Terauchi, M. Sato: J. Electrostat. **42**, 159 (1997)
52.8 N. Carmona, M.A. Villegas, J.M.F. Navarro: Sensors Actuat. A Phys. **116**(3), 398 (2004)
52.9 N. Carmona, M.A. Villegas, J.M.F. Navarro: Thin Solid Films **458**(1/2), 121 (2004)
52.10 C.L. Baban, Y. Toyoda, M. Ogita: Jpn. J. Appl. Phys. **143**(10), 7213 (2004)
52.11 I. Toma-Dasu, U. Dasu, M. Karlsson: Phys. Med. Biol. **49**(19), 4463 (2004)
52.12 S.E.J. Williams, P. Wootton, H.S. Mason, J. Bould, D.E. Iles, D. Riccardi, C. Peers, P.J. Kemp: Science **306**(5704), 2093 (2004)
52.13 F. Bender, C. Kim, T. Mlsna, J.F. Vetelino: Sensors Actuat. B Chem. **77**(1/2), 281 (2001)
52.14 B. Onida, L. Borello, S. Fiorilli, B. Bonelli, C.O. Arean, E. Garrone: Chem. Commun. **21**, 2496 (2004)
52.15 S. Sen, K.P. Muthe, N. Joshi, S.C. Gadkari, S.K. Gupta, M. Jagannath: Roy, S. K. Deshpande, J. V. Yakhmi: Sensors Actuat. B, Chem **98**(2/3), 154 (2004)
52.16 B.H. Timmer, K.M. van Delft, R.P. Otjes, W. Olthuis, A. van den Berg: Anal. Chim. Acta **507**(1), 137 (2004)
52.17 M. Beekmann, G. Angellet, G. Megie, H.G.J. Smith, D. Kley: J. Atmos. Chem. **19**, 259 (1994)
52.18 L. Xie, T.J. Lu, H.Q. Yan: Electroanal. **10**, 842 (1998)
52.19 T. Becker, L. Tomasi, C. Bosch v. Braunmühl, G. Müller, G. Sberveglieri, G. Fagli, G. Fagli: L. Tomasi, C. Bosch v. Braunmühl, G. Müller, G. Sberveglieri, G. Fagli, E. Comini, Sensors Actuat. A Phys **74**, 229 (1999)
52.20 H. Schulz, G.B. De Melo, F. Ousmanov: Combust. Flame **118**, 179 (1999)
52.21 B. Zielinska, J.C. Sagebiel, G. Harshfield, A.W. Gertler, W.R. Pierson: Atmos. Environ. **30**, 2269 (1996)
52.22 F.B. Reig, J.V. Adelantado, V.P. Martinez, M.C. Moreno, M.T. Carbo: J. Molec. Struct. **480/481**, 529 (1999)
52.23 W. Groves, E.T. Zellers, G.C. Frye: Anal. Chim. Acta **371**, 131 (1998)
52.24 W.A. Groves, E.T. Zellers: Ann. Occup. Hyg. **45**, 609 (2001)
52.25 N. Kasai, I. Sugimoto, M. Nakamuro, T. Katoh: Biosens. Bioelectron. **14**, 533 (1999)
52.26 G. Sberveglieri: Sensors Actuat. B **23**, 103 (1995)
52.27 D. Manno, A. Serra, M. Di Giulio, G. Micocci, A. Tepore: Thin Solid Films **324**, 44 (1998)
52.28 N. Taguchi: Japanese Patent Application No. 45-38200 (1962)
52.29 C.D. Kohl, A. Eberheim, P. Schieberle: Tech. Mess. **71**(5), 298 (2004)
52.30 A. Eberheim, D. Kohl, P. Schieberle: Phys. Chem. Chem. Phys. **5**(23), 5203 (2003)
52.31 J. Gutierrez, J. Getino, M.C. Horrillo, L. Ares, J.I. Robla, C. Garcia, I. Sayago: Thin Solid Films **317**, 429 (1998)
52.32 M. Graf, D. Barrettino, S. Taschini, C. Hagleitner, A. Hierlemann, H. Baltes: Anal. Chem. **76**(15), 4437 (2004)
52.33 E. Comini, G. Faglia, G. Sberveglieri, Z. Pan, Z.L. Wang: Appl. Phys. Letts. **81**(10), 1869 (2002)
52.34 A. Gramm, A. Schutze: Sensors Actuat. B Chem. **95**(1-3), 58 (2003)
52.35 O.K. Varghese, D. Gong, M. Paulose, K.G. Ong, C.A. Grimes: Sensors Actuat. B Chem. **93**, 338 (2003)
52.36 A.K. Prasad, P.I. Gouma: J. Mater. Sci. **38**(21), 4347 (2003)
52.37 K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, Y. Zhang, S.V. Dubonos, I.V. Grigorieva, A.A. Firsov: Science **306**(5696), 666 (2004)
52.38 J.F. Ping, Y.B. Zhou, Y.Y. Wu, V. Papper, S. Boujday, R.S. Marks, T.W.J. Steele: Biosens. Bioelectron. **64**, 373 (2015)
52.39 S.G. Chatterjee, S. Chatterjee, A.K. Ray, A.K. Chakraborty: Sens. Actuator B-Chem **221**, 1170 (2015)
52.40 Y.Y. Shao, J. Wang, H. Wu, J. Liu, I.A. Aksay, Y.H. Lin: Electroanalysis **22**(10), 1027 (2010)
52.41 M.J. Lu, J. Li, X.Y. Yang, C.A. Zhang, J. Yang, H. Hu, X.B. Wang: Chin. Sci. Bull. **58**(22), 2698 (2013)
52.42 J. Byun: J. Microbiol. Biotechnol. **25**(2), 145 (2015)
52.43 T. Kuila, S. Bose, P. Khanra, A.K. Mishra, N.H. Kim, J.H. Lee: Biosens. Bioelectron. **26**(12), 4637 (2011)
52.44 N.B. McKeown: *Phthalocyanine Materials. Synthesis, Structure and Function* (Cambridge Univ. Press, Cambridge 1998)
52.45 S. Tabuchi, H. Tabata, T. Kawai: Surf. Sci. **571**(1-3), 117 (2004)
52.46 M. Szybowicz, T. Runka, M. Drozdowski, W. Bala, A. Grodzicki, P. Piszczek, A. Bratkowski: J. Molec. Struct. **704**(1-3), 107 (2004)
52.47 A. Boguta, D. Wrobel, A. Bartczak, R. Swietlik, Z. Stachowiak, I.M. Ion: Mater. Sci. Eng. B Solid **113**(1), 99 (2004)
52.48 J. Spadavecchia, G. Ciccarella, S. Capone, R. Rella: Chem. Mater. **16**(11), 2083 (2004)
52.49 H. Nishimura, M. Iizuka, S. Kuniyoshi, M. Nakamura, K. Kudo, K. Tanaka: Electron. Commun. Jpn. **87**(2), 18 (2004)
52.50 Y.L. Lee, H.Y. Wu, C.H. Chang, Y.M. Yang: Thin Solid Films **423**(2), 169 (2003)
52.51 M.C. Petty: *Langmuir-Blodgett Films: An Introduction* (Cambridge Univ. Press, Cambridge 1996)
52.52 M.J. Cook: J. Mater. Sci. Electron. **5**, 117 (1994)
52.53 M.J. Cook: Int. J. Electron. **76**, 727 (1994)
52.54 C.G. Claessens, W.J. Blau, M. Cook, M. Hanack, R.J.M. Nolte, T. Torres, D. Wohrle: Monatsh. Chem. **132**(1), 3 (2001)
52.55 M.J. Cook: J. Mater. Chem. **6**, 677 (1996)
52.56 T.H. Richardson, C.M. Dooling, O. Worsfold, L.T. Jones, K. Kato, K. Shinbo, F. Kaneko, R. Tregonning, M.O. Vysotsky, C.A. Hunter: Colloid Surf. A **198**, 843 (2002)
52.57 M.J. Cook, J. McMurdo, D.A. Miles, R.H. Poynter, J.M. Simmons, S.D. Haslam, R.M. Richardson, K. Welford: J. Mater. Chem. **4**, 1205 (1994)
52.58 A.K. Hassan, A.K. Ray, A.V. Nabok, S. Panigrahi: IEE Proc.-Sci. Meas. Technol. **147**, 137 (2000)
52.59 D. Meyerhofer: J. Appl. Phys. **49**, 3993 (1978)
52.60 P.C. Sukanek: J. Elechtrochem. Soc **138**, 1712 (1991)

52.61 P.C. Sukanek: J. Electrochem. Soc. **144**, 3959 (1997)
52.62 A.K. Hassan, A.V. Nabok, A.K. Ray, A. Lucke, K. Smith, C.J.M. Stirling, F. Davis: Supramol. Sci. Mater. Sci. Eng. C **8/9**, 251 (1999)
52.63 S.M. Critchley, M.R. Willis, M.J. Cook, J. McMurdo, Y. Maruyama: J. Mater. Chem. **2**, 157 (1992)
52.64 G.C. Bryant, M.J. Cook, C. Ruggier, T.G. Ryan, A.J. Thorne, S.D. Haslam, R.M. Richardson: Thin Solid Films **243**, 316 (1994)
52.65 X. Li, H. Xu, Q. Zhou, D. Jiang, L. Zhang, A. Lu: Thin Solid Films **324**, 277 (1998)
52.66 K. Bandyopadhyay, S.G. Liu, H.Y. Liu, L. Echegoyen: Chem. Eur. J. **6**, 4385 (2000)
52.67 C.C. Hsueh, M.T. Lee, M.S. Freund, G.S. Ferguson: Angew. Chem. Int. Ed. **39**, 1228 (2000)
52.68 H.G. Hong, M. Jiang, S.G. Sligar, P.W. Bohn: Langmuir **10**, 153 (1994)
52.69 K.F. Kelly, Y.B.S. Shon, T.R. Lee, N.J. Halas: J. Phys. Chem. B **103**, 8639 (1999)
52.70 V. Poderys, A. Selskis, R. Rotomskis: Solid State Phenom. **97/98**, 221 (2004)
52.71 A. Ulman: *An Introduction to Ultrathin Films: From Langmuir–Blodgett to Self-Assembly* (Academic, SanDiego 1991)
52.72 T.R.E. Simpson, D.A. Russell, I. Chambrier, M.J. Cook, A.B. Horn, S.C. Thorpe: Sensors Actuat. B **29**, 353 (1995)
52.73 M.J. Cook: Pure Appl. Chem. **71**(11), 2145 (1999)
52.74 K. Ozoemena, P. Westbroek, T. Nyokong: J. Porphyr. Phthalocyan. **6**(2), 98 (2002)
52.75 X. Zhang, J.C. Shen: Adv. Mater. **11**(13), 1139 (1999)
52.76 Y. Lvov, G. Decher, H. Möhwald: Langmuir **9**, 481 (1993)
52.77 Y. Lvov, K. Ariga, I. Ichinose, T. Kunitake: J. Am. Chem. Soc. **117**, 6117 (1995)
52.78 Y.M. Lvov, G. Decher: Crystallogr. Rep. **39**, 628 (1994)
52.79 X. Zhang, M.L. Gao, X.X. Kong, Y.P. Sun, J.C. Shen: J. Chem. Soc. Chem. Commun. **9**, 1055 (1994)
52.80 C.Q. Sun, X.Y. Zhang, D. Jiang, Q.A. Gao, H.D. Xu, Y.P. Sun, X. Zhang, J.C. Shen: J. Electroanal. Chem. **411**(1/2), 73 (1996)
52.81 D.J. Revell, I. Chambrier, M.J. Cook, D.A. Russell: J. Mater. Chem. **10**(1), 31 (2000)
52.82 M.J. Cook, I. Chambrier: Phthalocyanine properties. In: *The Porphyrin Handbook*, Vol. 17, ed. by K.M. Kadish, K.M. Smith, R. Guilard (Academic, New York 2003) p. 38
52.83 S. Antohe, N. Tomozeiu, S. Gogonea: Phys. Stat. Sol. (a) **125**, 397 (1991)
52.84 M.J. Cook, M.F. Daniel, K.J. Harrison, N.B. Mckeown, A.J. Thomson: J. Chem. Soc. Chem. Commun. **15**, 1148 (1987)
52.85 M.J. Cook, M.F. Daniel, K.J. Harrison, N.B. Mckeown, A.J. Thomson: J. Chem. Soc. Chem. Commun. **14**, 1086 (1987)
52.86 A.K. Ray, A.V. Nabok, A.K. Hassan, O. Omar, R. Taylor, M.J. Cook: Philos. Mag. B **78**(1), 53 (1998)
52.87 X.B. Huang, Y.Q. Liu, S. Wang, S.Q. Zhou, D.B. Zhu: Chem. Eur. J. **8**(18), 4179 (2002)
52.88 C. Granito, L.M. Goldenberg, M.R. Bryce, A.P. Monkman, L. Troisi, L. Pasimeni, M.C. Petty: Langmuir **12**(2), 472 (1996)
52.89 B.N. Achar, P.K. Jayasree: Can. J. Chem./Rev. Can. Chim. **77**(10), 1690 (1999)
52.90 H.R. Kerp, E.E. van Faassen: Chem. Phys. Lett. **332**(1/2), 5 (2000)
52.91 P.D. Hooper, M.I. Newton, G. McHaleand, M.R. Willis: Semicond. Sci. Technol. **12**, 455 (1997)
52.92 D. Markovitsi, T.H. Tran-Thi, R. Even, J. Simon: Chem. Phys. Lett. **137**, 107 (1987)
52.93 T. Basova, E. Kol'tsov, A.K. Hassan, A. Nabok, A.G. Gurek, A.K. Ray, V. Ahsen: J. Mater. Sci. Mater. Electron. **15**(9), 623 (2004)
52.94 W. Snow, W.R. Barger: Phthalocyanine films in chemical sensors. In: *Phthalocyanines. Properties and Applications*, ed. by C.C. Leznoff, A.B.P. Lever (Wiley, New York 1989)
52.95 M. Nicolau, B. del Rey, T. Torres, C. Mingotaud, P. Delhaes, M.J. Cook, S.C. Thorpe: Synth. Met. **102**(1–3), 1462 (1999)
52.96 R. Zhou, F. Josse, W. Gopel, Z.Z. Öztürk, Ö. Bekaroğlu: Appl. Organomet. Chem. **10**, 557 (1996)
52.97 A. Chyla, A. Lewandowska, J. Soloducho, A. Gorecka-Drzazga, M. Szablewski: IEEE Trans. Dielect. El In **8**(3), 559 (2001)
52.98 S. Gao, H. Zhao, L.H. Huo, J.G. Zhao, Y.Q. Wu, S.Q. Xi: Sensors Actuat. B Chem. **97**(2/3), 319 (2004)
52.99 S. Suslick, N.A. Rakow, M.E. Kosal, J.-H. Chou: J. Porphyr. Phthalocyan. **4**, 407 (2001)
52.100 Y. Lee, B.K. Oh, M.E. Meyerhoff: Anal. Chem. **76**(3), 536 (2004)
52.101 J. Charvatova, O. Rusin, V. Kral, K. Volka, P. Matejka: Sensors Actuat. B **76**(1–3), 366 (2001)
52.102 O. Ikeda, H. Koyama, K. Kijima, T. Komura, A. Itajima, M. Miyake, K. Yamamoto, A. Yamatodani: Detection of Nitric Oxide with the Iron-Porphyrin Doped (III) Nafion|Glassy Carbon Electrode. In: *Proc. 27th Chem. Sensor Symp*, Vol. 14 (Supplement B), (Japan Association of Chemical Sensor, Nagaoka University of Technology 1998) p. 89
52.103 V.C. Smith, T. Richardson, H.L. Anderson: Supramolec. Sci. **4**(3/4), 503 (1997)
52.104 V. Arima, R.I.R. Blyth, F. Della Sala, R. Del Sole, F. Matino, G. Mele, G. Vasapollo, R. Cingolani: Mater. Sci. Eng. C Bio. Solids **24**(4), 569 (2004)
52.105 H. Imahori, K. Hosomizu, Y. Mori, T. Sato, T.K. Ahn, S.K. Kim, D. Kim, Y. Nishimura, I. Yamazaki, H. Ishii, H. Hotta, Y. Matano: J. Phys. Chem. B **108**(16), 5018 (2004)
52.106 Q.L. Li, S. Surthi, G. Mathur, S. Gowda, Q. Zhao, T.A. Sorenson, R.C. Tenent, K. Muthukumaran, J.S. Lindsey, V. Misra: Appl. Phys. Lett. **85**(10), 1829 (2004)
52.107 K.E. Splan, J.T. Hupp: Langmuir **20**(24), 10560 (2004)
52.108 J.R.C. da Rocha, G.J.F. Demets, M. Bertotti, K. Araki, H.E. Toma: J. Electroanal. Chem. **526**(1/2), 69 (2002)
52.109 J.D. Wright: Prog. Surf. Sci. **31**(1/2), 1 (1989)
52.110 J.P. Germain, A. Pauly, C. Maleysson, J.P. Blanc, B. Schöllhorn: Thin Solid Films **333**, 235 (1998)

52.111 B. Schöllhorn, J.P. Germain, A. Pauly, C. Maleysson, J.P. Blanc: Thin Solid Films **326**, 245 (1998)
52.112 J. Travis, A.K. Ray, S.C. Thorpe, M.J. Cook, S.A. James: Meas. Sci. Technol. **6**(7), 988 (1995)
52.113 A. Cole, R.J. McIlroy, S.C. Thorpe, M.J. Cook, J. McMurdo, A.K. Ray: Sensors Actuat. B **13**, 416 (1993)
52.114 D. Crouch, S.C. Thorpe, M.J. Cook, I. Chambrier, A.K. Ray: Sensors Actuat. B **18–19**, 411 (1994)
52.115 A. Tepore, A. Serra, D.P. Arnold, D. Manno, G. Micocci, A. Genga, L. Valli: Langmuir **17**(26), 8139 (2001)
52.116 Y.L. Lee, C.Y. Sheu, R.H. Hsiao: Sensors Actuat. B Chem **99**(2/3), 281 (2004)
52.117 Y.L. Lee, C.H. Hsiao, C.H. Chang, Y.M. Yang: Sens. Actuat. B **94**, 169 (2003)
52.118 M.I. Newton, T.K.H. Starke, G. McHale, M.R. Willis: Thin Solid Films **360**(1/2), 10 (2000)
52.119 M.I. Newton, T.K.H. Starke, M.R. Willis, G. McHale: Sens. Actuat. B **67**, 307 (2000)
52.120 Q. Zhou, R.D. Gould: Thin Solid Films **317**(1/2), 436 (1998)
52.121 W.F. Qiu, W.P. Hu, Y.Q. Liu, S.Q. Zhou, Y. Xu, D.B. Zhu: Sensors Actuat, B Chem. **75**(1/2), 62 (2001)
52.122 M. Bouvet, A. Leroy, J. Simon, F. Tournilhac, G. Guillaud, P. Lessnick, A. Maillard, S. Spirkovitch, M. Debliquy, A. Haan, A. Decroly: Sensors Actuat, B Chem. **72**(1), 86 (2001)
52.123 M. Bouvet, G. Guillaud, A. Leroy, A. Maillard, S. Spirkovitch, F.G. Tournilhac: Sensors Actuat, B Chem. **73**(1), 63 (2001)
52.124 C.Q. Sun, Y.P. Sun, X. Zhang, H.D. Xu, J.C.K. Shen: Anal. Chim. Acta **312**(2), 207 (1995)
52.125 Y.P. Sun, X. Zhang, C.Q. Sun, Z.Q. Wang, J.C. Shen, D.J. Wang, T.J. Li: Chem. Commun. **20**, 2379 (1996)
52.126 K.F. Schoch, J. Greggi, T.A. Temofonte: J. Vac. Sci. Technol. A **6**(1), 155 (1988)
52.127 A.K. Hassan, A.K. Ray, J.R. Travis, Z. Ghassemlooy, M.J. Cook, A. Abass, R.A. Collins: Sensors Actuat. B Chem. **49**(3), 235 (1998)
52.128 J. Mårtensson, H. Arwin, I. Lundstrom: Sensors Actuat. B Chem. **1**(1–6)), 134 (1990)
52.129 J.M. Pedrosa, C.M. Dooling, T.H. Richardson, R.K. Hyde, C.A. Hunter, M.T. Martin, L. Camacho: J. Mater. Chem. **12**(9), 2659 (2002)
52.130 O. Worsfold, C.M. Dooling, T.H. Richardson, M.O. Vysotsky, R. Tregonning, C.A. Hunter, C. Malins: Colloid Surf. A **198**, 859 (2002)
52.131 L. Gaffo, O.D.D. Couto, R. Giro, M.J.S.P. Brasil, D.S. Galvao, F. Cerdeira, O.N. de Oliveira, K. Wohnrath: Solid State Commun. **131**(1), 53 (2004)
52.132 T. Richardson, V.C. Smith, A. Topacli, J. Jiang, C.H. Huang: Supramol. Sci. **4**, 465 (1997)
52.133 T.R.E. Simpson, M.J. Cook, M.C. Petty, S.C. Thorpe, D.A. Russell: Analyst **121**(10), 1501 (1996)
52.134 T.R.E. Simpson, D.J. Revell, M.J. Cook, D.A. Russell: Langmuir **13**(3), 460 (1997)
52.135 E. Kretschmann: Z. Phys. **241**, 313 (1971)
52.136 I. Pockrand: Surf. Sci. **72**, 577 (1978)
52.137 J.M. Rooney, E.A.H. Hall: Anal. Chem. **76**(23), 6861 (2004)
52.138 T. Basova, E. Kol'tsov, A. Hassan, A. Tsargorodskaya, A.K. Ray, I. Igumenov: Phys. Stat. Sol. (b) **242**(4), 822 (2005)
52.139 J.D. Wright, A. Cado, S.J. Peacock, V. Rivalle, A.M. Smith: Sens. Actuat. B **29**, 108 (1995)
52.140 J.P. Lloyd, C. Pearson, M.C. Petty: Thin Solid Films **160**(1/2), 431 (1988)
52.141 M.J. Jory, P.S. Cann, J.R. Sambles: J. Phys. D **27**(1), 169 (1994)
52.142 K. Kato, C.M. Dooling, K. Shinbo, T.H. Richardson, F. Kaneko, R. Tregonning, M.O. Vysotsky, C.A. Hunter: Colloid Surf. A **198**, 811 (2002)
52.143 D.S. Ballantine, R.M. White, S.I. Martin, A.J. Ricco, E.T. Zellers, G.C. Fry, H. Wohltjen: *Acoustic Wave Sensors. Theory, Design, and Physico-Chemical Applications* (Academic Press, New York 1997)
52.144 K.D. Schierbaum, R. Zhou, S. Knecht, R. Dieing, M. Hanack, W. Göpel: Sensors Actuat. B **24**, 69 (1995)
52.145 H. Ding, V. Erokhin, M.K. Ram, S. Paddeu, L. Valkova, C. Nikolini: Thin Solid Films **379**, 279 (2000)
52.146 Z.Z. Öztürk, R. Zhou, U. Weimar, V. Ahsen, O. Bekaroğlu, W. Göpel: Sensors Actuat. B Chem. **26–27**, 208 (1995)
52.147 J. Souto, M.L. Rodriguez, J.A. Desaja, R. Aroca: Int. J. Electron. **76**(5), 763 (1994)
52.148 T. Basova, C. Tasaltin, A.G. Gurek, M.A. Ebeoğlu, Z.Z. Öztürk, V. Ahsen: Sensors Actuat. B Chem. **96**(1/2), 70 (2003)
52.149 C. Bariain, I.R. Matias, C. Fernandez-Valdivielso, F.J. Arregui, M.L. Rodriguez-Mendez, J.A. de Saja: Sensors Actuat. B Chem **93**(1–3), 153 (2003)
52.150 J. Spadavecchia, G. Ciccarella, R. Rella, S. Capone, P. Siciliano: Sensors Actuat. B Chem. **96**(3), 489 (2003)
52.151 J. Spadavecchia, G. Ciccarella, A. Buccolieri, G. Vasapollo, R. Rella: J. Porphyr. Phthalocyan. **7**(8), 572 (2003)
52.152 S. Nardis, D. Monti, C. Di Natale, A. D'Amico, P. Siciliano, A. Forleo, M. Epifani, A. Taurino, R. Rella, R. Paolesse: Sensors Actuat. B Chem. **103**(1/2), 339 (2004)
52.153 J. Spadavecchia, G. Ciccarella, G. Vasapollo, P. Siciliano, R. Rella: Sensors Actuat. B Chem. **100**(1/2), 135 (2004)
52.154 C. Granito, J.N. Wilde, S. Houghton, P.J. Iredale: Thin Solid Films **284/285**, 98 (1996)
52.155 E.K. Quirk, R. Mogg, D.D. Brown, M.A. Lally, D.V. Mehrotra, M.J. DiNubile, M.N. Robertson: Clin. Infect. Dis. **47**(12), 1593 (2008)
52.156 J.D. Huang, S.Q. Wang, P.C. Lo, W.P. Fong, W.H. Ko, D.K.P. Ng: New. J. Chem. **28**, 348 (2004)
52.157 T. Basova, S. Paul, D. Paul, P. Vadgama, A.G. Gürek, V. Ahsen, A.K. Ray: J. Bionanosci. **2**, 114 (2008)
52.158 S. Paul, D. Paul, T. Basova, A.K. Ray: J. Phys. Chem. C **112**(31), 11822 (2008)
52.159 S. Paul, D. Paul, T. Basova, A.K. Ray: IET Nanobiotechnology **4**(1), 1 (2010)
52.160 T. Basova, A.G. Gurek, V. Ahsen, A. Ray: Opt. Mater. **35**(3), 634 (2013)
52.161 G. Jin, C.O. Too, J. Norrish, G.G. Wallace: Synth. Met. **135**(1–3)), 29 (2003)

52.162 V. Papes, S. Brodska: Sensors Actuat. B Chem. **40**, 143 (1997)
52.163 A. Riul, A.M.G. Soto, S.V. Mello, S. Bone, D.M. Taylor, L.H.C. Mattoso: Synth. Met. **132**(2), 109 (2003)
52.164 J.S. Do, W.B. Chang: Sensors Actuat. B Chem. **101**(1/2), 97 (2004)
52.165 G. Anitha, E. Subramanian: Sensors Actuat. B Chem. **92**(1/2), 49 (2003)
52.166 E. Segal, R. Tchoudakov, M. Narkis, A. Siegmann, Y. Wei: Sensors Actuat. B Chem. **104**(1), 140 (2005)
52.167 E. Milella, F. Musio, M.B. Alba: Thin Solid Films **285**, 908 (1996)
52.168 Y. Kunugi, K. Nigorikawa, Y. Harima, K. Yamashita: J. Chem. Soc. Chem. Commun. **1994**(7), 873 (1994)
52.169 S. Christie, E. Scorsone, K. Persaud, F. Kvasnik: Sensors Actuat. B Chem. **90**(1–3)), 163 (2003)
52.170 M. Ando, C. Swart, E. Pringsheim, V.M. Mirsky, O.S. Wolfbeis: Solid State Ionics **152–153**, 819 (2002)
52.171 A. Penza, G. Cassano, A. Sergi, C. Lo Sterzo, M. Russo: Sensors Actuat. B Chem. **81**(1), 88 (2001)
52.172 T.A. Dickinson, J. White, J.S. Kauer, D.R. Walt: Nature **382**, 697 (1996)
52.173 D. Li, C.A. Mills, J.M. Cooper: Sensors Actuat. B Chem. **92**(1/2), 73 (2003)
52.174 T. Jeong, H.K. Lee, D.C. Jeong, S. Jeon: Talanta **65**(2), 543 (2005)
52.175 M.M. Ardakany, A.A. Ensafi, H. Naeimi, A. Dastanpour, A. Shamlli: Sensors Actuat. B Chem. **96**(1/2), 441 (2003)
52.176 S.Y. Huan, C.X. Jiao, Q. Shen, J.H. Jiang, G.M. Zeng, H.H. Guo, S.L. Guo, R.Q. Yu: Electrochim. Acta **49**(25), 4273 (2004)
52.177 P.A. Antunes, C.M. Santana, R.F. Aroca, O.N. Oliveira, C.J.L. Constantinoa, A. Riul: Synth. Met. **148**(1), 21 (2005)
52.178 T. Shimanouchi, S. Morita, H.S. Jung, Y. Sakurai, Y. Suzuki, R. Kuboi: Sensor Mater **16**(5), 255 (2004)
52.179 P.C. Ewbank, R.S. Loewe, L. Zhai, J. Reddinger, G. Sauve, R.D. McCullough: Tetrahedron **60**(49), 11269 (2004)
52.180 J.N. Wilde, J. Nagel, M.C. Petty: Thin Solid Films **327–329**, 726 (1998)
52.181 R. Casalini, J.N. Wilde, J. Nagel, U. Oertel, M.C. Petty: Sensors Actuat. B **57**, 28 (1999)
52.182 J.K. Abraham, B. Philip, A. Witchurch, V.K. Varadan, C.C. Reddy: Smart Mater. Struct. **13**(5), 1045 (2004)
52.183 B. Philip, J.K. Abraham, A. Chandrasekhar, V.K. Varadan: Smart Mater. Struct. **12**(6), 935 (2003)
52.184 R. Capan, A.K. Ray, A.K. Hassan, T. Tanrisever: J. Phys. D Appl. Phys. **36**, 1115 (2003)
52.185 R. Rego, N. Caetanoc, R. Vale, A. Mendes: J. Membr. Sci. **244**(1–2), 35–44 (2004)
52.186 M. Matsuguchi, A. Okamoto, Y. Sakai: Sensors Actuat. B Chem. **94**(1), 46 (2003)
52.187 J. Sutter, A. Radu, S. Peper, E. Bakker, E. Pretsch: Anal. Chim. Acta **523**(1), 53 (2004)
52.188 C. Ramesh, G. Velayutham, N. Murugesan, V. Ganesan, V. Manivannan, G. Periaswami: Ionics **10**(1/2), 50 (2004)
52.189 A. Calogirou, M. Duan, D. Kotzias, M. Lahaniati, B.R. Larsen: Atmos. Environ. **31**, 2741 (1997)
52.190 A.V. Nabok, A.K. Hassan, A.K. Ray, J. Travis, M. Hofton, A. Dalley: IEE Proc. Sci. Meas. Technol. **147**, 153 (2000)
52.191 A.K. Ray, A.V. Nabok, A.K. Hassan, M. Hofton, A. Dalley: Polyphenylsulphide thin films prepared by spin coating for ozone detection. In: *Sensors and their Appl. X Conf*, ed. by N.M. White, J.T. Augousti (1999)
52.192 A.K. Hassan, A.V. Nabok, A.K. Ray, G. Kiousis: Mater. Sci. Eng. C **22**, 197 (2002)
52.193 A.K. Hassan, J. Greenway, A.K. Ray, A.V. Nabok: J. Phys. D Appl. Phys. **36**(17), 2130 (2003)
52.194 B. Ding, J.H. Kim, Y. Miyazaki, S.M. Shiratori: Sensors Actuat. B Chem. **101**(3), 373 (2004)
52.195 L. Uzun, A.P.F. Turner: Biosens. Bioelectron. **76**(SI), 131 (2016)
52.196 J.D. Marty, M. Mauzac: Adv. Polym. Sci. **172**, 1 (2005)
52.197 H. Sambe, K. Hoshina, J. Haginaka: J. Chromatogr. A **1152**(1/2), 130 (2007)
52.198 R.A. Abram, S. Brand: J. Phys. D **48**(12), 125101 (2015)
52.199 Z. Cao, D. Cao, Z.G. Lei, H.G. Lin, R.Q. Yu: Talanta **44**, 1413 (1997)
52.200 C.D. Gutsche: *Calixarenes* (Royal Society of Chemistry, Cambridge 1989)
52.201 D.J. Cram, S. Karbach, H.–E. Kim, C.B. Knober, E.F. Maverick, J.L. Ericson, R.C. Hegelson: J. Am. Chem. Soc. **110**, 2229 (1988)
52.202 F.L. Dickert, U.P.A. Baumler, G.K. Zwissler: Synth. Met. **61**, 47 (1993)
52.203 P. Nelli, E. Delcanale, G. Faglia, G. Sberveglieri, P. Soncini: Sensors Actuat. B Chem. **13/14**, 302 (1993)
52.204 T. Weiss, K.D. Schierbaum, W. Göpel, U. Thoden van Velzen, U. Thoden van Velzen: U. Thoden van Velzen, D. N. Reinhoudt, Sensors Actuat. B Chem **26**, 203 (1995)
52.205 E. Dalcanale, J. Hartman: Sensors Actuat. B Chem. **24**, 39 (1995)
52.206 J. Rickert, T. Weiss, W. Gopel: Sensors Actuat. B Chem **31**, 45 (1996)
52.207 A.K. Hassan, A.K. Ray, A.V. Nabok, F. Davis: Sensors Actuat. B **77**, 638 (2001)
52.208 A.V. Nabok, N.V. Lavrik, Z.I. Kazantseva, B.A. Nesterenko, L.N. Markovskiy, V.I. Kalchenko, A.V. Shivaniuk: Thin Solid Films **259**, 244 (1995)
52.209 A.V. Nabok, A.K. Hassan, A.K. Ray: J. Mater. Chem. **10**, 189 (2000)
52.210 A.V. Nabok, A.K. Hassan, A.K. Ray, O. Omar, V.I. Kalchenko: Sensors Actuat. B Chem. **45**, 115 (1997)
52.211 A.K. Hassan, A.K. Ray, A.V. Nabok, T. Wilkop: Appl. Surf. Sci. **182**, 49 (2001)
52.212 S.J. Gregg, K.S.W. Sing: *Adsorption, Surface Area and Porosity* (Academic Press, New York 1967)
52.213 T. Wilkop, A.K. Ray: J. Phys. D Appl. Phys. **35**(20), 2661 (2002)
52.214 P. Gouma, G. Sberveglieri, R. Dutta, J.W. Gardner, E.L. Hines: MRS Bull. **29**(10), 697 (2004)

52.215 A.V. Shevade, M.A. Ryan, M.L. Homer, A.M. Manfreda, H. Zhou, K.S. Manatt: Sensors Actuat. B Chem **93**(1–3), 84 (2003)
52.216 A. Mills, S.L. Hunte: J. Photochem. Photobiol. A **108**, 1 (1997)
52.217 G.J. Wilson, G.D. Will: Curr. Appl. Phys. **4**(2–4), 351 (2004)
52.218 D. Chatterjee, A. Mahata: J. Photochem. Photobiol. **153**(1–3), 199 (2002)
52.219 L.R. Skubal, N.K. Meshkov, M.C. Vogt: J. Phototechnol. Photobiol. A **148**(1–3), 103 (2002)

53. Packaging Materials

Darrel Frear

This chapter is a high-level overview of the materials used in an electronic package including: metals used as conductors in the package; ceramics and glasses used as dielectrics or insulators; and polymers used as insulators and, in a composite form, as conductors. There is a need for new materials to meet the ever-changing requirements for high-speed digital and radio frequency (RF) applications. There are different requirements for digital and RF packages that translate into the need for unique materials for each application. The interconnecting and dielectric (insulating) requirements are presented for each application and the relevant material properties and characteristics are discussed. The fundamental material characteristics of materials are: dielectric constant, dielectric loss, thermal and electric conductivity, resistivity, moisture absorption, glass transition temperature, strength, time-dependent deformation (creep), and fracture toughness. The material characteristics and properties are dependent upon how they are processed to form the electronic package, so the fundamentals of electronic packaging processes are discussed including wirebonding, solder interconnects, flip chip interconnects, underfill for flip chip and overmolding. The relevant material properties are given along with requirements (including environmentally friendly Pb-free packages) that require new materials to be developed to meet future electronics needs for both digital and RF applications.

53.1 **Package Applications** 1312
53.2 **The Materials Challenge of Electronic Packaging** 1313
53.2.1 Materials Issues in High-Speed Digital Packaging 1313
53.2.2 RF Packaging Materials Issues 1314
53.3 **Materials Coefficient of Thermal Expansion** 1315
53.4 **Wirebond Materials** 1315
53.4.1 Wirebond for Digital Applications 1315
53.4.2 Wirebond for RF 1317
53.5 **Solder Interconnects** 1317
53.5.1 Flip Chip Interconnects 1319
53.5.2 Flip Chip for RF 1320
53.5.3 Pb-Free 1320
53.6 **Substrates** 1322
53.6.1 RF Substrate Materials 1322
53.7 **Underfill and Encapsulants** 1323
53.7.1 Underfill 1323
53.7.2 Encapsulation 1323
53.8 **Electrically Conductive Adhesives (ECAs)** 1324
53.8.1 Adhesive Polymers 1324
53.8.2 Metal Fillers 1325
53.8.3 Conduction Mechanisms 1325
53.8.4 Isotropic Versus Anisotropic Conduction 1326
53.8.5 Rework 1326
53.9 **Thermal Issues** 1326
53.9.1 Thermal Issues in Digital Packaging 1326
53.9.2 Thermal Issues in RF Packaging 1327
53.10 **Summary** 1327
References 1327

An electronic package is a configuration of materials that interconnects electronic signals from one area to another. This scheme must isolate these signals so that there is no interference and must also protect the electronics from a degrading environment. A package typically consists of the semiconductor mounted and interconnected (with solder or Au or Cu wires) to a dielectric substrate (with a leadframe or with metal traces) that is encapsulated to seal the device from the environment. The electronic package must serve four functions:

1. Provide electrical (or photonic) contact to and from the chip
2. Act as an electrical space transformer to take electronic functionality from the dense surface of the chip to the coarser pitched *outside world*

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_53

3. Provide environmental isolation to the semiconductor, and
4. Provide an avenue for conduction of heat away from the device.

In the past, advances in silicon semiconductors had little or no effect on the electronic package design or process. New advances in silicon devices have driven increases in chip speed, the device density (e.g., the number of transistors/mm^2), and power dissipation. These changes, and those planned in the future, require larger numbers of interconnects to and from the devices, improved electrical performance, more effective thermal management, software systems to design and model the assembly, and different sets of materials to make these changes happen. Furthermore, the package is shrinking to a minimalist version of the products that traditionally have been called packaging. The chip and board have become a conjugate packaging system that must provide the traditional levels of electrical, mechanical, thermal interfaces, and environmental protection, with the additional challenge of increased performance. Furthermore, the boundary between where the device ends and where the package begins has become blurred with passive devices (e.g., resistors, capacitors, etc.) in the package and redistribution layers that act as space transformers on the chip. Electronic packaging is becoming one of the greatest challenges in manufacturability, performance, and reliability in advanced electronics applications. With the tremendous growth of wireless telecommunication, RF applications are beginning to drive many areas of microelectronics traditionally led by the development of the microprocessor. An increasingly dominant factor in RF microelectronics is electronic packaging and the materials needed to create the package because the package materials strongly affect the performance of the RF electronics. Many challenges remain for packaging of microprocessors as well. These challenges include increased speed, numbers of input/outputs (I/O), decreased pitch, and decreased cost.

53.1 Package Applications

Electronic packages are divided into four levels, with three that exist beyond the integrated circuit:

- Level 0: Semiconductor chip level (integrated circuits). The materials are: the semiconductor (Si, SiGe, GaAs, etc.), metalization on the circuit (Al, Cu, Au) and interconnects on the die.
- Level 1: Chip in a carrier. The die is bonded (with a conductive or nonconductive adhesive or solder) into a package carrier (substrate or leadframe) and interconnected by either wirebonds or bulk conductive interconnects (flip chip) or solder of conductive adhesives. An example of a wirebonded package is shown in Fig. 53.1. Figure 53.2 is a representation of a flip chip package interconnect. The die is protected either by a lid or encapsulated with a polymer overmold.
- Level 2: The chip carrier mounted to a board. The package is solder, or conductive adhesive, attached to a circuit board. An example of a package mounted to a board using area array interconnects is shown in Fig. 53.2.

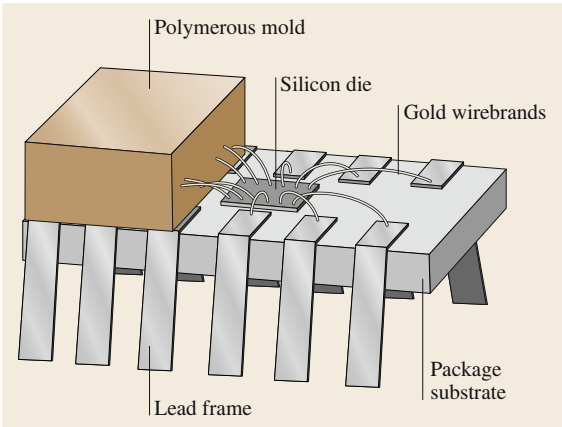


Fig. 53.1 Isometric section schematic that illustrates a peripheral package showing a dual-in-line-package (DiP) that is wirebonded

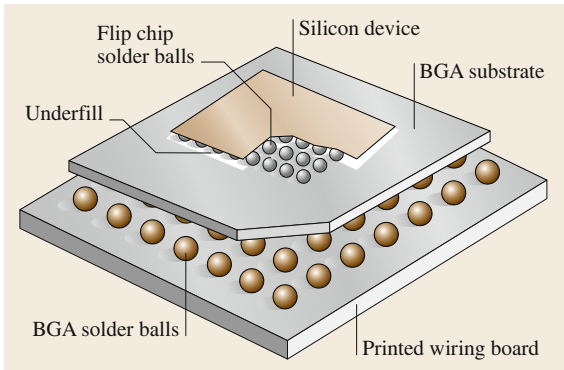


Fig. 53.2 Schematic illustration of an area array package with a flip chip device interconnect and a ball grid array package

- Level 3: Board to board interconnects. The boards are interconnected to the final electronic system using friction interconnects, solder interconnects, or fiber-optic connectors.

53.2 The Materials Challenge of Electronic Packaging

Electronic packaging is arguably the most materials-intensive application today. The families of materials included in a package include: semiconductors, ceramics, glasses, composites, polymers, and metals. A list of the types of materials used in an electronic package is shown in Table 53.1. The processes required to assemble a package are equally varied: welding, soldering, curing, cold and hot working, sintering, adhesive bonding, laser drilling, and etching.

Each of these materials and how they are used in an electronic package could be the topic of an entire book. This chapter is a high-level overview of the materials used in an electronic package. The focus is on the following classes of materials:

Table 53.1 Examples of materials used in electronic packaging

Semiconductors	Si, SiGe, GaAs
Metals	Solders for interconnects (Sn-Pb, Sn-Ag, Sn-Ag-Cu, Sn-Au, Sn-Sb) Au or Cu wirebonds Cu leadframes (Kovar, CuBe, Alloy 42) Cu traces in substrates W, Mo traces in cofired ceramics Ag, Au, Pd for thin/thick films on ceramics Ni diffusion barrier metalizations Al heat sinks
Ceramics	Al_2O_3 substrates modified with BaO, SiO_2, CuO, etc. LTCC substrates Al_2O_3 modified with low-temperature glass (e.g., PbO) SiN dielectrics Diamond heat sinks
Polymers	Epoxies (overmold) Filled epoxies (overmold) Silica-filled anhydride resin (underfills) Conductive adhesives (die bonding, interconnects) Laminated epoxy/glass substrates Polyimide dielectric Benzoyclobutene Silicones Photosensitive polymers for photomasks (acrylates, monomers, etc.)
Glasses	SiO_2 fibers for optoelectronics Silicate glasses for sealing Borosilicate glass substrates Glass fibers for epoxy/glass substrates (FR-4)

Metals are used as conductors in the package, primarily electrical but also thermal for power devices. This includes the thin metal interconnects on the integrated circuit made of Al, Cu or Au and the interconnects between the integrated circuit and the package that are either wirebond interconnects or solder joints. Metals are also used to act as heat sinks for power devices and as shields for RF applications

Ceramics and glasses are used as dielectrics or insulators. Ceramics are used in devices as dielectrics to form capacitors and inductors. In the package itself, the ceramics are used as insulating materials and as substrates that provide a structural base that electrically isolates lines and pads.

Polymers are used as insulators and, in a composite form, as conductors. As an insulator, polymers are encapsulants, underfills and substrates (note: in these applications, polymers are used as composites with silica or glass fillers). Polymers are also used as insulating adhesives to glue components to a substrate or board to provide mechanical strength. The addition of metal particles to the polymer can make it a conductive material used as a conductive adhesive interconnect.

Composite materials are a mix of materials that can be tailored for either mechanical behavior improvements, thermal enhancement or as an electrical conductor. Many of the composite materials used in packages are based on a polymer matrix as noted in the polymer section above.

There is a wealth of general information on materials in electronic packaging. Some additional resources are found in [53.1–8].

53.2.1 Materials Issues in High-Speed Digital Packaging

The ability to decrease line width and feature size in semiconductor technology is decreasing below 90 nm. As the feature size shrinks, the function per unit area on the die increases. This increased functionality means either smaller die or more I/Os per die. The reduced die size also reduces the available perimeter of the package for wirebond pads while increased functionality requires more I/Os. The decreased perimeter area also drives wirebond capability down to 44 µm pitch with 10 µm gold wire. An attractive solution to this decreasing size is to use the entire surface of the chip (flip chip) rather than just the periphery (wirebond). Figure 53.3

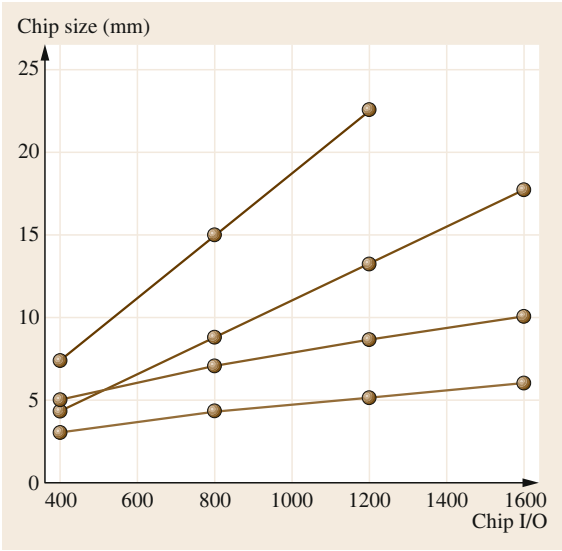


Fig. 53.3 Plot of bond pitch requirements as a function of I/Os and die size for current wirebond (75 μm) and flip chip (250 μm) and state-of-the art wirebond (44 μm) and flip chip (150 μm)

shows a comparison of available I/Os for a given chip size for standard, and state-of-the-art, wirebond and flip chip pitch.

From an overall digital package perspective, the trend is to move from perimeter-bonded die and packages to an array package and then array die bond in an array package that offers increased numbers of I/Os and performance potential. This trend is shown in Fig. 53.4. Materials used in the package and substrate (e.g., ceramic (Al_2O_3), low-temperature cofired ceramic [LTCC] and organic laminants) and package style affect performance as well as area array versus peripheral interconnects. Ceramic packages are the material of choice for hermetic high-performance applications. Ceramic is typically 96% Al_2O_3 with tungsten metalization. The ceramic is fired and sinters at high temperatures that preclude the use of Cu or Al metalization, which would melt in the ceramic sintering process. Multilayer ceramic packages with thin-film metalizations also offer a thermal expansion coefficient similar to the Si die (6×10^{-6} /°C for alumina ceramic versus 3×10^{-6} /°C for Si) so strain that can arise during thermal cycling can be minimized. However, Al_2O_3 ceramic is expensive and has limited use in commercial applications. LTCC is of interest because it offers the hermeticity and thermal expansion advantages of ceramics at a lower cost. The LTCC is an alumina ceramic/glass composite that is fired at sufficiently low temperatures that Cu or Ag can be used as the metallization.

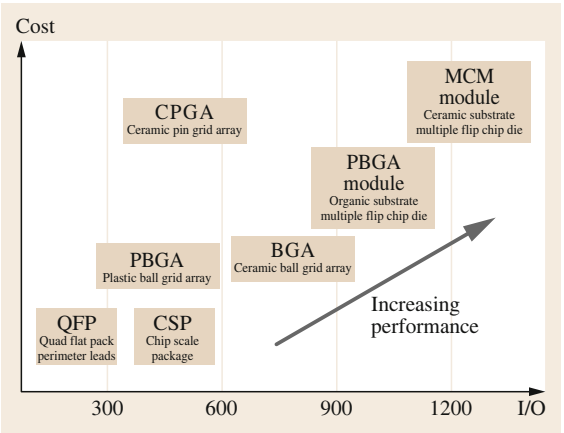


Fig. 53.4 Plot of package style as a function of cost and I/O

Table 53.2 Dielectric constants of packaging substrate materials

Material	Dielectric constant (ε_r)
Epoxy	5.0
FR-4	3.5–4.0
Other laminates	2.5–4.5
Kapton	3.1–3.5
Multilayer ceramic	5.0–9.5
Low-temperature cofired ceramic (LTCC)	3.9–7.8

For the interconnects in the package, flip chip appears to provide substantial improvements in I/O and pitch but wirebond will remain as a packaging solution because there is a great deal of capital invested in wirebond equipment that cannot be ignored.

In addition to smaller size driving finer pitch, increased signal speed will drive package requirements. Digital signal delays must be minimized. The total delay is a function of the total distance and the delay per unit length that is a function of the transmitting medium's dielectric constant and is the square root of the dielectric constant (ε_r) times the free space delay in vacuum (33 ps/cm). Therefore, a material with a high dielectric constant increases the delay. Table 53.2 shows a variety of dielectric constants used in packaging.

For minimal signal delays, an optimal dielectric substrate material is required. The signal length can also be shortened by changing from a wirebond solution to flip chip.

53.2.2 RF Packaging Materials Issues

High-speed and microwave circuits are defined for digital devices with clock speeds faster than 100 MHz

and 0.1 to 100 GHz for analog circuits. For analog, this is also termed radio frequency (RF) circuits. The microwave, or RF, circuit module is defined as microstrip elements composed of transmission lines and matching networks on a substrate with discrete components (resistors, inductors, capacitors and transistors) attached, or embedded, in the substrate. A monolithic microwave device has all the above elements integrated onto a semiconductor die. The semiconductor used for RF applications is typically GaAs because of its high resistivity and suitability for circuits that operate at high frequencies. The following is a discussion of critical packaging materials issues related to RF devices and modules.

One of the most significant differences between RF and digital packaging is that the package is part of the RF circuit due to the interaction of the RF electric field with all adjacent conductors and insulators. This is one of the most significant challenges in the design of RF circuits. This will be further exacerbated as the trend is to move from a single die RF package into multiple die packages (or RF modules) with passive components. For example, the wireless communication industry is striving toward a solution of a *phone-in-a-package* where the entire electronic functionality of a cellular phone can fit into a package with a size on the order of 1 cm^2. Materials used in the package of these RF applications are significant to the performance and cost of the solution.

It is clear that the electronic package is a complex materials system that is driven by a variety of thermal, mechanical and electrical performance requirements. These requirements are also application-dependent for both digital and RF packages. The remainder of this chapter is an overview of the key materials required for microelectronic packaging.

53.3 Materials Coefficient of Thermal Expansion

The coefficient of thermal expansion (CTE) is the length of increase of a macroscopic sample for a given temperature increase. The units of the CTE are length/length °C. The CTE is a critical physical property of materials used in an area array package. A wide variety of materials with a variety of CTEs, such as metals, ceramics, and polymers are joined together in an electronic package assembly. As processing (or in-use) temperatures change, the materials expand or contract to various degrees that could result in the formation of extensive, and nonuniform, strains in the package assembly. A set of materials typically used in electronic packages and their CTEs are shown in Table 53.3.

53.4 Wirebond Materials

Wirebonding is the process where a thin wire (can be less than 25 μm in diameter) made of Au or Al is bonded to the surface of an integrated circuit then to a pad or leadframe in the package (Fig. 53.1). Gold ball thermosonic bonding is the typical method used to form these interconnects. The thermocompression bond is a weld between two metals where thermal energy (preheating of the capillary tool and wire), force and ultrasonic energy are imparted to the wire causing it to melt on the surface and rapidly interdiffuse with the bond pad to form a joint. The integrated circuit pad surface finish is typically Al or Au and the package bond pad finish is typically Au or Cu.

One of the biggest transitions in materials for electronic packaging occurred when the price of Au dramatically increased in the early 2000s, which forced a change to dramatically increased use of Cu as the wirebond material. By the end of 2010, the installed base of copper wire capable bonders approached 25% of the total installed fine-pitch wirebonder base from 5% in 2009 [53.9], and by 2015 more than 33% of all wirebond machines bond Cu [53.10].

53.4.1 Wirebond for Digital Applications

The challenges for decreased wirebond pitch are primarily related to the process of accurate machine control to move bonding heads across shorter distances (44 μm pitch) with increased accuracy. However, there are materials challenges. The metallurgical challenges of wirebonding Au to Al have been well addressed. The classic, early failures were dominated by *purple plague* and the definitive work in this area is by *Philofsky* [53.11]. Purple plague resulted when the Au of the wirebond extensively reacted with Al to form brittle $AuAl_2$ that resulted in bond failures. (Purple plague derives its name from the purple tinge of the $AuAl_2$ intermetallic.) This was addressed by reducing the

Table 53.3 Coefficients of thermal expansion of a sampling of materials used in area array electronic packages

Material	CTE (10^{-6} /°C)
Metals	
304 Stainless Steel	17.8
Ag	19.7
Al	23.5
Alloy 42	4.9
Au	14.2
Cu	16.8
Invar	1.6
Kovar	5.5
Mo	5.1
Ni	13–15
Solder: 63Sn-37Pb	25
Solder: 95Pb-5Sn	28
Ti	10
W	4.5
Ceramics	
AlN	4.3
Alumina (96%)	6.4
Alumina (99.5%)	6.5
BeO	7.8
BN	3.7
Fused silica glass	0.56
Quartz	13
SiC	3.8
SiN	3
Semiconductors	
GaAs	5.8
Si	2.7
Organic materials	
Epoxy resins	50–80
FR-4 (x–y plane)	15.8
FR-4 (z-axis)	80–90
Polycarbonates	50–70
Polyimide glass (x–y plane)	12–14
Polyimide glass (z-axis)	60
Polyimides	40–50
Polyurethanes	180–250
RTV	800

wirebond process temperature to below 300 °C, which reduced the extent of the $AuAl_2$ formation. Current wirebond failures are the result of surface impurities or corrosion. These are addressed by plasma or ultraviolet (UV)-ozone cleaning of the surface prior to wirebond.

Wirebonding problems have also been caused by plating impurities in Au bond pads. The impurities in the Au diffuse to the molten front that occurs during wirebonding and concentrate in a plane in the bond. If there are a sufficient number of impurities, they precipitate and act as sinks for vacancies that can become voids and lower the bond strength. The impurities have been identified as Ni, Fe, Co and B [53.12]. The contaminants are present in the Au film as a result of the plating process. This can be addressed by optimizing the plating process to form pure metallic Au.

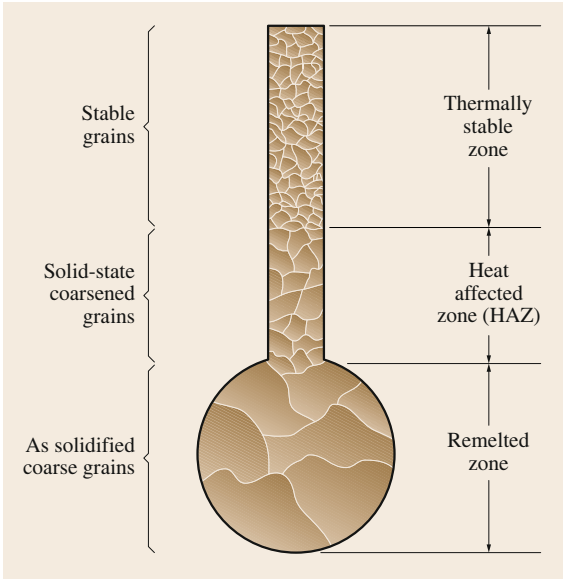


Fig. 53.5 Schematic illustration of the grain size in a Au wirebond. (After [53.13])

With finer pitch wirebond the Au wire must also decrease in size, approaching 10 μm in diameter. This thin wire poses a materials challenge. The wire must have sufficient ductility to deform to the required wirebond shape but retain sufficient strength to not deform or move during the molding process (wire sweep). The Au alloy must be optimized for these opposing requirements. Au wire is typically stabilized with Cu and Ag dopants at levels less than 100 ppm by weight. These dopants keep the grain size small to retain strength without deleteriously affecting ductility. As the Au wire decreases in width, the levels of dopants may need to be increased reducing the Au to less than the current 99.9% level. Proposed dopants include Be and Cu. The critical region in the wirebond is above the ball and below the wire called the heat affected zone (HAZ). A schematic of the zone is shown in Fig. 53.5. This zone forms above the ball that was molten and cools to form a very large as-solidified grain size. The HAZ forms a large grain size while coarsening in the solid-state while the molten ball is created. (The ball is formed under a flame or high current, solidifies, and is then placed on a Au or Al pad and with the addition of energy (pressure/heat/vibration) forms a solid-state ball bond.) The strength in the HAZ is reduced by the Hall–Petch relationship that states strength is inversely proportional to grain size. The HAZ would be susceptible to poten-

tially high levels of strain during molding and would be the most likely failure location. In thinner wires, an increased grain size is even more critical and alloying elements are needed to stabilize the structure and retain sufficient strength without affecting the ductility required to form the wirebond.

The use of Cu as the wirebond material requires significant process changes because Cu is harder than Au and oxidizes much more quickly, requiring specialized copper wirebond process parameters (e.g., a reducing atmosphere at wirebond or higher pressure/energy at wirebond). Cu is harder and has lower ductility than Au so the bonding process window is more narrow and requires more effort to optimize the process so damage to the Si under the bond pad is minimized. The metallurgical reaction between the Cu wire and the Al bond pad is not as extensive as the Au-Sn intermetallic that forms in Au wirebonding. The thinner intermetallic is more stable and results in a more robust bond. The electrical performance of Cu is better than Au wirebonds because copper enables better heat dissipation and can carry more power with thinner wire diameters.

53.4.2 Wirebond for RF

At RF frequencies the wirebonds affect electrical performance. At sub-RF frequencies, the wirebonds can be neglected as their resistance and inductance are in the noise with respect to performance. In RF applications, the wirebonds experience parasitic losses due to inductance with adjacent wirebonds resulting in signal disturbance. RF circuit designers typically add at least 1 nH inductance per wirebond, but the effect can be variable and must be physically compensated by tuning the circuit. In effect, the wirebonds act as discrete inductors. The materials issues for RF wirebond are the same as described above for digital, but the effect of the wirebonds themselves is significant in RF compared to digital and must be accounted for in the design of the device.

53.5 Solder Interconnects

For surface mount and area array applications, the role of solder in the package is significant. In these advanced designs, the solder is an electrical interconnect, a mechanical bond and must often serve as a thermal conduit to remove heat from the joined device. The interconnects become more critical as chip size, chip carrier size, and the number of I/Os increase, while the solder joint size and cost decreases. Furthermore, some of the use environments are becoming increasingly severe. For example, some automotive electronics are being placed in locations under the hood with temperature extremes that could range from −55 to 180 °C. Portable devices, such as cellular telephones are expected to withstand severe shock environments caused by dropping the phone.

A solder interconnect consists of the solder alloy and the pads to which it is joined. The solder alloy is a low-temperature-melting metal that wets and reacts with metal pads or leads to form a metallurgical bond. During the wetting process the solder reacts with the pad metalization forming an intermetallic compound (IC) on the pad surface that continues to grow when the solder is in the solid state.

The most commonly used solder alloy for electronic interconnects is near-eutectic Sn-Pb with a composition of 63Sn-37Pb (by wt%). The microstructure of this alloy is shown in the optical micrograph in Fig. 53.6 and consists of Sn- and Pb-rich lamella. The lamella are oriented as cells after solidification with the slightly coarsened regions that separate the individual cells as the last part of the solder to solidify. This alloy melts at 183 °C. Pb-rich Sn-Pb alloys (such as 95Pb-5Sn or 97.5Pb-2.5Sn) are used for soldering temperature hierarchy as these alloys melt at temperatures in excess of 300 °C. The microstructure of Pb-rich alloys consists of a matrix of large Pb grains decorated with Sn precipitates. In flip chip applications die are soldered to substrates with high-Pb solder. The substrates are soldered to a board with near eutectic Sn-Pb. The Sn is the active component of these solder alloys and reacts

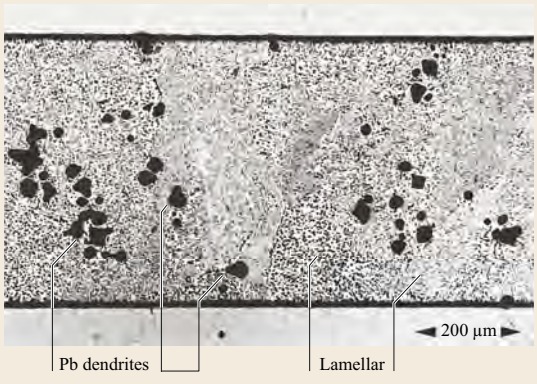


Fig. 53.6 Optical micrograph of the near-eutectic Sn–37Pb solder microstructure with the lamellar Sn- and Pb-rich phases and dendrites indicated

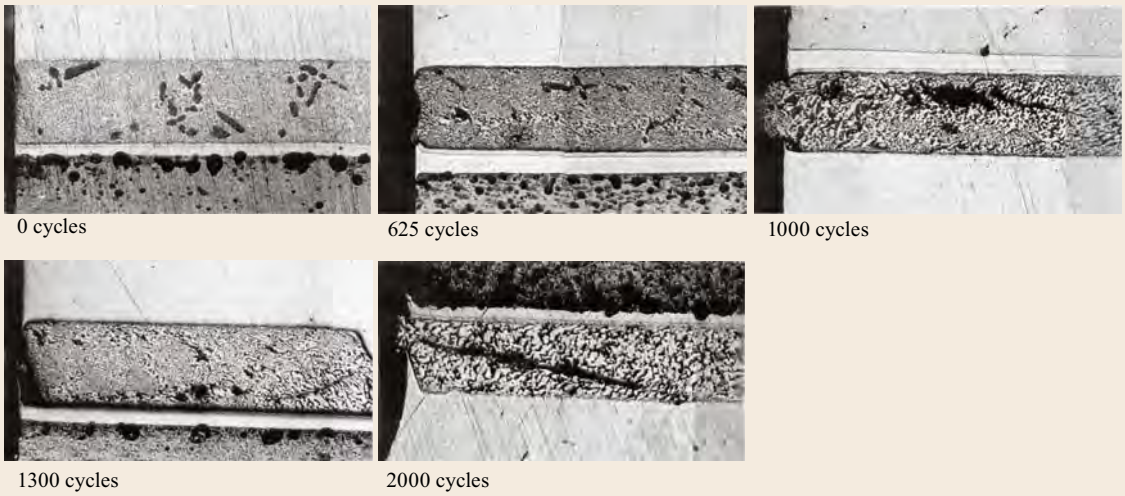


Fig. 53.7 Optical micrographs that show heterogeneous coarsening of the eutectic Sn-Pb microstructure at colony boundaries after thermomechanical fatigue

with metalizations such as Cu or Ni to form intermetallic compounds.

Other solder alloys have found some use in electronic packaging, including Sn-Ag, Pb-In, Sn-Sb, etc. These alloys are used for a variety of applications including creating a soldering temperature hierarchy (higher or lower than Sn-Pb) or slowing the reaction rate with the metalization (e.g., In), or changing the mechanical properties of the joint. These alloys either form as a single-phase matrix with precipitates or as two-phase structures similar to eutectic Sn-Pb. The focus of this section, however, is on the most commonly used solder alloy, eutectic Sn-Pb.

The microstructure of the solder influences the mechanical behavior and reliability of the interconnects by governing the deformation and failure process. Typical deformation depends on the motion of dislocations and the growth and reconfiguration of grains and this is more sensitive to the structure of the solder than its chemistry. Therefore, depending upon processing conditions, the same solder alloy can respond differently to imposed stress and strain. Additionally, solders are low-melting-temperature materials that are used at a significant fraction of their melting point and are thermodynamically unstable. As the joint ages (at temperature and under strain) the microstructure changes and the mechanical properties, and reliability, change. The microstructure changes by grain (or phase) coarsening that may be heterogeneous or homogeneous. Furthermore, the interfacial intermetallic grows in thickness and coarsens with time at temperature. All of this must be comprehended to determine the reliability of solder interconnects.

Thermomechanical fatigue occurs when materials with different CTEs are joined and used in an environment that experiences cyclic temperature fluctuations resulting in imposed cycling strain. Thermomechanical fatigue is a major deformation mechanism concerning solder interconnects in electronic packages. Even small temperature fluctuations can have a large effect, depending upon the joint thickness and CTE difference of the joined materials. The strain imposed on the solder joints follows the relation

$$\Delta\gamma = \frac{\Delta\alpha\Delta Ta}{h} ,$$

where $\Delta\gamma$ is the shear strain imposed, $\Delta\alpha$ is the difference in coefficient of expansion between the joined materials, ΔT is the temperature change, a is the distance from the neutral expansion point of the joined materials, and h is the thickness of the interconnect.

After a critical number of thermal excursions, such as machine on/off cycles, solder joints experience fatigue failure. The type and magnitude of strains in solder joints under conditions of thermomechanical fatigue are often quite complex. For surface mount applications, the strain is nominally in shear. However, tensile and mixed-mode strains can occur due to bending of the chip carrier or board.

The combination of strain and temperature during thermomechanical fatigue has a large effect on the microstructure and microstructural evolution of eutectic Sn-Pb solder joints. The microstructural evolution of 60Sn-40Pb solder as a function of the number of thermal cycles (−55−125 °C) is shown in Fig. 53.7. The

microstructure evolves through deformation that concentrates at the colony boundaries closely parallel to the direction of imposed shear strain, causing the cells to slide or rotate relative to one another. The structure within the cell boundaries becomes slightly coarsened relative to the remaining solder-joint microstructure and, thus are the *weak links* of the joint. Damage (in the form of defects or dislocations) is created at the cell boundaries at the low-temperature portion of a thermal cycle. As the temperature rises, the deformation is annealed by recrystallization or stress-assisted diffusion where material diffuses to regions of high stress. This results in coarsening the Sn-rich and Pb-rich grains and phases in colony boundaries. The heterogeneously coarsened colony boundaries are weaker than the rest of the joint and any further deformation concentrates in the coarsened regions resulting in further coarsening. Failure eventually occurs due to cracks that form in the coarsened regions of a joint. The first indications of impending failure are associated with cracking of coarsened Sn-rich grains in the heterogeneous region whose initial as-solidified grain size is in the submicron range. When cracks initiate during thermomechanical fatigue, the Sn-grain size grows to a range of about $5-10\,\mu m$. Failure occurs when grains can no longer slide and rotate to accommodate the imposed strain resulting in intergranular separation.

Lead-rich Pb-Sn alloys undergo thermomechanical fatigue but show little evidence of microstructural evolution. These solders undergo intergranular failure caused by void coalescence and growth at Pb grain boundaries. The strain imposed during thermomechanical fatigue cannot be accommodated by large Pb grains, resulting in intergranular failure. However, Pb loses the work-hardening effect quickly so that damage in the form of cracks does not propagate easily.

Other solder alloys, such as the Pb-free Sn-3.5Ag eutectic-based solders, experience thermomechanical fatigue damage and failure at Sn grain boundaries. The microstructural evolution in these alloys tends to be phase coarsening with minimal grain size coarsening. Sn-Ag-X alloys tend to have longer thermomechanical fatigue lifetimes than near-eutectic Sn-Pb solders.

Intermetallic compounds form between pad metallization and the active components of the molten solder (typically Sn). For a Cu metalization, the Sn reacts to form Cu_3Sn and Cu_6Sn_5 intermetallics. For Ni, the Sn reacts to form Ni_3Sn_4. After solidification, the intermetallic compounds continue to grow by solid-state diffusion. Over long periods of time, the intermetallic layers can grow to significant thicknesses ($> 20\,\mu m$) and the solder–intermetallic interfaces may constitute easy sites for crack initiation and propagation. Excessive growth also consumes the base metal, or finish, that can result in the loss of adhesion to the underlying metal that is not solder wettable or create a plane of weakness owing to the stress generated from an intermetallic layer that is too thick. The metalized pad thickness generally must be greater than that consumed by the solder.

The transformation of solder-wettable coatings into intermetallics by solid-state reactions can also result in excessive intermetallic growth that degrades mechanical properties. The interfacial intermetallics are brittle and may fracture when strain is imposed, especially if the strain is tensile in nature. Solder-joint interfacial intermetallics are brittle because they typically have complex crystal structures with few crystallographic planes available to accommodate stress by strain relief, i. e., plastic deformation via a slip mechanism. The failures are characteristically brittle and occur through the intermetallic or at the intermetallic/solder interface under low-load conditions.

53.5.1 Flip Chip Interconnects

Recent developments in the under bump metallurgy (UBM) and solder joints have resulted in lower cost and higher performance flip chip interconnects. The UBM serves as an electrical, thermal, and mechanical interface between the silicon bond pads and the package substrate. The structure of the UBM is designed to adhere to the Al on the Si, act as a diffusion barrier between the solder and Si, and be a wettable surface to join to the solder. The most common UBM is the evaporated Cr/Cr-Cu/Cu/Au developed by IBM [53.14] for use with an evaporated high-Pb-content Pb-Sn alloy joined to a ceramic substrate. This UBM is expensive and new UBM systems with sputtered or plated metalizations have been developed. Plated metalizations are the lowest cost UBM.

A significant change in flip chip interconnects is the move toward a lower-melting-point solder alloy, such as eutectic 63Sn-37Pb (wt%) that can be processed below 220 °C as opposed to 350 °C for high-Pb solder. New materials on the die (low-κ dielectric) and substrate (organic materials) require this lower processing temperature. The change in solder alloy also requires a change in UBM. Traditional UBM systems are based on a thin Cu layer that dissolves into high-Sn solders resulting in spalling and dewetting. Eutectic solder UBM systems utilize a layer of Ni for solder wetting. The Ni is wet by Sn-Pb solder but reacts much slower than Cu. The addition of $\approx 9\%$ V to Ni reduces the ferromagnetic behavior of the Ni and enables the use of sputtering of the Ni. Electroless plating of Ni-9%P is the lowest cost UBM and has the advantage that plating occurs only where desired on the metal pads on the silicon. To plate onto Al, the pads are zincated then

immediately plated with Ni. One issue with Ni-based UBMs is the intermetallic layer between the solder and Ni (Ni_3Sn_4). Although very thin, this intermetallic layer can be brittle and the long-term reliability must be fully characterized and understood.

There are three solder deposition techniques available for flip chip. Traditionally, high-Pb solder is evaporated onto the UBM then reflowed to form metallurgical bonds. Evaporation is expensive because the masks must be regularly cleaned and there is excessive scrap solder. The composition of evaporated solder is difficult to control outside the high-Pb regime. Solder paste is a lower cost alternative to evaporation and involves screening the paste onto the UBM then reflowing. Solder paste reduces waste, cleaning is inexpensive, and the paste comes in a variety of compositions, including eutectic Sn-Pb. The difficulty with paste arises at fine pitch applications where the rheology of the paste makes it difficult to force into the required small holes in the screen. Solder may also be plated, which is inexpensive (no waste and easy cleaning) and has excellent fine pitch coverage. Solder bumps of a pitch of 25 mm have been achieved by plating. The difficulty with plating is achieving ball and compositional uniformity.

The trend in flip chip interconnect pitch is that the current 250 μm pitch will continue to shrink. The driving force for this pitch shrink is to satisfy the requirements for high-performance silicon devices. These requirements include a dramatic increase in the number of I/Os due to increases in the number of signal lines and power requirements. Higher power devices require more signal and ground lines and, to limit point sources of heat, the power and ground interconnects should be spread evenly across the area array.

There are a number of materials and processing challenges associated with finer pitches. As the pitch shrinks, the methods to deposit the solder become more limited. Solder paste is very difficult to deposit using a silk-screen method at pitches below 150 μm due to rheological limitations of forcing a semisolid (the paste) into small holes (the silk screen). Evaporation is difficult because developing a metal screen mask with the required tolerances is prohibitively expensive. Solder plating is still a good option but the solder must be very uniform across each die. Solder ball uniformity is critical because large variations between die could result in electrical opens, for small balls, and shorts, for large balls. At 250 μm, ball uniformity across a die is 10%, at 150 μm this decreases to 5% and the change in dimension tolerance is 12.5 μm down to 3.75 μm. An additional issue with a decrease in ball size is the joint gap between the substrate and the die decreases to the point that it may become very difficult to flow underfill completely under the die. At 100 μm pitch the gap between the die and substrate could be significantly less than 25 μm, below the limit of underfill flow. For these very fine pitch applications an alternative underfill technique will need to be developed because flow under the chip will not be possible. One alternative would be to deposit the underfill material on the wafer immediately after flip chip solder bumping then underfill cure would simultaneously occur during solder relfow. New underfill materials and processes must be developed to implement this process.

53.5.2 Flip Chip for RF

Flip chip provides substantial improvements in off-circuit RF performance because the inductance is decreased as the ball height and shape is very consistent and predictable. Wirebond lengths can vary (typically $+/-100\,\mu m$) resulting in parasitic variations not observed in flip chip interconnects. The number of I/Os on an RF circuit is generally small so the pitch of the interconnects is generally quite large.

The die interconnects on GaAs are typically Au whereas on Si they are Al. A UBM is required for Si applications because solder does not wet Al. The solder would wet the Au of the GaAs but the Au layer is very thin and would dissolve the entire Au layer very quickly so a UBM is also required for GaAs. The UBMs available for Si can also be applied to GaAs but can be simpler because the adhesion of other metals to Au is easier to achieve than it is on Al.

The solder alloys for GaAs flip chip bumps tend to be similar to that for Si but with a greater emphasis on Au-based alloys because of the desire to use a highly conductive material such as Au for very-high-speed applications. Eutectic Au-Sn, Ag and In alloys have been commonly used but the wetting of these solders is typically poor. For finer pitch applications on GaAs (due to small die size, not large numbers of I/Os) other materials with good wetting are needed. Conductive adhesives have been explored for GaAs bumps. Work by *Lin* et al. [53.15] found no significant difference in electrical performance up to 2 GHz between a conductive adhesive and Au-Sn solder. The use of underfill for flip chip packages can cause a major loss in RF energy because the high dielectric constant of the underfill polymer is greater than that of air. With a conductive adhesive localized to just the bond pads, not the entire surface of the die, this RF loss will not be observed.

53.5.3 Pb-Free

The electronics industry extensively uses Pb-Sn solder alloys in flip chip applications as well as in many

other interconnects in the electronic package. However, medical studies have shown that Pb is a heavy metal toxin that can damage the kidneys, liver, blood, and central nervous system. Less than one percent per year of the global Pb consumption is used in solder alloys for electronic products but electronics and electrical systems make up an increasingly larger fraction of landfills [53.16]. The issue of Pb leaching from landfills into the water table has raised alarm as a potential source of long-term contamination of soil and ground water. Concerns about the presence of Pb in the environment and potential exposure scenarios that could result in the ingestion of Pb by humans and wildlife has prompted a concerted effort to limit the use of Pb in manufactured products (notably gasoline, plumbing solders, and paint). International laws have recently been proposed to expand Pb control laws to limit or ban the use of Pb in manufactured electronics products. The most aggressive and well-known effort is the European Union's Waste in Electrical and Electronic Equipment (WEEE) directive that proposed a ban on Pb in electronics by 2006. The Japanese Environmental Agency has proposed that Pb-containing scrap must be disposed of in sealed landfills to prevent Pb leaching. Electronics manufacturers have responded to these proposed bans in a variety of ways. Many companies have not taken a stance hoping that legislation will not be enacted. Other companies have aggressively pursued solutions to the proposed bans and are using Pb-free products as a *green* marketing strategy. Extensive research on Pb-free solders has been published. A comprehensive review of the status of Pb-free solders can be found in the literature that is primarily focused on carrier-to-board (surface mount and through-hole) interconnects [53.17–21]. A growing requirement is Pb-free solders for flip chip interconnects.

One benefit of a Pb-free flip chip interconnect is the reduction of Pb^{210}-created alpha particle radiation. All mined Pb contains a small amount of radioactive Pb^{210} that decays and emits alpha particles. When an alpha particle enters an active element of the Si (such as a memory cell) it has sufficient energy to cause the stored charge to be released with the result of changing stored memory from a 1 to a 0 state. There is no permanent damage to the Si itself so this radiation induced fault is termed a *soft error*. The alpha particles have a low energy that is dissipated over relatively short distances. However, due to their proximity to active elements, the flip chip solder interconnects have sufficient levels of alpha particle radiation to induce soft errors in CMOS technology that becomes more critical as cell size on the die is reduced [53.22, 23]. For Pb-Sn solders, one solution to alpha particle radiation is to use elemental Pb that was mined many, many years ago where the majority of Pb^{210} has decayed. The source of this Pb is typically found as the ballast of shipping vessels that sank almost 2000 years ago and is relatively expensive. The elemental constituents of Pb-free solders (Sn, Cu, Ag, Bi, In, Sb), however, do not radioactively decompose so alpha particle radiation is minimal.

Flip chip interconnects are the electrical and mechanical connections between the semiconductor integrated circuit and the package (or board for direct chip attach DCA). These interconnects are formed on the periphery or in an area array on the top surface of an active die. Flip chip interconnects are formed by depositing solder onto a metalized Si wafer in the form of discrete balls, solder paste or by directly plating onto the pads on the wafer. The solder must wet and join to the pads on the Si devices so an under bump metallurgy (UBM) is typically deposited on the Al or Cu pads on the Si. The UBM typically consists of a barrier metal (e.g., Ti or W) followed by a solder wettable layer (e.g., Cu or Ni). The top layer of the Ni metalization is covered with a noble metal, such as Au, to prevent oxidation that inhibits solder wetting. The UBM also acts as a diffusion barrier between the Si and the solder and must be thick enough to withstand interactions (intermetallic formation) between the solder and UBM. Flip chip interconnects are smaller (on the order of 100 μm diameter) than surface mount joints and are projected to have pitches that shrink below 150 μm. Flip chip interconnects have a unique set of requirements. These joints must be able to withstand a potentially high level of strain mismatch between Si and an organic substrate. Flip chip technology has moved from ceramic packaging with high-Pb solder (97.5Pb-2.5Sn) to an organic package that requires lower-temperature reflow ($< 260\,^{\circ}C$). This can be accomplished by bumping the die with high-Pb solder then joining it to an organic board with eutectic Sn-Pb but this is a cost adder that is eliminated with a monolithic solder. The joints must withstand board-level reflow environments compatible with joining to organic substrates that, again, have a maximum reflow temperature of $260\,^{\circ}C$. The Pb-free solder must meet these requirements and perform at, or above, the level of performance of the Sn-Pb solder it is intended to replace. The flip chip solder alloy is typically deposited on Si wafers either as solder paste stencil printed on the defined UBM pads or direct plating on the UBMs. The simplicity and low cost of plating the solder for flip chip interconnects makes electrochemical deposition the most attractive choice for flip chip bumping.

53.6 Substrates

The move to an organic substrate is the focus of an extensive development effort worldwide by both printed circuit board manufacturers and traditional suppliers that requires significant amounts of capital. Cost savings are expected by replacing ceramic substrate with organic substrates. One stated reason for moving to an organic substrate is based on the belief that, in volume production, organic materials are less expensive than ceramics due to the nature of printed circuit board processing. The organic substrate also offers an increase in electrical performance due to a lower dielectric constant (therefore, less capacitance and better speed) of 3.5 for organic versus 5.0 for ceramics.

Advanced substrates for flip chip must also provide increased wireability while delivering improved electrical performance with reliability levels at least equivalent to current surface mount applications. The board interconnect density for flip chip substrates must accommodate the increasing density of off-chip interconnects, and cost-effective substrate capability that combines the necessary fine line and microvia features must be developed. Microvias are the metal-filled holes that provide a conduction path between copper lines in the multiple layers of an organic substrate. Microvia capability must also scale with line width/spacing in order to provide the *via in line* structures that will be necessary to support bump pitch densification. At 250 µm pitch, microvias can be created using a photolithographic process. At a finer pitch, laser drilling techniques are used to create the required small via size to support the smaller capture pads and facilitate the dense routing required for flip chip.

Materials with dielectric constants approaching 2.0 with coefficients of thermal expansion approaching 6.0 ppm/°C are necessary to meet fine pitch requirements. Furthermore, uniformity and flatness requirements must be improved on organic substrates to ensure uniform joint size and bonding across each die as the interconnect pitch decreases.

The change in substrate materials drives the need for a lower-melting-temperature solder to replace the high-Pb-content Sn-Pb alloys that are processed in excess of 300 °C for ceramic substrates. The organic materials in a substrate are typically epoxy-based and char or burn at temperatures in excess of 250 °C. The most common solder alloy candidate is near-eutectic Sn-Pb that can be processed well below 250 °C ($T_{mp} = 183$ °C). However, this could result in a situation where the flip chip interconnect melts every time a subsequent solder reflow occurs. Multiple reflows of the solder interconnects accelerates intermetallic growth between the tin of the solder and the metalized pads for the UBM and substrate. One way to circumvent this issue is to deposit low-melting-temperature eutectic Sn-Pb solder on the substrate lands, or on the high-melting-temperature solder bump. Reflow to the substrate can be performed at eutectic Sn-Pb soldering temperatures but the mixing between the high-Pb content and eutectic solder results in a higher melting temperature composition. Subsequent reflows, at eutectic Sn-Pb processing temperatures, will not remelt the *composite alloy* flip chip joint. This addresses the low-temperature processing requirements for organic substrates but does not address the low-κ dielectric low-temperature processing requirements because the high-Pb-content solder must still be reflowed on the silicon.

53.6.1 RF Substrate Materials

Optimum electrical performance is achieved in an RF module when the signal-carrying conductor is adjacent to the ground plane to contain the electric field. In microstrip designs the conductor is above the ground plane and in coplanar designs the conductor is sandwiched between ground plans as shown schematically in Fig. 53.8. These designs affect the materials in the package. The impedance of the lines is desired to be 50 Ω (which matches the radiation impedance of the antenna to air). The impedance is affected by the conductor material, typically Au, and the dielectric material (Si, SiGe, GaAs, for the die and LTCC, Al_2O_3 or epoxy (organic) for the substrate).

Like its digital counterpart, RF packaging is moving away from the Al_2O_3 ceramic substrates used in RF

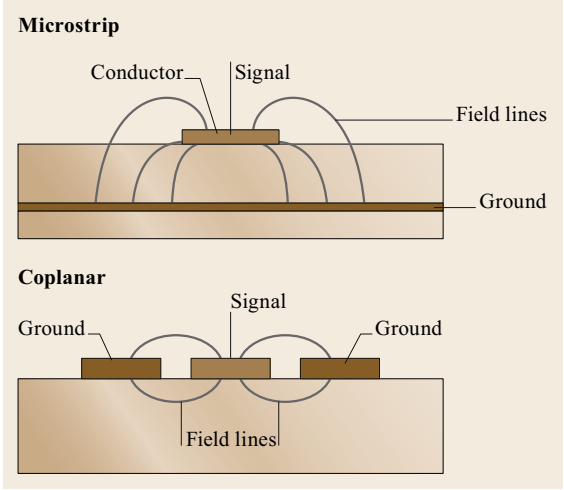


Fig. 53.8 Schematic illustration of RF electric field lines in the microstrip and coplanar designs

applications to the lower-cost solution of LTCC and organic. The metalizations used are Cu on the interior of the substrates and Au as the top layer metalization to improve RF performance.

Organic substrates offer a lower-cost solution. This substrate is built in a laminate structure with punched or laser drilled holes that are subsequently plated for interlayer via interconnects. The laminant layers are variations of epoxy and glass optimized for low dielectric constants. LTCC substrates are also laminant structures of sheets of LTCC with laser drilled and plated vias. Recently, LTCC materials have been optimized to not shrink in the *x*-*y* plane during the sintering process (*z*-direction shrinkage, however, is significant). The LTCC laminants are extremely thin compared to organic, and offer the opportunity to build complex RF substrates.

An additional trend in RF substrate technology is to move the discrete functions of RF module circuits from surface-mounted components to embedding them in the substrate. These embedded functions are typically capacitors and inductors and occasionally resistors. This can be done because the substrate material is dielectric and metal; the basic components of passive devices. The advantages of embedding components are reduced size, reduced cost and improved performance. Size can be reduced because surface-mounted components take up a great deal of the surface of the substrate. By embedding these components, they are thinner and area is reduced. Cost is reduced because the embedded passives are formed as part of the laminant layers by patterning the metal/dielectric and is essentially *free* because the patterning would be done regardless. Performance can be increased through the smaller size of the module. The embedding of components in substrates does offer opportunity for improved materials. Performance could be improved by finding means to alter the material properties of regions of the laminant layers to optimize the electrical performance of discrete components that are to be embedded (i. e., *seed* regions of the substrate for optimal capacitance, inductance and resistance).

53.7 Underfill and Encapsulants

53.7.1 Underfill

Flip chip technology demands that the space between the chip and the substrate be filled with a dielectric organic material to help mechanically interlock the chip to the substrate. The majority of underfill material available is a silica-filled anhydride resin polymer. The filler provides the strength in the composite polymer and the resin bonds with the die and substrate. Without the underfill the difference in coefficient of thermal expansion between the chip (3 ppm/°C) and the substrate ($\approx$ 17 ppm/°C) would quickly result in fracture of the solder interconnects when variations in temperature occurred. New underfills must be developed to accommodate increased bump density and shrinking bump height. As bump density increases and pitch decreases, the underfill must flow into smaller and smaller spaces. A 250 μm pitch requires an underfill with a thickness of 110 μm. At a 150 μm pitch, the dimension between the chip and substrate will be less than 50 μm but 100% fill of underfill is still required. Traditional flow dispensing and cure may not be suitable in these thin spaces due to a lack of capillary driving force (current underfill materials flow through a minimum thickness of 75 μm). For very thin fill spaces, wafer level deposition of the underfill needs to be developed. This wafer level deposition involves spin-on of an uncured polymer to a uniform thickness slightly less than the bump height. The wafer would then be diced and die-mounted on substrates. Reflow would form the solder joints and cure the underfill simultaneously. The filled polymer must be spun to a uniform thickness of 25 μm or less, bond well with silicon and substrate materials, and be compliant.

53.7.2 Encapsulation

Encapsulation of electronic packages is required to protect the electronic components from environmental degradation such as moisture contamination and mechanical damage, and provide electrical isolation. The encapsulation material is a polymer that flows over and covers the electronic components then cures into a solid. For electronic packaging, the most common method to apply the encapsulation is transfer molding. Figure 53.9 is a schematic illustration of the transfer mold process. The transfer mold compound material is typically a thermoset cross-linking polymer. The polymer resin portion consists of 25–35% of the total mold compound (by weight) and consists of epoxy resin, a hardener and a curing agent. The mold compound material is filled with powders of silica (SiO_2), alumina (Al_2O_3) or glass (65–75% of the total weight) to reduce the thermal expansion coefficient of the polymer and add strength. Amorphous silica is used for a minimum thermal expansion coefficient but results

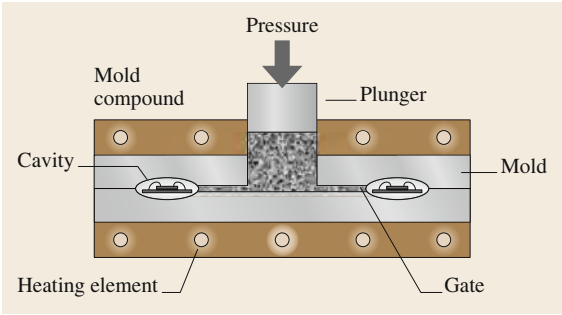


Fig. 53.9 Schematic illustration of the transfer mold process

in a decrease in thermal conductivity. Crystalline silica improves thermal conductivity but is abrasive to the molds. Small amounts of pigments (to color the mold compound), accelerants, flame retardants, mold release agents, getterers, and antioxidants are also added to the final mold compound. Pellets of the mold compound are placed into a heated cavity at the top of the mold and pressure is applied by a ram that pushes the liquefying materials into the mold. The material is forced through channels and through a constriction (called the gate) that provides back pressure and regulates the flow rate of the molten mold compound. After passing through the gate, the mold material flows across the component being encapsulated. The heat is turned off and upon cooling the polymer cross-links, with the use of a curing agent (typically an amine) that is present in the mold compound. The mold compound material has good adhesion, good stability with regards to temperature, chemicals, and the environment. They cure rapidly, have low permeability to moisture and high dielectric strength.

The transfer mold process has the advantage that part dimensions are very well controlled. The process and materials are low cost and high-volume manufacturing is possible.

An issue with mold compound encapsulation processing is wire sweep where the force of the viscous encapsulant can push bond wires to short with one another. This has been addressed with optimized filler size and shape and the mold temperature and pressure. After molding, the difference in thermal expansion between the cured mold compound and the rest of the package (the substrate, leadframe, semiconductor device, etc.) can result in considerable stresses at the interfaces. Furthermore, the CTE of mold compounds changes when the material goes above the glass transition temperature, T_g. T_g is the temperature at which the polymer changes from a rigid structure to a softer more glass-like structure. Below T_g, the CTE of typical mold compounds is $20\times10^{-6}\,(^{\circ}\mathrm{C})^{-1}$ but above T_g this can increase to $60\times10^{-6}\,(^{\circ}\mathrm{C})^{-1}$ or $70\times10^{-6}\,(^{\circ}\mathrm{C})^{-1}$. The T_g of mold compounds varies but is usually 175–200 °C, below the solder processing temperatures that are typically in excess of 220 °C. After cooling from mold processing, the mold compound forms a compression seal with the lower expanding materials (leadframe, substrate, semiconductor). However, upon heating to solder process temperatures the stress is reversed to tensile and can result in interfacial cracking, or delamination. The interfacial cracks provide a conduit for the ingress of water or ionic contaminants that could result in corrosion of metals in the package (particularly Al bond pads). This issue must be addressed by optimizing the mold compound to the package application (materials and size) to minimize strain and to maximize interlocking mechanisms between the mold compound and the substrate/leadframe.

Mold compound encapsulation is generally detrimental to RF device performance. The encapsulant materials are nonuniform and degrade performance because the electric field generated by the RF circuit in the encapsulant is nonuniform resulting in lossy behavior. Many high-performance RF packages are designed with a cavity and are made out of ceramic with a metal lid to form the cavity. However, plastic encapsulation is required for suitable, low-cost, reliable packages, so a trade-off must be made for performance, reliability and cost.

53.8 Electrically Conductive Adhesives (ECAs)

Electrically conductive adhesives (ECAs) are composite materials consisting of a dielectric curable polymer and metallic conductive particles. ECAs are low-process temperature alternatives to solder alloys. The polymer is an adhesive material that chemically reacts with metals to form a bond. The metallic particles in the adhesive form a network in the cured joint that forms a conduction path in area-array applications.

53.8.1 Adhesive Polymers

The polymer portion of the adhesive is available in two forms: thermoplastic and thermosetting. A schematic illustration of these two types of polymers is shown in Fig. 53.10. A thermoplastic polymer is a linear chain of linked monomers (nominally, hydrogen-carbon bonds). After curing, thermoplastic polymers become

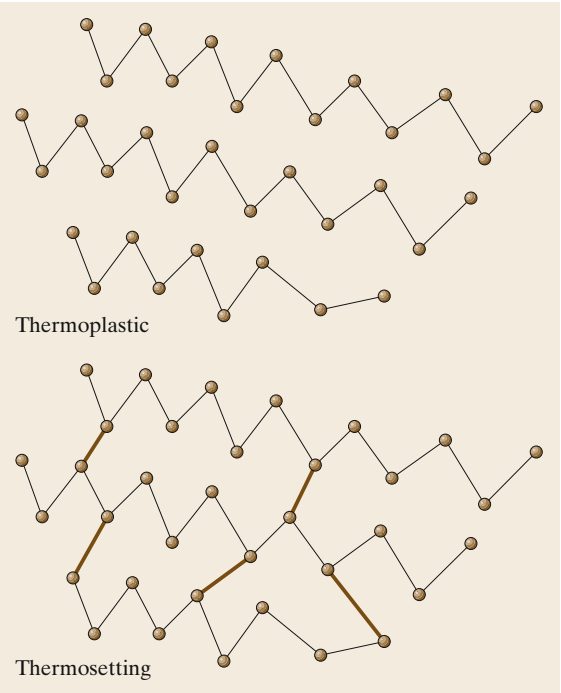


Fig. 53.10 Schematic illustration of polymer bonds for thermoplastic (linear bonds) and thermosetting (cross-linked bonds) electrically conductive adhesives

more rigid, but remain flexible. As temperature is increased, these polymers essentially become molten. This provides a method to repair or replace components with thermoplastic joint area-array joints. Thermosetting materials are similar to thermoplastic, except that monomer chains become cross-linked in three dimensions during the curing stage. The number of cross-links determines the rigidity and the glass transition temperature (T_g) of a thermoset polymer. In general, the more cross-links, the higher the T_g. A curing agent (e.g., a hardener) can be added to thermoplastic polymers to cause the cross-linking reaction in these materials as well. Curing can be performed using UV light, heat, or catalysts. Thermosetting materials are harder than thermoplastic, but can be processed and cured below their T_g and do not become *molten* above the T_g. However, the curing operation is not reversible, making repair difficult. Examples of thermosetting materials include epoxies and acrylics. ECAs are typically made using either thermosetting or thermoplastic polymers, or a combination of the two.

53.8.2 Metal Fillers

A number of metals have been used to provide the electrical conductivity of ECAs. The metals typically are in the form of flakes, plates, rods, fibers, or spheres. The size of the metal particles ranges between 5 and 20 μm. For the ECA to have suitable strength and electrical conduction requires a trade-off in metal filler content. The greater the metal content, the better the electrical conduction, but the poorer the strength of the adhesive. The metal content depends upon whether the adhesive is an isotropic or anisotropic conductor (described below), but does not usually exceed 40% by volume.

The metals most often used in ECAs are silver, gold, nickel, copper or variations of these (such as gold on nickel, silver on glass, or nickel on polymers). Silver-filled ECAs are the most common, because the cost of the metal is moderate and has good electrical conductivity and low reactivity with oxygen. Gold has better physical properties than silver but the cost is prohibitive for most applications. Nickel is less expensive, but also has lower conductivity than either Au or Ag and has been found to corrode when aged in a humid environment resulting in poor adhesion between the nickel and the epoxy matrix. Copper fillers oxidize rapidly and also delaminate from the polymer matrix.

53.8.3 Conduction Mechanisms

The mechanism of conduction in ECAs is hypothesized to occur via percolation where current travels through a three-dimensional matrix of contacting particles. There must be a sufficient quantity of particles present in the adhesive to provide a suitable conduction path. The bulk resistivity of 60Sn-40Pb solder is 15×10^{-6} Ωcm, while that of ECAs range from 7×10^{-5} to 5×10^{-4} Ωcm [53.24]. Therefore, the conduction of solder is 5–33 times better than ECAs. This difference in conductivity is due to the fact that only a portion of the joint cross-section is available to provide the percolation path in ECAs whereas the entire cross-section of solder joints conducts electricity. The loss of electrical conductivity does not severely affect second-level interconnects but could be an issue for first-level interconnects where good conductivity and switching speed may be required. For the current to pass from one metal particle to another, it must pass through, at a minimum, the oxide that forms on the metals or, at worst, some distance through the polymer matrix. It has been proposed that contact resistance is overcome by tunneling or bulk conduction through the semiconducting metal oxide layer. However, measured electrical conduction has been determined to be better than any of the mechanisms described above [53.25]. It is possible that, in silver-loaded ECAs, silver diffuses out of the particles into microcracks in the polymer matrix, thereby creating additional conduction paths. To date, the details of the conduction mechanisms remain unresolved.

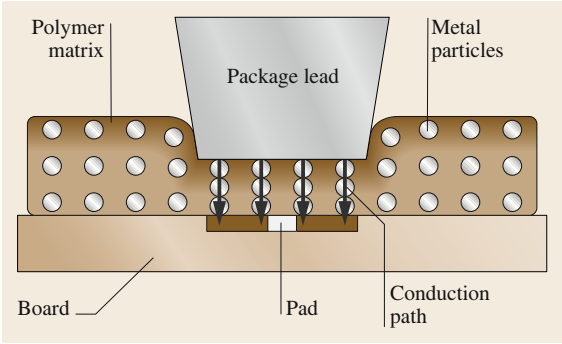


Fig. 53.11 Schematic illustration of an anisotropic electrically conductive adhesive. The adhesive is compressed causing the conductive particles to contact one another in the *z*-axis forming an electrical path between terminal pads. Outside the pad region, the metal particles are separated by the polymer and the AECA acts as a dielectric

53.8.4 Isotropic Versus Anisotropic Conduction

ECAs are available in two types of conductors, isotropic and anisotropic. In isotropic adhesives, the conduction path is uniform in all directions with a filler metal content sufficiently high to create a percolation path for the conductor. Anisotropic ECAs (AECAs, also called a *z*-axis conductors) only have one direction of conduction, useful for fine-pitched applications. In general, anisotropic ECAs are prepared by dispensing metal particles in the polymer matrix at a level far below the percolation threshold. Pressure is imposed on the contact so that the particles come into contact with one another in one direction (parallel to the direction of imposed pressure). In the areas where no pressure is imposed, there is no conduction path and the adhesive is electrically insulated. This is illustrated in Figure 53.11. In area-array applications, anisotropic ECAs can be applied to the board but conduction will only take place between package and mating board pads due to the compression resulting from the cured adhesive causing particle-to-particle contact between mating pads. No metal particle contact is made beyond the terminal pads, thus the adhesive is an insulator in those areas.

The terminal pad surface to which ECAs or AECAs are to be joined must be oxide free for the adhesive to form a cohesive joint. Bare copper typically provides very poor adhesion [53.26] because adhesives do not wet or react well with metal oxides, and in this way adhesives are similar to solder alloys. To overcome this problem, copper can be plated with nickel ($\approx 0.5\,\mu m$) and a thin layer of gold ($\approx 500\,Å$). Gold is an excellent bonding surface because it is virtually free of surface oxide and provides a diffusion barrier for nickel. Unlike solders, ECAs do not allow the luxury of using flux to remove oxides during processing so oxide removal surface treatments must precede dispensing the adhesive. The stability of ECA joints is directly related to the quality of the bond interface with terminal pad metalizations. In general, the less surface oxide present, the better the electrical stability. If surface oxide is present, there can be de-adhesion at the interface resulting in a reduced ability to carry electrical signals. The bonding of ECAs to other polymers, such as printed wiring board surfaces (FR-4 or polyimide) is excellent. Area-array component requirements are more stringent in the case of ECA joints compared with solder bumps. Solders only wet metal surfaces, so special processing steps are not needed to define solder-joint locations. For ECAs, the method of depositing the ECA must be precise (or use an anisotropic ECA) so that electrical shorts do not occur and this is exasperated in area-array fine pitch applications and the interconnects not easily inspected because they are hidden beneath the joined component.

53.8.5 Rework

The ability to perform rework on conductive adhesive interconnects may limit the use of these materials for many applications. Thermoplastic adhesives can be reworked by heating and *melting* the joint allowing components to be separated from the assembly. Thermosetting bonds are not reversible but can be softened by heating well above T_g and then removed. Some acrylic adhesives can be cleaved at room temperature and physically removed, although the brittle nature of the acrylics that allows for removal may also represent too great a joint reliability risk.

53.9 Thermal Issues

53.9.1 Thermal Issues in Digital Packaging

Increasing circuit and component density creates thermal complications that must be resolved by the packaging solution. Excessive temperatures in the devices result in degraded circuit performance or reliability issues. The thermal performance can be improved by using packaging materials with high thermal conductivity.

The operating temperature of Si directly impacts switching speed and timing. In the operating temper-

ature range of 50–100 °C, a 5 °C drop in operating temperature results in a 1% increase in CMOS switching speed. From a reliability perspective, each 10 °C increase in operating temperature results in a 2× increase in the die failure rate. Therefore, packaging materials and processes that improve thermal performance are very important.

New materials are the area where the most significant improvements in thermal performance are possible. Improved thermal interface materials (greases, adhesives, tapes, phase change materials) will improve the heat flow out of the package. Improved adhesives and compliant sheets inside the package will improve heat dissipation from the die and can better accommodate CTE mismatch strains within the package.

Composite materials for package lids, heat spreaders and heat sinks are needed. Currently, these components face a dilemma of conflicting material properties of good thermal conduction materials (e.g., Al) have very high CTE (e.g., Al = 25×10^{-6} /°C) that can result in poor reliability. Furthermore, most low-cost thermal conductors are also electrical conductors making their use in dielectric packages problematic. The development of composite materials with anisotropic conductivities can provide excellent thermal conduction with matched CTE and the potential for a good electrical isolation. At the die level, the use of low-cost thin-film diamond must increase. Diamond has excellent thermal conductivity and isolation but is currently too expensive for extensive use in microelectronic packaging.

53.9.2 Thermal Issues in RF Packaging

RF die can generate a great deal of heat in operating at high frequencies. This is particularly an issue with GaAs die because it has poor thermal conductivity (30W/mK). The generated heat must be dissipated through the package. The preferred method to remove heat from wirebonded GaAs die is to thin the GaAs wafer by back-grinding to 100–125 μm. The thinner the die, the easier it is to transmit heat through the GaAs. Through-die vias are also used to help dissipate heat. The vias are formed by laser drilling holes through the GaAs then plating the side walls with Au and thinning the wafer. Thinning the wafer not only benefits heat dissipation but also reduces the distance from the conducting layer on the surface of the die to the ground plane on the die backside that improves impedance match in microstrip designs.

The problems with through-die vias and wafer thinning are cost and die breakage. Laser drilling and thinning processes increase the cost of the die. Furthermore, GaAs is a very brittle material, so that the presence of vias and thinning exacerbates potential cracking.

One method to reduce the thermal and mechanical problems associated with thermal issues on GaAs is to use flip chip interconnects. For flip chip, the bumps act as electrical, mechanical and thermal interconnects. The bumps are on the front side of the die so full thickness die can be used. The challenge with flip chip as a thermal conductor is to put sufficient numbers of bumps on the die to dissipate the heat.

53.10 Summary

Materials are one of the most important aspects of microelectronic packaging. There is a need for new materials to be developed to meet the ever-changing requirements for high-speed digital and RF applications. Considerable resources are required to develop and characterize these materials. These materials must also be tested to determine their compatibility with assembly processes and device performance and meet industry standards. The fundamental material characteristics that must be understood, depending upon the material, include: dielectric constant, dielectric loss, conductivity, resistivity, moisture absorption, glass transition temperature, strength, time dependent deformation (creep), and fracture toughness.

References

53.1 D.R. Frear, W.B. Jones, K.R. Kinsman: *Solder Mechanics: A State of the Art Assessment* (TMS, Warrendale 1991)

53.2 K. Puttlitz, P. Totta (Eds.): *Area Array Packaging* (Kluwer, Dordrecht 2001)

53.3 P.A. McCluskey (Ed.): *High Temperature Electronics* (CRC, Boca Raton 1996)

53.4 J.H. Lau: *Ball Grid Array Technology* (McGraw–Hill, New York 1995)

53.5 G. Harmann: *Wire Bonding in Microelectronics Materials, Processes, Reliability, and Yield* (McGraw–Hill, New York 1997)

53.6 R.J. Klein Wassink: *Soldering in Electronics* (Electrochem. Publ., Ayr, Scottland 1989)

53.7 C.A. Harper: *Electronic Packaging and Interconnection Handbook* (McGraw–Hill, New York 1991)
53.8 M.L. Minges (Ed.): *Electronic Materials Handbook: Packaging*, Vol. 1 (ASM–Int., Materials Park 1989)
53.9 H. Clauberg, I. Qin, P. Reid, B. Chylak: Fine pitch copper wire bonding, Chip Scale Rev. **14**(6) (2010)
53.10 P. Chauhan, A. Choubey, Z.W. Zhong, M. Pecht: *Copper Wire Bonding* (Springer, New York 2014)
53.11 E. Philofsky: Intermetallic formation in Au–Al systems, Solid State Electron. **13**, 1391 (1970)
53.12 C. Horsting: *Purple Plague and Au Purity, Proceedings 10th Annual IRPS* (IEEE, Westmoreland 1972) p. 155
53.13 L. Levine, M. Sheaffer: Wirebonding strategies to meet thin film packaging requirements – Part 1, Solid State Technol. **36**, 63 (1993)
53.14 L.S. Goldman: Geometric optimization of controlled collapse interconnects, IBM J. Res. Dev. **13**, 251 (1969)
53.15 J.–K. Lin, J. Drye, W. Lytle, T. Scharr, R. Sharma: *Conductive Polymer Bump Interconnects, Proceedings of the 46th Electronic Components Technology Conference* (IEEE, Piscataway 1996) p. 1059
53.16 N.C. Lee: *Pb–free soldering – Where the world is going, Adv. Microelectron* (IEEE, Piscataway, NJ 1999) p. 29
53.17 J. Glazer: Microstructure and mechanical properties of Pb–free solder alloys for low–cost electronic assembly: A review, J. Electron. Mater. **23**, 693 (1994)
53.18 J. Glazer: Metallurgy of low temperature Pb–free solder for electronic assembly, Int. Mater. Rev. **40**, 65 (1995)
53.19 M. Abtew, G. Selvardery: Pb–free solder in microelectronics, Mater. Sci. Eng. **27**, 95 (2000)
53.20 H.K. Seelig, D. Suraski: The status of Pb–free solder alloys, Proc. 50th Electron. Comp. Tech. Conf. (IEEE, Piscataway 2000) p. 1405
53.21 K.G. Snowden, C.G. Tanner, J.R. Thompson: Pb–free soldering interconnects: Current status and future developments, Proc. 50th Electron. Comp. Tech. Conf. (IEEE, Piscataway 2000) p. 1416
53.22 Z. Hasnain, A. Ditali: *Building–in reliability: Soft errors – a case study, Ann. Proc. Reliab. Phys* (IEEE, Westmoreland 1992) p. 276
53.23 M.W. Roberson, P.A. Deane, S. Bonafede, A. Huffman, S. Nangalia: Conversion between standard and low–alpha Pb in solder bumping production lines, J. Electron. Mater. **29**, 1274 (2000)
53.24 H.L. Hvims: Conductive adhesives for SMT and potential applications, IEEE Trans. Components Hybrids Manuf. Technol. **18**, 284 (1995)
53.25 L.C. Li, H. Lizzul, I. Kim, J.E. Sacolick, J.E. Morris: Electrical, structural and processing properties of electronic conductive adhesives, IEEE Trans. Components Hybrids Manuf. Technol. **16**, 843 (1993)
53.26 D.D.L. Chang, J.A. Fulton, H.C. Ling, M.B. Schmidt, R.E. Sinitiski, C.P. Wong: Accelerated life test of *z*–axis conductive adhesives, IEEE Trans. Component Hybrids Manuf. Technol. **16**, 836 (1993)

54. Organic Solar Cells

Masahiro Hiramoto, Yusuke Shinmura

One of the responsibilities of scientists is to help develop new ways in which we can generate energy, since gaining control over energy resources such as petroleum has been one of the main reasons for conflict between nations. Renewable energy generated by solar cells is one of the potential solutions to the problem of maintaining our energy supply and these have been studied intensively for about half-century. Today, silicon solar cells have already been commercialized and have become an indispensable source of electricity. However, the price of electricity produced by silicon solar cells is still higher than that produced by petroleum. In order to increase the production of energy by solar cells, the price of electricity produced by solar cells needs to be lower than that produced by petroleum. Organic solar cells have the potential to be part of the next generation of low-cost solar cells. There was a steep increase in the power-conversion efficiency of organic solar cells around the year 2000, indicating that the technology needed to bring them to a commercial level would be established by around 2020, taking into consideration the example of organic electroluminescent devices for which scientific breakthroughs were made in 1987 and commercialization occurred around 2010. Now, in 2015, the power-conversion efficiency of organic solar cells has reached 12%.

Organic solar cells have many advantages; they are flexible, printable, light weight, and low cost, can be fashionably designed, and can be fabricated by roll-to-roll production, etc. Printed organic solar cells can be attached to the roofs, windows, and walls of houses and buildings. Automobiles wrapped with colorfully printed organic solar cells can be fabricated. Moreover, they are suitable for constructing solar power plants in space, since their light weight allows them to be easily put into orbit. In this section, the history, fundamental principles, and recent progress in organic solar cells are summarized.

The most essential factor for organic solar cells is the existence of excitons, that is, strongly bound electron–hole pairs. To efficiently generate photocarriers from excitons, donor–acceptor sensitization is used. Fullerenes acting as acceptors are used in present organic solar cells. Since the diffusion length of excitons is extremely small, blended junctions are used. Route formation both for photogenerated electrons and holes to the respective electrodes by phase separation is required for organic blended junctions. The magnitude of the photovoltage that can be obtained is determined by the difference between the lowest unoccupied molecular orbital (LUMO) of the acceptor molecules and the highest occupied molecular orbital (HOMO) of the donor molecules. Utilization of tandem cells has been effective in increasing the power-conversion efficiency. Today, the power-conversion efficiency of organic solar cells has reached 12%. For the organic semiconductor films used in organic solar cells, both small molecule films deposited by the dry process of vacuum evaporation and polymer films deposited by the wet process of spin coating are used.

54.1 History 1330
54.2 Excitons 1331
54.3 Donor–Acceptor Sensitization 1332
54.4 Exciton Diffusion 1332
54.5 Blended Junctions 1333
54.6 Route Formation 1334
54.7 π–π Stacking 1335
54.8 HOMO–LUMO Gap 1335
54.9 Tandem Cells 1336
54.10 Conclusions 1337
References 1337

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_54

54.1 History

The first organic solar cell was fabricated by *Calvin* in 1958 (Fig. 54.1) [54.1]. For a long while, single films of small molecule organic semiconductors deposited by vacuum evaporation were used. Typical organic semiconductors were phthalocyanines (Fig. 54.2) [54.2] and merocyanine [54.3]. Most organic semiconductor films had p-type characteristics since oxygen molecules from the ambient air, acting as acceptors, inevitably doped the films. A photocurrent could be generated at the Schottky junction between the p-type organic film and a low-work function metal, such as Al. However, in the early stages of development, organic solar cells had little photocurrent, typically less than several microamperes [54.4, 5].

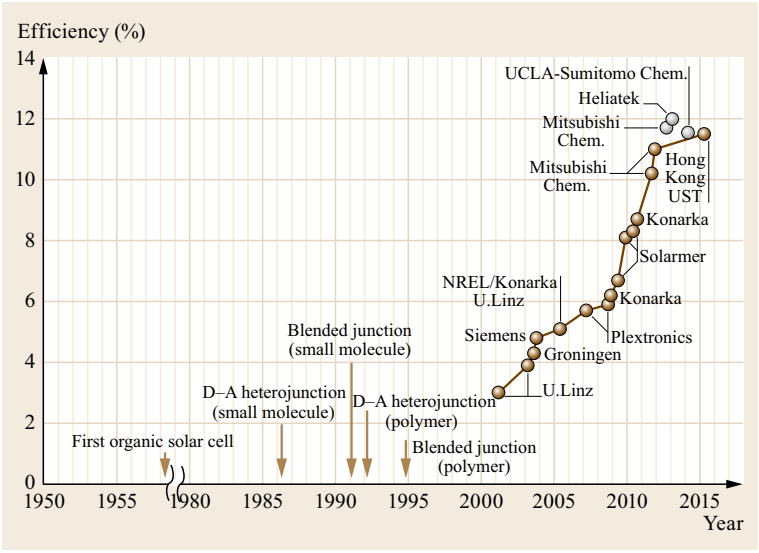


Fig. 54.1 The history of organic solar cells. The main breakthroughs are indicated by *arrows*. The efficiencies after 2000 (*dots*) are plotted according to the NREL chart. *Grey dots* are plotted according to the recent press releases

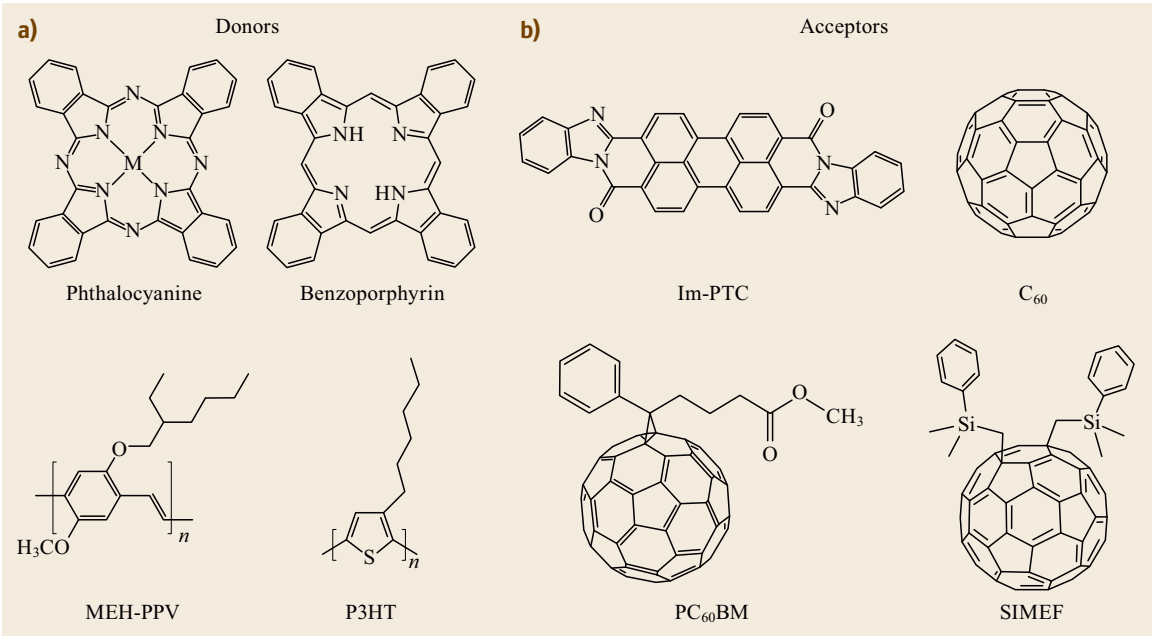


Fig. 54.2a,b Chemical structures of typical organic semiconductors acting as donors (**a**) and acceptors (**b**)

A breakthrough occurred in 1986, namely, the development of a two-layer organic photovoltaic cell, which had a large photocurrent density, of the order of mA/cm^2, and an efficiency of 1% (Fig. 54.1). This was as a result of the donor–acceptor sensitization reported by *Tang* [54.6], which has had a big impact on the field of organic solar cells. A blended junction for small molecule cells was proposed in 1991 by *Hiramoto* [54.7, 8]. In 1992, *Sariciftci* [54.9] reported a polymer heterojunction cell composed of C_{60} and MEH-PPV (poly[2-methoxy-5-(2-ethylhexyloxy)-1,4-phenylenevinylene]) (Fig. 54.2), which was followed in 1995 by a bulk polymer heterojunction cell reported by *Yu* [54.10]. Typical combinations of donors and acceptors both for small molecular systems and polymer systems are shown in Fig. 54.2. Based on these fundamental studies, after 2000, the efficiency rose from 1% and started to increase rapidly and reached 12% by 2013. Nowadays, the photocurrent density of organic solar cells produced by solar radiation is approaching $20\,mA/cm^2$, which is comparable to the values produced by inorganic solar cells.

The main contributory factors to the increase in efficiency of organic solar cells can be summarized as follows:

1. Donor–acceptor sensitization
2. The use of blended junctions
3. Phase separation in the blended junctions
4. The use of tandem cells
5. The use of fullerenes acting as acceptors
6. Syntheses of new organic semiconductors.

54.2 Excitons

The dissociation of photogenerated electron–hole pairs (excitons) is a key factor for generating carriers in organic semiconductors. Exciton dissociation is affected by the relative permittivity of the solid (ε). The attractive force between an electron–hole pair is given by Coulomb's law

$$F = \left(\frac{1}{4\pi\varepsilon\varepsilon_0}\right)\left(\frac{q_1 q_2}{r_2}\right). \tag{54.1}$$

Here, ε_0, q_1, q_2, and r are the absolute permittivity, the elementary charges on the electron and the hole, and the distance between them, respectively. In a solid with a small value of ε, the positive and negative charges experience a strong attractive force, whereas, in a solid with a large value of ε, the positive and negative charges experience a relatively weak attractive force. Inorganic semiconductors have large values for ε. For example, the value for Si is 11.9 and the exciton has a large diameter of 9.0 nm and is localized over about 10^4 Si atoms (Fig. 54.3a). The thermal energy at room temperature is sufficient for this Wannier-type exciton to dissociate immediately into a free electron and a hole, thereby generating photocurrent. On the other hand, organic semiconductors have small values for ε. For example, the value for C_{60} is 4.4 and the exciton diameter is just 0.50 nm and is localized over a single C_{60} molecule (Fig. 54.3b). The thermal energy at room temperature is hardly sufficient for these Frenkel-type excitons to dissociate into free electrons and holes and they can easily relax into the ground state (Fig. 54.4a). Therefore, organic semiconductors generate few photocarriers. This is the reason why organic solar cells fabricated before the work of *Tang* [54.6] had extremely low photocurrents, of the order of less than a microampere.

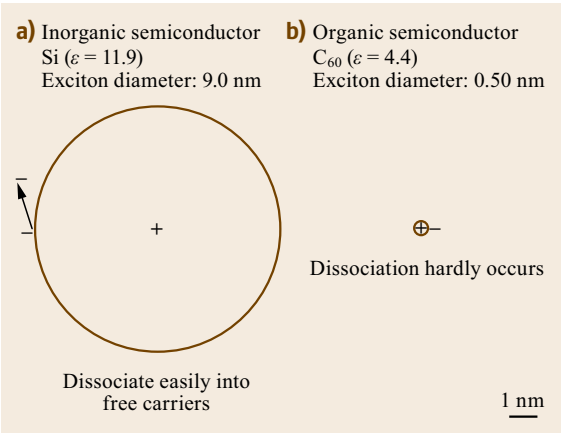


Fig. 54.3a,b Size of excitons for **(a)** an inorganic semiconductor (Si) and **(b)** an organic semiconductor (C_{60}). The former is a Wannier type and easily dissociates into free carriers. The latter is a Frenkel type and dissociates into free carriers with difficulty

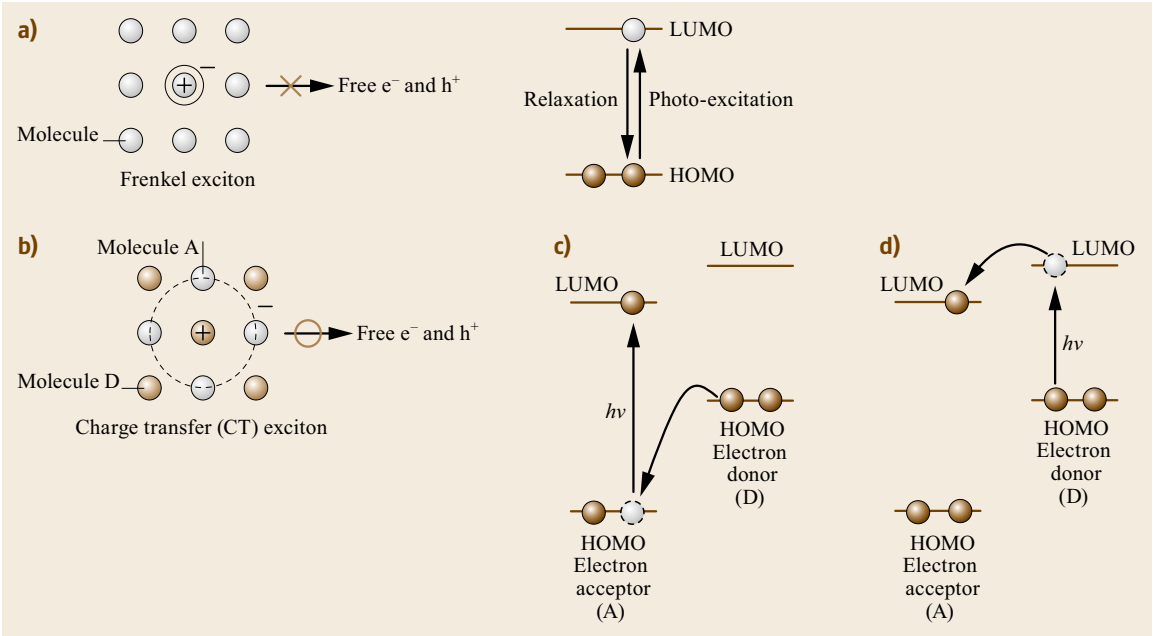


Fig. 54.4a–d Carrier generation in organic semiconductors. **(a)** Single molecule solids. **(b)** Donor(D)/acceptor(A) sensitization for carrier generation by mixing two kinds of organic semiconductor molecules. Efficient free carrier generation occurs from the charge transfer (CT) exciton. **(c)** Photoinduced electron transfer from the HOMO of the donor molecule (D) to the HOMO of the acceptor molecule (A). **(d)** Photoinduced electron transfer from the LUMO of the donor molecule (D) to the LUMO of the acceptor molecule (A)

54.3 Donor–Acceptor Sensitization

Today's organic solar cells have overcome the above problem by combining two kinds of organic semiconductor. An electron-donating molecule (D) and an electron-accepting molecule (A) for which the energetic relationship of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are brought together in contact or blended. When the electron-accepting molecule (A) is excited, electron transfer from the HOMO of the electron-donating molecule (D) to the HOMO of the electron-accepting molecule (A) occurs, and as a result, the A molecule charges negatively (A^-) and the D molecule charges positively (D^+) (Fig. 54.4c). On the other hand, when the electron-donating molecule (D) is excited, electron transfer from the LUMO of the electron-donating molecule (D) to the LUMO of the electron-accepting molecule (A) occurs, and as a result, the A and D molecules charge negatively (A^-) and positively (D^+), respectively (Fig. 54.4d). Irrespective of the excitation of molecules (A) and (D), the charge transferred states (D^+A^-) obtained are the same. Thus, a CT exciton is formed in which the positive and negative charges are separated on neighboring D and A molecules due to photoinduced electron transfer (Fig. 54.4b). This CT exciton can dissociate to a free electron and a hole due to the thermal energy at room temperature. By utilizing this donor–acceptor (D–A) sensitization, organic semiconductors have become capable of generating photocurrents of significant magnitude; of the order of milliamperes, by solar radiation.

54.4 Exciton Diffusion

A two-layer organic solar cell (Fig. 54.5) [54.6] utilizes D–A sensitization at the heterojunction. The width of the photoactive region (*shaded brown*) is, however, limited to around 10 nm in the vicinity of the heterojunction

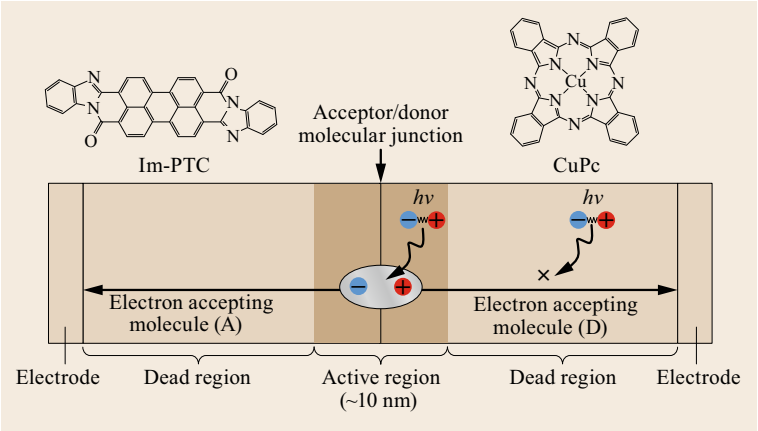


Fig. 54.5 Schematic illustration of a two-layer cell composed of perylene pigment (Im-PTC) acting as an acceptor molecule (A) and copper phthalocyanine (CuPc) acting as a donor molecule (D). Photocurrent is generated only in the active region (*shaded brown*) close to the heterojunction and all other parts of the organic films act as dead regions

due to the extremely small exciton diffusion length of just several nanometers [54.11, 12]. So, when the thicknesses of the organic layers are increased, the so-called *masking effect* occurs, namely, a dead region develops in front of the active region in which the incident solar light is absorbed and no photocurrent is generated, and as a result, the magnitude of the photocurrent is severely suppressed. A 10 nm-thick organic film can only absorb a small part of the incident solar light. However, in order to increase the efficiency of organic solar cells, it is necessary for the whole of the incident solar light to be absorbed in the 10 nm-thick active layer.

54.5 Blended Junctions

In order to overcome this problem, blended junctions in organic semiconductors, which contain both donors and acceptors, were fabricated using the codeposition technique for small molecule organic solar cells (Fig. 54.6a) [54.7, 8]. From the standpoint of photocarrier generation occurring at the molecular level, there are donor–acceptor molecule contacts acting as photocarrier generation sites due to donor–acceptor sensitization in the bulk of the codeposited layer. Blended junctions, usually called *bulk heterojunctions*, have become crucial for polymer organic solar cells [54.10].

Blended junctions are fundamentally physically the same as the porous TiO_2 in the dye-sensitized solar cells proposed by *Grätzel* in 1991 [54.13]. For both solid–solid and solid–electrolyte junctions, by transmitting the incident light through a vast number of heterointerfaces, the whole of the incident solar light can be absorbed in extremely thin active layers.

Figure 54.6b shows the energetic structure of the blended junction in a small molecule cell. Photogenerated electrons in donor molecules transfer from the LUMO of the donor molecule to that of the acceptor molecule and photogenerated holes in acceptor molecules transfer from the HOMO of the acceptor molecule to that of the donor molecule. This D–A sensitization promotes photocurrent generation. The original concept is that the positive and negative charges from ionized donors and acceptors in n- and p-type organic semiconductors, respectively, are compensated by each

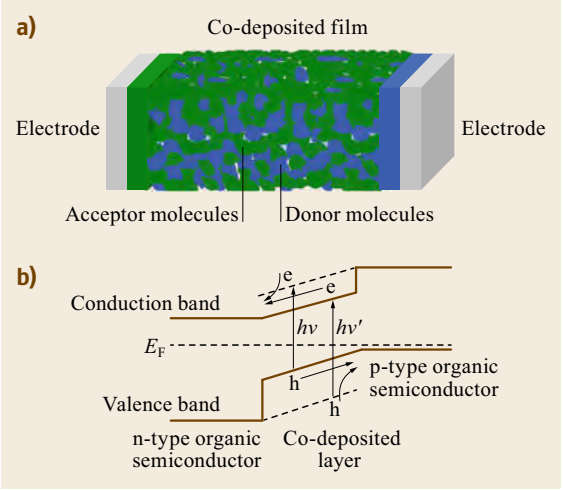


Fig. 54.6 **(a)** A small molecule cell consisting of blended junctions in an organic semiconductor, containing both donor and acceptor molecules. The entire bulk of the blended layer acts as an active layer for photocarrier generation. **(b)** Energetic structure of a blended junction in a small molecule cell

other, and the resulting codeposited interlayer behaves like an intrinsic semiconductor. With regard to the formation of the built-in potential in a molecular solid, the built-in electric field is distributed across an i-interlayer sandwiched between n- and p-layers, similar to the case of amorphous silicon incorporating a p-i-n junction.

54.6 Route Formation

Even if exciton dissociation occurs, nanostructure control of the codeposited films, that is, route formation for electrons and holes generated by exciton dissociation, is crucial in order to extract a significant portion of the photogenerated charge to the external circuit. Once the two kinds of organic semiconductors acting as donor and acceptor layers are blended, extraction of photogenerated holes and electrons inevitably becomes a tough problem, since they can only move through the electron-donating molecules (D) and the electron-accepting molecules (A):

1. *Molecular mixture*: When the two kinds of organic molecules are blended by the codeposition technique at room temperature, the codeposited films usually have an amorphous molecularly blended structure (Fig. 54.6a). Only a small number of carriers can be extracted since there are few routes for electron and hole transport.
2. *Phase-separated structure*: Routes for the electrons and holes are formed by percolation due to self-assembly promoted by annealing, while maintaining D–A molecular contacts in the bulk of the organic film (Fig. 54.7a) [54.15, 16].
3. *Vertical superlattice structure*: An ideal nanostructure is the *vertical superlattice* structure (Fig. 54.7b). This structure enables the efficient dissociation of photogenerated excitons at the D–A interfaces within the exciton diffusion length (5−10 nm) and the transport of electrons and holes

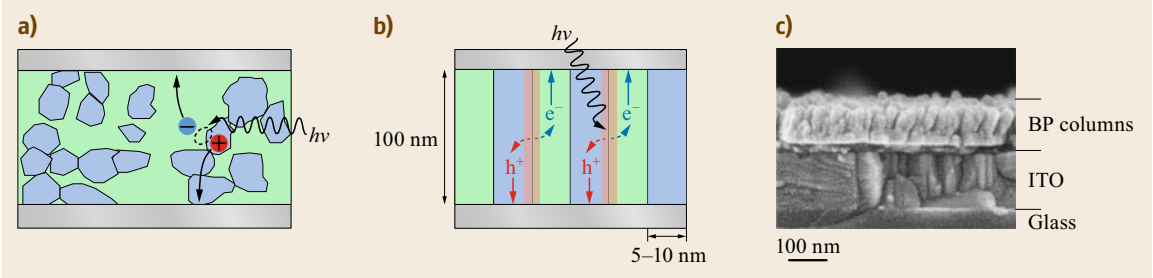


Fig. 54.7 (a) Phase-separated structure. (b) Vertical superlattice structure. (c) Interdigitated structure of the round columns of crystalline benzoporphyrine (BP) with diameters of around 20 nm standing almost vertically, used in highly efficient cells. (Reprinted with permission from [54.14]. © American Chemical Society)

Fig. 54.8 (a) Commercial flexible see-through organic solar cell module. (Courtesy of Mitsubishi Chemical Corporation). (b) Building using these modules on the exterior walls and windows. (Courtesy of Taisei Corporation)

to the respective electrodes. By changing the layer width, we can determine the exciton diffusion length to several nanometers [54.12]. It is a difficult task to construct such a structure by artificial design over a large area.

There are reports of the successful construction of such ideal nanostructures. A pseudo-vertical superlattice structure was fabricated by combining a new type of C_{60} derivative (bis(dimethylphenylsilylmethyl)[60]fullerene (SIMEF)), Fig. 54.2) with benzoporphyrin (BP, Fig. 54.2) [54.14]. BP is formed by thermal conversion from a precursor. Fabrication by wet processing was adopted, that is, the blended structure (Fig. 54.6a) was fabricated by a spin-coating technique. After removal of the C_{60} derivative (SIMEF) by washing in toluene, the desired interdigitated structure is made clearly visible (Fig. 54.7c). The round columns of crystalline BP with diameters of around 20 nm stand almost vertically. This structure allows excitons to dissociate within their diffusion length and ostensibly inhibits the recombination of photogenerated electrons and holes. In 2013, Mitsubishi Chemical Co. Ltd. succeeded in obtaining a power-conversion efficiency of 12% with this system. A commercial version using this system is shown in Fig. 54.8a,b, which shows a flexible see-through organic solar cell module and a demonstration building using these modules as exterior walls and windows [54.17]. This is a clear demonstration of the ideal vertical structure (Fig. 54.7c) showing practical conversion efficiency.

54.7 π–π Stacking

For efficient hole transport in blended junctions, π–π stacking of the organic semiconductor is necessary. The formation of π–π stacking of poly(3-hexylthiophene) (P3HT) in a [6,6]-Phenyl-C61 butyric acid methyl ester (PCBM):P3HT (Fig. 54.2) blend by means of a regular arrangement of long-chain substituents allowed the fabrication of highly efficient polymer solar cells [54.18].

Since the organic films are sandwiched between two electrodes, perpendicular orientation of the π–π stack to the substrate is more favorable to hole transport in the blended junctions. That is, face-on orientation (Fig. 54.9a) is more suitable than edge-on orientation (Fig. 54.9b). Blended cells using polymers with face-on orientation enabled high photocurrents to be obtained with thicker active layers without reducing the fill factor, resulting in an increase in efficiency [54.19].

54.8 HOMO–LUMO Gap

The maximum open-circuit voltage (V_{oc}) of photovoltaic cells consisting of a combination of donor–acceptor organic semiconductors is determined by the energetic difference between the LUMO of the acceptor molecule and the HOMO of the donor molecule ($V_{LUMO–HOMO}$) (Fig. 54.10a) [54.20]. In order to accomplish efficient exciton dissociation, about 0.3 eV is reported to be necessary, that is, the observed V_{oc} is around 0.3 V smaller than $V_{LUMO–HOMO}$.

In other words, in order to obtain high V_{oc}, a combination of donor–acceptor organic semiconductors having a larger $V_{LUMO–HOMO}$ difference is necessary (Fig. 54.10). One of the main reasons for the recent increase in power-conversion efficiency is the increase in V_{oc} due to the newly synthesized donor molecules having deeper HOMO levels (Fig.54.10b) [54.20] and the acceptor molecules having shallower LUMO levels (Fig. 54.10c) [54.21].

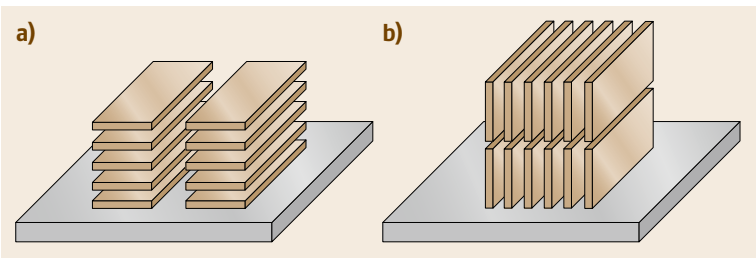


Fig. 54.9 (**a**) Face-on and (**b**) edge-on orientations of π–π stacking. Molecular planes are indicated in the figure. π–π stackings are formed perpendicular to the molecular planes

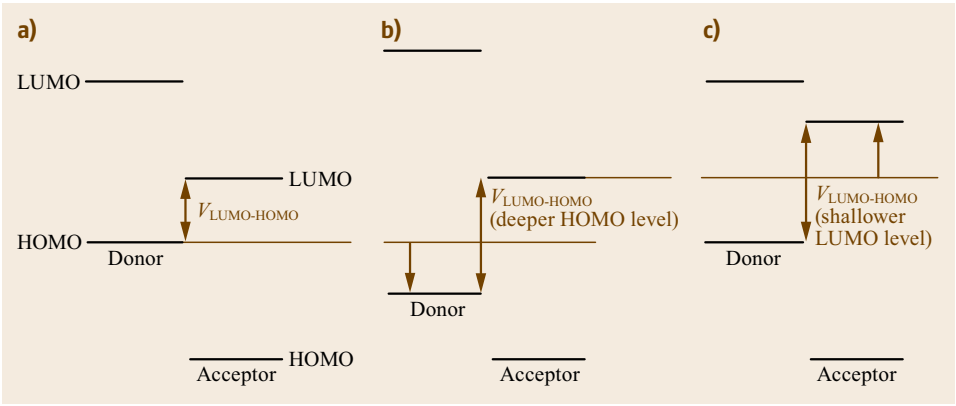


Fig. 54.10 **(a)** $V_{LUMO–HOMO}$ difference of donor and acceptor molecules. **(b)** $V_{LUMO–HOMO}$ difference using a donor molecule having a deeper HOMO level. **(c)** $V_{LUMO–HOMO}$ difference using an acceptor molecule having shallower LUMO level

54.9 Tandem Cells

The first organic tandem cell in which two unit cells were connected by metal nanoparticles was reported in 1990 (Fig. 54.11a) [54.22]. V_{oc} can be doubled by connecting two unit cells. Due to the very small diffusion length of excitons, even two very thin donor–acceptor heterojunction cells with identical absorption spectra joined together can increase the total effective thickness of the cells (Fig. 54.11b) [54.23]. Moreover, an efficiency of 5.3% was obtained with a tandem cell comprising two thin blended cells with identical absorption spectra (Fig. 54.11c) [54.24].

However, a more essential target of cells is to utilize the whole of the solar spectrum by combining cells that absorb light from different parts of the spectrum. Polymer tandem cells having double junctions [54.25] and triple junctions [54.26] utilizing the wide wavelength region from 300 to 900 nm had a V_{oc} of 2.28 V and an efficiency of 11.55%.

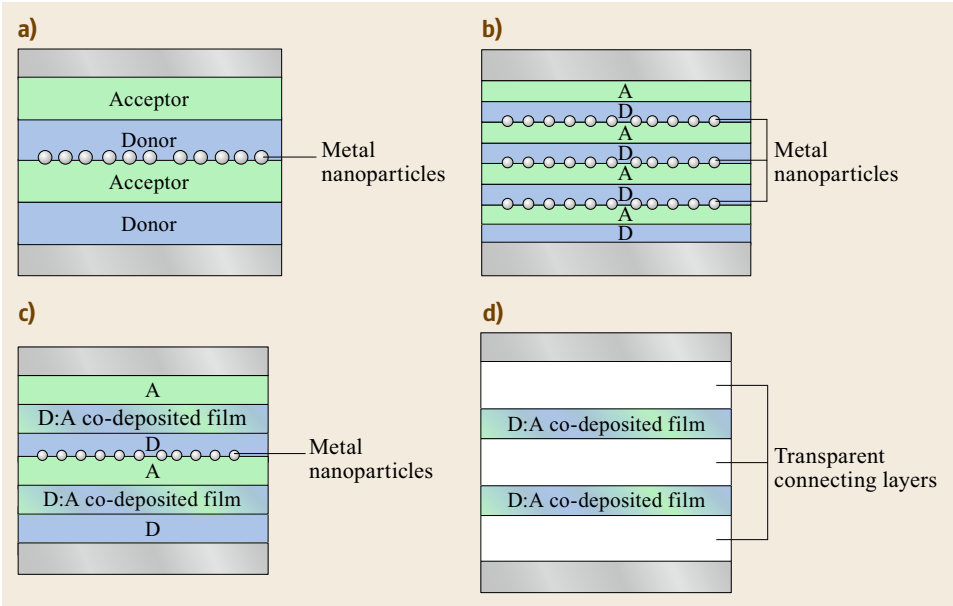


Fig. 54.11a–d Structures of organic tandem cells. **(a)** Two heterojunction cells connected together. **(b)** A series of many extremely thin heterojunction cells with identical absorption spectra. **(c)** Two thin blended cells with identical absorption spectra connected together. **(d)** Transparent connection layers with the optimum thicknesses

More sophisticated small molecule tandem solar cells (Fig. 54.11d) using transparent connecting layers of optimum thickness in which the electric field of light inside the cell is maximized in the photoactive layers of the codeposited films had an efficiency of 12% [54.27].

54.10 Conclusions

The basic components of organic solar cells, that is, excitons, D–A sensitization, exciton diffusion, blended junctions, route formation, π–π stacking, the HOMO–LUMO gap, and tandem cells, are summarized in this section. These components will help us in our search for further breakthroughs in the performance of organic solar cells. For example, (i) if we can utilize organic semiconductors with high ε values, exciton dissociation would no longer be a limiting factor for organic solar cells. (ii) If we can find organic semiconductors with significantly longer exciton diffusion lengths, exciton diffusion would no longer be a limiting factor for organic solar cells. On the other hand, the p-n-control of organic semiconductors by doping has become significant in the fabrication of organic solar cells [54.27–29].

Organic solar cells have the potential to provide the next generation of low-cost solar cells following on from silicon solar cells. Organic solar cells are representative of the organic electronics industry, which includes organic electroluminescent devices and organic transistors, and is set to dominate in the twenty-first century, following on from the dominance of the inorganic electronics industry in the twentieth century. The field of organic solar cells is closely related to many other research fields such as device physics, electronics, and synthetic chemistry, in which there are a vast number of researchers. The author fervently hopes that young researchers will take up the challenge of working in the interdisciplinary field of organic solar cells to help develop the next generation of cells for renewable energy. Some suggested books [54.30–33] are indicated in the references for new and young researchers entering the field of organic solar cells.

References

54.1 D. Kearns, M. Calvin: J. Chem. Phys. **29**, 950 (1958)
54.2 A.K. Ghosh, D.L. Morel, T. Feng, R.F. Shaw, C.A. Rowe Jr.: J. Appl. Phys. **45**, 230 (1974)
54.3 A.K. Ghosh, T. Feng: J. Appl. Phys. **49**, 5982 (1978)
54.4 G.A. Chamberlain: Solar Cells **8**, 47 (1983)
54.5 D. Wohrle, D. Meissner: Adv. Mater. **3**, 129 (1991)
54.6 C.W. Tang: Appl. Phys. Lett. **48**, 183 (1986)
54.7 M. Hiramoto, H. Fujiwara, M. Yokoyama: Appl. Phys. Lett. **58**, 1062 (1991)
54.8 M. Hiramoto, H. Fujiwara, M. Yokoyama: J. Appl. Phys. **72**, 3787 (1992)
54.9 N.S. Sariciftci, L. Smilowitz, A.J. Heeger, F. Wudl: Science **285**, 1474 (1992)
54.10 G. Yu, J. Gao, J.C. Hummelen, F. Wudl, A.J. Heeger: Science **270**, 1789 (1995)
54.11 P. Peumans, A. Yakimov, S.R. Forrest: J. Appl. Phys. **93**, 3693 (2003)
54.12 M. Hiramoto, T. Yamaga, M. Danno, K. Suemori, Y. Matsumura, M. Yokoyama: Appl. Phys. Lett. **88**, 213105 (2006)
54.13 B. O'Regan, M. Grätzel: Nature **353**, 737 (1991)
54.14 Y. Matsuo, Y. Sato, T. Niinomi, I. Soga, H. Tanaka, E. Nakamura: J. Am. Chem. Soc. **131**, 16048 (2009)
54.15 F. Padinger, F.R.S. Rittberger, N.S. Sariciftci: Adv. Funct. Mater. **13**, 85 (2003)
54.16 G. Li, V. Shrotriya, J. Huang, Y. Yao, T. Moriarty, K. Emery, Y. Yang: Nat. Mater. **4**, 864 (2005)
54.17 Mitsubishi Chemical/Taisei Corporation: http://www.m-kagaku.co.jp/newsreleases/00018.html, in Japanese (2014)
54.18 Y. Kim, S. Cook, S.M. Tuladhar, S.A. Choulis, J. Nelson, J.R. Durrant, D.D.C. Bradley, M. Giles, I. McCulloch, C.-S. Ha, M. Ree: Nat. Mater. **5**, 197 (2006)
54.19 I. Osaka, M. Saito, T. Koganezawa, K. Takimiya: Adv. Mater. **26**, 331 (2014)
54.20 M.C. Scharber, D. Mühlbacher, M. Koppe, P. Denk, C. Waldauf, A.J. Heeger, C.J. Brabec: Adv. Mater. **18**, 789 (2006)
54.21 Y. Matsuo: Chem. Lett. **41**, 754 (2012)
54.22 M. Hiramoto, M. Suezaki, M. Yokoyama: Chem. Lett. **1990**, 327 (1990)
54.23 A. Yakimov, S.R. Forrest: Appl. Phys. Lett. **80**, 1667 (2002)
54.24 J. Xue, S. Uchida, B.P. Rand, S.R. Forrest: Appl. Phys. Lett. **85**, 5757 (2004)
54.25 J. You, L. Dou, K. Yoshimura, T. Kato, K. Ohya, T. Moriarty, K. Emery, C. Chen, J. Gao, G. Li, Y. Yang: Nat. Commun. **4**, 1446 (2012)
54.26 C. Chen, W. Chang, K. Yoshimura, K. Ohya, J. You, J. Gao, Z. Hong, Y. Yang: Adv. Mater. **26**, 5670 (2014)
54.27 Heliatek: Heliatek consolidates its technology leadership by establishing a new world record for organic solar technology with a cell efficiency of 12% (Press Release), http://www.heliatek.com/en/press/press-releases/details/heliatek-

consolidates-its-technology-leadership-by-establishing-a-new-world-record-for-organic-solar-technology-with-a-cell-effi (2013)
54.28 B. Lüssem, M. Riede, K. Leo: Phys. Status. Solidi A **210**, 9 (2013)
54.29 M. Hiramoto, M. Kubo, Y. Shinmura, N. Ishiyama, T. Kaji, K. Sakai, T. Ohno, M. Izaki: Electronics **3**, 351 (2014)
54.30 C. Brabec, V. Dyakonov, U. Scherf (Eds.): *Organic Photovoltaics, Materials, Device Physics, and Manufacturing Technologies* (Wiley, Weinheim 2008)
54.31 F.C. Krebs (Ed.): *Stability and Degradation of Organic and Polymer Solar Cells* (Wiley, West Sussex 2012)
54.32 C. Hoth, A. Seemann, R. Steim, T. Ameri, H. Azimi, C.J. Brabec: Printed organic solar cells. In: *Solar Cell Materials, Developing Technologies*, ed. by G. Conibeer, A. Willoughby (Wiley, West Sussex 2014) p. 217
54.33 S.-S. Sun, N.S. Sariciftci (Eds.): *Organic Photovoltaics, Mechanisms, Materials, and Devices* (Taylor Francis, London 2005)

55. Materials for Terahertz Engineering

Roger Lewis

Metals reflect, plastics transmit, and water absorbs terahertz-frequency electromagnetic radiation. Such diverse responses open up a vast range of applications for terahertz materials spanning art, science, engineering, and medicine. The three main components of terahertz devices are sources, detectors, and the intervening optics. Sources include solid-state emitters, typically involving in their operation either the lattice (nonlinear optics) or the charge carriers (transient dipoles). Quantum cascade lasers, built of multiple semiconductor layers, represent a rapidly developing solid-state terahertz source. Detectors typically depend on either the crystal lattice (electro-optical detection) or the charge carrier reservoir (electronic detection) being sensitive to terahertz radiation. Terahertz components encompass metal-coated mirrors, plastic (machined, molded, or three-dimensional (3-D) printed) lenses, and waveguides, filters, and polarizers of many different materials and designs. An emerging class of components are the terahertz metamaterials.

55.1 **Overview** 1339
55.1.1 Bridging Electronics and Photonics 1340
55.1.2 Focus on Frequency 1340
55.1.3 Physical Phenomena at Terahertz Frequencies 1341
55.1.4 Terahertz Technology Timeline 1342
55.1.5 Entry into the Terahertz Literature 1342
55.2 **Terahertz Sources** 1343
55.2.1 Thermal 1343
55.2.2 Vacuum Electronic 1344
55.2.3 Solid-State Electronic 1344
55.2.4 Laser-Pumped 1345
55.2.5 Mechanical Excitation 1345
55.3 **Terahertz Detectors** 1345
55.3.1 Thermal 1345
55.3.2 Electro-Optic 1345
55.3.3 Electronic 1346
55.4 **Terahertz Components** 1346
55.4.1 Mirrors 1346
55.4.2 Lenses 1346
55.4.3 Polarizers 1346
55.4.4 Metamaterials 1346
55.5 **Conclusion** 1347
References 1347

55.1 Overview

Most materials have properties in the terahertz range of the electromagnetic spectrum quite different to the properties they possess in the visible range. Many organic materials, such as plastic, paper and cardboard, which are opaque to visible light, transmit light in the terahertz range. This makes terahertz techniques ideal for the testing of packaged materials, *seeing through* envelopes and wrapping. On the other hand water, often thought of as transparent to visible light, absorbs terahertz radiation strongly. This equips terahertz-frequency electromagnetic radiation for a variety of uses that involve detecting water in such areas as agriculture, biology, food processing, and medicine. Metals, which reflect well in the visible, reflect even better at terahertz frequencies. In combination, more exotic applications of terahertz frequencies are possible. For example, terahertz radiation can penetrate the plastic coating of an electrical cable, bounce back from the copper core, and give an idea of the integrity of the cable – even while it is in-service and live. Similarly, terahertz radiation can penetrate layers of paint on an automobile panel, reflect off the metal below, and provide information about the type and thickness of paint as well as of incipient corrosion of the hidden metal before it becomes visible. All these applications are nondestructive: the material being probed by the terahertz radiation remains intact and in situ throughout.

In brief, the physical mechanisms in play at terahertz frequencies may be classified into two broad categories. Conductors, such as metals, tend to reflect

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_55

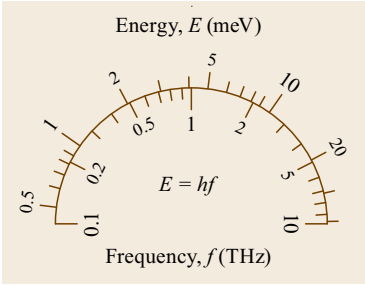


Fig. 55.1 Photon energies in the terahertz range. Roughly speaking, terahertz frequencies correspond to meV energies. The precise relationship is $E = hf$, where h is the Planck constant. To obtain the energy (in meV), multiply the frequency (in THz) by 4.135667516. To obtain the frequency (in THz), multiply the energy (in meV) by 0.2417989348. These numerical conversion factors are approximate but correct to the number of digits given

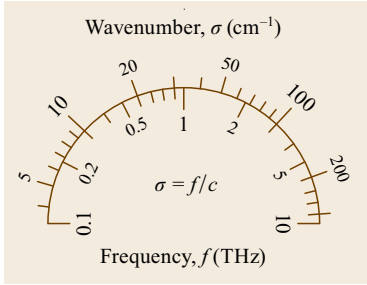


Fig. 55.2 Spatial frequencies (that is, numbers of waves per unit length, abbreviated as wavenumbers) in the terahertz range. Roughly speaking, terahertz frequencies correspond to spatial frequencies of tens per centimeter (or tens of inverse centimeters). The precise relationship is $\sigma = f/c$, where c is the speed of light. The figure is prepared for light in a vacuum. To obtain the spatial frequency (in cm^{-1}), divide the frequency (in THz) by 0.0299792458. To obtain the frequency (in THz), multiply the spatial frequency (in cm^{-1}) by 0.0299792458. These conversion factors are exact for light in a vacuum

terahertz-frequency radiation. Insulators, such as ionic crystals and many organic solids, tend to transmit terahertz radiation of many frequencies, but absorb strongly at certain frequencies, corresponding to allowed oscillations in the material. These characteristic oscillations – differing in frequency from material to material – serve as a spectral signature, and allow identification of the material under inspection.

55.1.1 Bridging Electronics and Photonics

Electronic devices conventionally operate at Hz, kHz, MHz – and even GHz frequencies – but struggle when it comes to THz frequencies. Visible light, on the other hand, is found at much higher frequencies still, of the order or 10^{14}–10^{15} Hz (0.1–1 PHz). It is difficult to extend conventional optical techniques to lower frequencies of 10^{13}, 10^{12}, or 10^{11} Hz. The terahertz region, for convenience here taken to span 10^{11} to 10^{13} Hz, thus lies in-between electronics and photonics. This fascinating frequency space, rich in science but impecunious in technology, is often referred to as the *terahertz gap*.

As set out in Fig. 55.1, terahertz-frequency photons have energies of the order of milli-electron volts. This corresponds, for light in empty space, to spatial frequencies of the order of tens of inverse centimeters, and to wavelengths of hundreds of micrometers, as illustrated in Figs. 55.2 and 55.3 respectively.

55.1.2 Focus on Frequency

Electromagnetic radiation is naturally characterized by its frequency, rather than by its wavelength [55.1, 2]. The frequency is directly proportional to the photon energy, the constant of proportionality being the

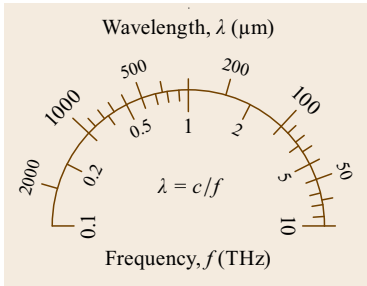


Fig. 55.3 Wavelengths of photons in the terahertz range. Roughly speaking, terahertz frequencies correspond to submillimeter wavelengths. The precise relationship is $\lambda = c/f$, where c is the speed of light. The figure is prepared for light in a vacuum and the conversion factors given are exact for a vacuum. To obtain the wavelength (in μm), divide 299.792458 by the frequency (in THz). To obtain the frequency (in THz), divide 299.792458 by the wavelength (in μm)

Planck constant, that is, $E = hf$. This is a fundamental relationship. Proportionality constant aside, energy *is* frequency. Likewise, photon frequency, considered as twice the thermal energy per degree of freedom, may be directly related to the thermodynamic temperature, T, by $T = [h/k_B]f$. In vacuum, the photon spatial frequency (or wavenumber) is a third quantity that depends directly on frequency: $\sigma = [1/c]f$, where c is the speed of light. In a medium, this relation needs to be modified by the material's properties, expressed by the refractive index. In contrast to the other quantities mentioned so far, the wavelength varies not directly,

Table 55.1 Conversion of units for terahertz photons. The quantities are related by $f = E/h$, $E = hf$; $f = k_B T/h$, $T = hf/k_B$; $f = c\sigma$, $\sigma = f/c$; $f = c/\lambda$, $\lambda = c/f$. Here c is the speed of light, h the Planck constant, and k_B the Boltzmann constant

	f Frequency (THz)	E Energy (meV)	T Temperature (K)	σ Wavenumber (in vacuum) (cm^{-1})	λ Wavelength (in vacuum)
1 K	0.02084	0.08617	1	0.6950	14.39 mm
	0.02418	0.1	1.160	0.8066	12.40 mm
1 cm	0.02998	0.1240	1.439	1	10 mm
One wavenumber	0.02998	0.1240	1.439	1	10 mm
Helium boils	0.08751	0.3619	4.2	2.919	3.426 mm
	0.1	0.4136	4.799	3.336	2.998 mm
	0.2084	0.8617	10	6.950	1.439 mm
1 meV	0.2418	1	11.60	8.066	1.240 mm
1 mm	0.2998	1.240	14.39	10	1 mm
	0.3	1.241	14.40	10.01	999.3 µm
1 THz	**1**	**4.136**	**47.99**	**33.36**	**299.8 µm**
Nitrogen boils	1.613	6.670	77.4	53.80	185.9 µm
	2.084	8.617	100	69.50	143.9 µm
	2.418	10	116.0	80.66	124.0 µm
	2.998	12.40	143.9	100	100 µm
	3	12.41	144.0	100.1	99.93 µm
Triple point	5.692	23.54	273.16	189.9	52.67 µm
Room temperature	6.108	25.26	293.15	203.7	49.08 µm
	6.251	25.85	300	208.5	47.96 µm
	10	41.36	479.9	333.6	29.98 µm
	24.18	100	1160	806.6	12.40 µm
	29.98	124.0	1439	1000	10 µm
1 µm	299.8	1240	14 388	10 000	1 µm

but inversely, with frequency: $\lambda = c/f$. Table 55.1 provides representative conversions between the various photon properties for some round numbers for each variable and for some technically important temperatures. The focus is on the two decades of frequency around 1 THz.

55.1.3 Physical Phenomena at Terahertz Frequencies

Terahertz frequencies offer a cornucopia of physical phenomena. Table 55.2 lists some of these. Representative numerical values are given for an illustrative material, the donor-doped III-V semiconductor, *n*-GaAs. Being crystalline, it exhibits distinct lattice modes; being polar, these modes readily interact with oscillating electromagnetic fields; being doped with charge carriers, it is in addition a platform for electronic phenomena.

Vibrational

Mechanical vibrations in crystalline lattices, or phonons, typically occur at terahertz frequencies. For example, in the common ionic compound NaCl the transverse-optical phonon frequency, f_t, is 4.9 THz and the longitudinal-optical phonon frequency, f_l, is 8.0 THz. These frequencies are related to the low-frequency dielectric constant, $\epsilon(0)$, and the high-frequency dielectric constant, $\epsilon(\infty)$, by $f_l^2\epsilon(\infty) = f_t^2\epsilon(0)$. Lattice properties are thus intimately related to the electric and so to the optical properties of the material.

Electronic

Oscillations of charge carriers often occur at terahertz frequencies. Free electrons oscillate at the plasma frequency, $f_p = (nq^2)/(4\pi^2\varepsilon_0 m^*)$. Here n is the density, m^* the effective mass, q the charge, and ε_0 the electric constant. Surface, interface, magneto-, and gap plasmons all occur at terahertz frequencies.

Charge carriers confined in potential wells also often have characteristic oscillations at terahertz frequencies. Donor or acceptor atoms give rise to one charge carrier moving in the Coulombic potential well of a fixed charge of opposite sign. This produces a hydrogen-like system with the characteristic energy, the Rydberg, scaled for the effective mass of the moving charge and for the dielectric constant of the material.

Table 55.2 Physical phenomena at terahertz frequencies. The numerical values given are based on *n*-GaAs

Phenomena	Equations	Examples
Vibrational		
Lattice–phonons		
– Transverse optical	f_t	$f_t = 8.2\,\text{THz}$
– Longitudinal optical	$f_l^2 = \epsilon(0) f_t^2/\epsilon(\infty)$	$f_l = 8.8\,\text{THz}$
Electronic		
Plasma frequency	$f_p = (nq^2)/(4\pi^2\varepsilon_0 m^*)$	
Electrons confined in Coulomb potential, hydrogenic systems	$Ry^* = (m^*/m_e)(1/\epsilon^2)Ry$	$Ry^* = 5.72\,\text{meV}$
Electrons confined in quantum wells	$E_n = (n^2h^2)/(8ma^2)$	width 10 nm, $E_1 \approx 1.4\,\text{THz}$
Free electrons in a magnetic field – cyclotron resonance	$\hbar\omega_c = \hbar eB/m^*$	$\hbar\omega_c \approx 1.74\,\text{meV/T}$
Bound electrons in a magnetic field – Zeeman effect	$\Delta E = \mu_B g B$	$\Delta E \approx 0.87\,\text{meV/T}$
Combined vibrational and electronic		
Polarons		
Fano resonances		

While optical transitions of hydrogen occur in the ultraviolet and the visible, the scaling means the transitions in the solid-state analogs are at terahertz frequencies. Likewise, electrons or holes confined in quantum wells, such as thin layers containing charge carriers sandwiched between barriers (made by molecular-beam epitaxy), typically have confinement energies corresponding to terahertz photons.

Charge carriers in magnetic fields undergo oscillatory motion and in many practical cases the frequency of this motion is in the terahertz range. For *free* electrons (which may, in fact, be confined to a solid, or to a two-dimensional sheet therein) the result is cyclotron motion. For *bound* electrons, such as those in a Coulombic potential, the effect is named after Zeeman.

Combined Vibrational and Electronic

Furthermore, the vibrations of the lattice may interact with the oscillations of the charge carriers. This gives rise to phenomena such as the polaron, the combination of an electron and the lattice distortion it produces, and Fano resonances, the combination of a phonon and an electronic transition between bound states being resonant with a continuum of electron states.

55.1.4 Terahertz Technology Timeline

The dramatic developments that have taken place in terahertz science and technology in recent times have been based on advances in materials science and materials engineering. Terahertz sources, for example, include vacuum devices (backward-wave oscillator, free-electron laser, synchrotron, and gyrotron) as well as gas lasers; but much of the recent development has been in solid-state sources, such as Gunn oscillators, high-electron mobility transistors, solid-state lasers of various species, and solid sources driven by ultrashort laser pulses, the pulses typically emanating from Ti:sapphire lasers. Likewise, while there exist gas-based terahertz detectors, such as the pneumatic Golay cell, most modern terahertz detectors are based on solid-state devices: Si or Ge bolometers, oxide pyroelectric materials, and, in the case of the detection of ultrafast terahertz pulses by nonlinear optical processes in polar crystals or by photoconductive devices. An overview of the progress of terahertz technologies since 1945 is given in Fig. 55.4.

55.1.5 Entry into the Terahertz Literature

Roughly speaking, the number of terahertz documents in the Scopus database has doubled every three years starting with one document in 1975. More precisely, the number of documents, D, published in any year, Y, from 1975 to 2013 inclusive is expressed well by the equation

$$D = \alpha \exp\left(\frac{Y - 1975}{\tau}\right) \qquad (55.1)$$

with parameters $\alpha = 1.55 \pm 0.42$ and $\tau = 4.66 \pm 0.18$ [55.5].

Review Articles

Overviews of terahertz science and technology appear in three key articles. Each has been cited more than one thousand times:

- *Terahertz technology*, *Siegel* [55.6]
- *Materials for terahertz science and technology*, *Ferguson* and *Zhang* [55.7]
- *Cutting-edge terahertz technology*, *Tonouchi* [55.8].

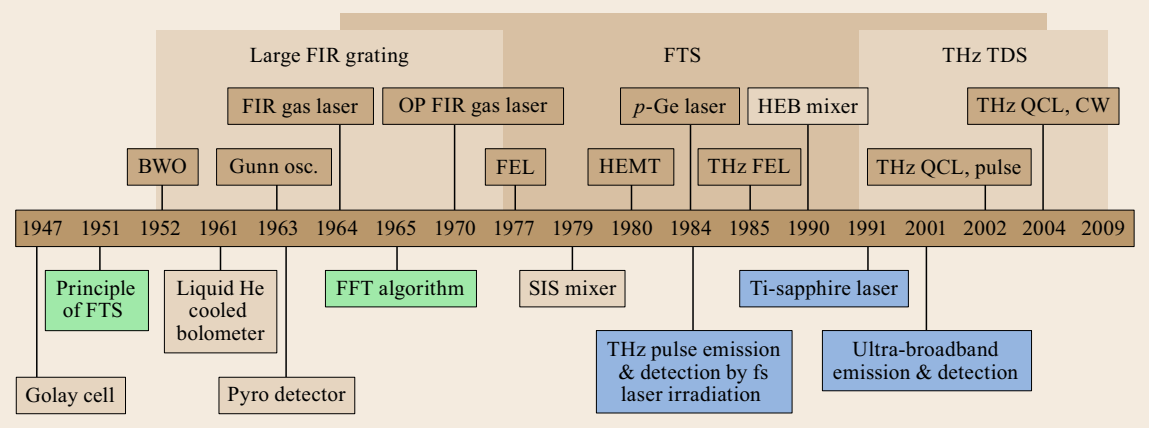


Fig. 55.4 Timeline indicating the emergence of principal terahertz technologies. (After [55.3, 4])

Further reviews cover a range of applications of terahertz science and technology:

- *Terahertz spectroscopy*, *Beard* et al. [55.9]
- *Imaging with terahertz radiation*, *Chan* et al. [55.10]
- *Pulsed terahertz tomography*, *Wang* and *Zhang* [55.11]
- *Biomedical applications of terahertz technology*, *Pickwell* and *Wallace* [55.12]
- *Present and future of terahertz communications*, *Song* and *Nagatsuma* [55.13].

Books

As the field has matured, edited volumes have been succeeded by multi-author monographs and, eventually, single-author monographs:

- 2003 *Sensing with Terahertz Radiation*, *Mittleman* [55.14]
- 2005 *Terahertz Optoelectronics*, *Sakai* [55.15]
- 2006 *Intense Terahertz Excitation of Semiconductors*, *Ganichev* and *Prettl* [55.16]
- 2008 *Terahertz Spectroscopy, Principles and Applications*, *Dexheimer* [55.17]
- 2009 *Principles of Terahertz Science and Technology*, *Lee* [55.18]
- 2010 *Introduction to Terahertz Wave Photonics*, *Zhang* and *Xu* [55.19]
- 2012 *Terahertz Techniques*, *Brundermann* et al. [55.20]
- 2012 *Terahertz Physics*, *Lewis* [55.21].

Conferences

The major conference series in the area is *The International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*. The conference series originated in the USA in 1974 with *The First International Conference on Submillimeter Waves and their Applications*. It was later renamed *The International Conference on Infrared and Millimeter Waves*. In 2004 this stream joined with the *International Conference on Terahertz Electronics*, a series based in Europe and then in its twelfth edition. IRMMW-THz is now held annually, rotating between USA, Europe, and Asia on a three-year cycle. Further information is available at www.irmmw-thz.org.

Journals

- *Journal of Infrared, Millimeter and Terahertz Waves*, previously (until 2009) the *International Journal of Infrared and Millimeter Waves*, was established in 1980. It is closely associated with the IRMMW-THz conference series.
- *IEEE Transactions of Terahertz Science and Technology* has been published since 2011 by the Institute of Electrical and Electronics Engineers.

55.2 Terahertz Sources

Terahertz sources have been reviewed recently [55.5]. A sampling of terahertz sources is presented in Table 55.3. The area is one of rapid current advancement.

55.2.1 Thermal

Thermal sources include emission from gases as in the mercury lamp [55.26], from solid-state sources such as

Table 55.3 Some sources of terahertz radiation

Source type	Representative references
Thermal	
Cosmic background radiation	*Thacker* et al. [55.22]
Globar	*Zolotarev* [55.23], *Friedel* and *Sharkey* [55.24, 25]
Mercury lamp	*Charrada* et al. [55.26]
Vacuum electronic	*Booske* et al. [55.27]
Backward-wave oscillator	*Mineo* and *Paolini* [55.28], *Xu* et al. [55.29], *He* et al. [55.30]
Extended-interaction klystron	*Shin* et al. [55.31]
Traveling-wave tube	*Bhattacharjee* et al. [55.32]
Gyrotrons	*Bratman* et al. [55.33], *Idehara* et al. [55.34]
Free-electron lasers	*Knyazev* et al. [55.35], *Dem'yanenko* et al. [55.36]
Synchrotrons	*Byrd* et al. [55.37]
Solid-state electronic	
Gunn diodes	*Perez* et al. [55.38]
Transistors	*Lusakowski* et al. [55.39]
Frequency multiplication	*Maestrini* et al. [55.40]
Superconductor	*Ozyuzer* et al. [55.41]
Lasers	
Gas	*Dodel* [55.42]
Semiconductor	*Hübers* et al. [55.43], *Chassagneux* et al. [55.44]
Quantum cascade	*Köhler* et al. [55.45], *Williams* [55.46]
Sources pumped by lasers	
Continuous	
Photomixer	*McIntosh* et al. [55.47]
Mechanical resonance	*Wu* et al. [55.48]
Pulsed	
Photoconductive switches	*Auston* [55.49], *Shen* et al. [55.50]
Air	*Kim* et al. [55.51], *Kress* et al. [55.52]
Magnetic dipoles	*Beaurepaire* et al. [55.53]
Terahertz parametric oscillator	*Kawase* et al. [55.54]
Optical rectification	*Radhanpura* et al. [55.55]
Bulk	*Hargreaves* et al. [55.56], *Rowley* et al. [55.57]
Surface	*Reid* et al. [55.58], *Urbanowicz* et al. [55.59]
Transient currents	*Heyman* et al. [55.60], *Hargreaves* and *Lewis* [55.61]
Diffusion	*Dember* [55.62–64], *Klatt* et al. [55.65, 66]
Drift	*Lloyd-Hughes* et al. [55.67]
Mechanical excitation	
Peeling tape	*Horvat* and *Lewis* [55.68]
Surface formation	*Cortie* and *Lewis* [55.69]

silicon carbide forming a *globar* [55.23–25], and from the universe itself, via the cosmic background radiation [55.22]. Such sources can be modeled as black (or more accurately, as gray) bodies.

55.2.2 Vacuum Electronic

Charge carriers, typically electrons, accelerated freely, typically in a vacuum, emit terahertz-frequency electromagnetic radiation. Vacuum electronic sources range from relatively small units, such as backward-wave oscillators [55.28–30], to facilities that take up a whole room or building, such as free-electron lasers [55.35, 36] and synchrotrons [55.37]. These yield the highest power of all present terahertz sources, but are accompanied by limitations related to cost, size, and portability.

55.2.3 Solid-State Electronic

Electrons in solids, especially in semiconductors, also may be manipulated to produce terahertz radiation. Diodes [55.38], transistors [55.39], and superconducting elements [55.41] may be adapted for this purpose. Solid-state devices are also the basis of compact and robust terahertz lasers. These fall into two broad categories, bipolar [55.43, 44] and unipolar (the quantum cascade laser) [55.45, 46].

55.2.4 Laser-Pumped

Recent advances in terahertz sources have been associated with laser pumping. Two continuous laser sources of slightly different frequencies may be mixed in a suitable material to yield terahertz radiation at the frequency difference [55.47]. Continuous laser sources may also be used to excite mechanical resonances, which lead in turn to terahertz emission [55.48].

Much work has been done on pulsed laser sources, as these serve as the basis for terahertz time-domain spectroscopy. Pulses of near-infrared radiation can be downconverted to terahertz frequencies by such processes as photoconductive switching [55.49, 50], the ionization of air [55.52], or the reorientation of magnetic dipoles [55.53]. A widely used method is by optical rectification realized in nonlinear (electro-optic) crystals [55.55]. Both bulk [55.56, 57] and surface electric-field-induced [55.58, 59] effects contribute to the terahertz signal. A second widely used method is by the initiation of transient currents by photo-ionization of electron-hole pairs in semiconductors. The two principal mechanisms associated with the development of transient currents are charge-carrier drift [55.67] and diffusion [55.65, 66].

55.2.5 Mechanical Excitation

Terahertz radiation may also be produced by purely mechanical means, much as visible light is produced by friction (triboluminescence) or by the cleavage of surfaces (fractoluminescence). Two examples are charge separation on the peeling of adhesive tape leading to terahertz emission [55.68] and terahertz emission from rapidly-forming surfaces [55.69].

55.3 Terahertz Detectors

The mechanisms of terahertz detection in some sense reflect the mechanisms of terahertz emission: thermal, electronic, and electro-optic physical processes are common to both. A sampling of terahertz detector technologies is given in Table 55.4.

55.3.1 Thermal

Thermal detectors rely on the energy of incoming radiation raising their temperature and this temperature change leading to a change in another physical property. The physical property changed in a thermocouple is the electrical potential generated across it. Thermocouples are conventionally made of alloys of Bi and Sb [55.70, 82] but GaAs [55.83] and CMOS (CMOS) structures [55.84] are other materials employed. Cascading together a number of thermocouples results in a thermopile [55.71]. These also are often based on Bi [55.85] and its alloys [55.86] but other materials, such as Si/Ti, have been used [55.87]. Micromachining is now often utilized [55.88] and arrays of thermopiles are used for imaging [55.89]. Likewise, a voltage is set up across pyroelectric crystals, typically titanates or niobates, when heated [55.72] and these also are now micromachined [55.90] and employed for imaging [55.77] and even holography [55.91]. Rather than measure the change in electric potential on heating, the change in electrical resistance might be employed. This is the principle of operation of the bolometer [55.75, 92, 93]. Bolometers are conventionally made from Si, Ge or alloys such as SiGe on SiN [55.94, 95]. Superconductors, cooled to near the superconducting transition, are also used, but not widely [55.76, 96]. Bolometers are the most sensitive detectors of terahertz radiation.

Table 55.4 Some detectors of terahertz radiation

Detector type	Representative references
Thermal	
Thermocouple	*Roess* and *Dacus* [55.70]
Thermopile	*Moll* [55.71]
Pyroelectric	*Whatmore* [55.72]
Thermomechanical	*Alves* et al. [55.73, 74]
Bolometer	*Richards* [55.75]
Superconductor	*Gao* et al. [55.76]
Imaging arrays	*Yang* et al. [55.77]
Electro-optic	
Optical rectification	*Wu* et al. [55.78]
Electronic	
Direct electronic	*Song* et al. [55.79], *Pelling* et al. [55.80]
Photoconductive gating	*Seliuta* et al. [55.81]

55.3.2 Electro-Optic

Electro-optical detection relies on having a material whose birefringence is altered on being exposed to terahertz radiation [55.78]. The change in birefringence is then detected by a second light beam, typically a pulse from a Ti:sapphire laser, traversing the same material. This detection method may be described as upconversion; the terahertz response of the material is probed by a higher-frequency beam.

55.3.3 Electronic

Terahertz radiation may induce direct electronic transitions in suitable materials and the change in conductivity, or the change in the electric potential produced, may be used to monitor the intensity of the radiation [55.79, 80]. Alternatively, in an inversion of the photoconductive switch for terahertz emission, a photoconductive material may be used for terahertz detection, typically gated by a Ti:sapphire laser pulse [55.81].

55.4 Terahertz Components

Conventional materials utilized in the transfer of radiation of terahertz frequencies include windows and filters. Windows (of cryostats, for example) are typically made of inorganic crystals, including alkali halides, or plastics, such as polyethylene, polypropylene, or Mylar. Simple filters are based on size effects: for example, transmission filters incorporating diamond or other particles of suitable size, and reflection filters based on metal plates abraded with grits of appropriate grades. Likewise, wire grids, either free standing or supported by a semiconductor or plastic substrate, serve as filters with a cut-on frequency determined by the grid spacing. Other terahertz components include mirrors, lenses, polarizers, and metamaterials, a sampling of which is given in Table 55.5.

55.4.1 Mirrors

Metals are an obvious choice of material for terahertz mirrors. Al is commonly used; Ag has a higher conductivity and so a higher reflectivity; Au, while expensive, is less liable to tarnish. To save expense, and sometimes weight, an Ag or Au coating may be laid over a less expensive or lighter form. Recent innovations in terahertz mirror technology include semiconductor [55.97] and hybrid mirrors [55.98] and mirrors based on the concept of the photonic crystal [55.99].

Table 55.5 Some components for terahertz radiation

Component	Representative references
Mirrors	
Semiconductor	*Ye* et al. [55.97]
Hybrid	*Krumbholz* et al. [55.98]
Photonic crystal	*Tecimer* et al. [55.99]
Lenses	
Molded	*Scherger* et al. [55.100]
Variable focal length	*Scherger* et al. [55.101]
Fresnel zone plates	*Hristov* et al. [55.102]
Plasmonic resonance	*Jiang* et al. [55.103]
Paper	*Siemion* et al. [55.104]
Polarizers	
Reconfigurable	*Cheng* et al. [55.105]
Metamaterials	
Absorber	*Tao* et al. [55.106]
Modulator	*Chen* et al. [55.107]

55.4.2 Lenses

Terahertz lenses have conventionally been machined on a lathe from a suitable plastic such as polyethylene or TPX. More recently, molding has been employed, allowing for the rapid production of a large number of lenses [55.100]. Other innovations include Fresnel zone plates [55.102] and plasmonic resonances [55.103], variable focal length lenses [55.101], and even terahertz lenses made of paper [55.104].

55.4.3 Polarizers

Polarizers in the terahertz region are based on the same principle as in the visible, for example, by using a material that absorbs differently depending on the direction of the incident electric field. Given the greater wavelengths of terahertz radiation compared to visible radiation, wire-grid polarizers are simpler to realize in the terahertz than in the visible region. Recently, reconfigurable polarizers have been reported [55.105].

55.4.4 Metamaterials

Metamaterials are artificially engineered structures possessing properties not found in natural materials. Metamaterials allow the manipulation of terahertz radiation in various ways. A typical configuration is Au on GaAs forming split-ring resonators [55.108, 109]. Separate control of the electrical permittivity and the magnetic permeability has been achieved in terahertz-absorbing metamaterials [55.106] and terahertz modulators have also been demonstrated [55.107].

55.5 Conclusion

Most materials have properties in the terahertz spectral range significantly different to their properties in the lower-frequency (electronic) and higher-frequency (photonic) frequency ranges that bracket it. Terahertz materials thus form the basis for unique methods that underpin science and technology in this relatively under-explored frequency region. A current and growing area of materials research centers on sources and detectors of terahertz radiation, as well as the components that manipulate terahertz radiation, involving metals, semiconductors and crystalline and amorphous insulators, in a multitude of configurations.

Acknowledgments. K. Sakai is thanked for provision of the diagram on which Fig. 55.4 is based.

References

55.1 R.A. Lewis: Let's talk terahertz!, Am. J. Phys. **79**, 341 (2011)
55.2 S.M. Stewart: Terahertzing visible light, Am. J. Phys. **79**, 797 (2011)
55.3 K. Sakai: Private communication (2014)
55.4 K. Sakai: Pioneering studies and cutting-edge approaches in the terahertz-waves region, Proc. 34th Int. Conf. Infrared, Millimeter, and Terahertz Waves, IRMMW-THz 2009 (2009)
55.5 R.A. Lewis: A review of terahertz sources, J. Phys. D: Appl. Phys. **47**, 374001 (2014)
55.6 P.H. Siegel: Terahertz technology, IEEE Trans. Microw. Theory Tech. **50**, 910 (2002)
55.7 B. Ferguson, X.C. Zhang: Materials for terahertz science and technology, Nat. Mat. **1**, 26 (2002)
55.8 M. Tonouchi: Cutting-edge terahertz technology, Nat. Photonics **1**, 97 (2007)
55.9 M.C. Beard, G.M. Turner, C.A. Schmuttenmaer: Terahertz spectroscopy, J. Phys. Chem. B **106**, 7146 (2002)
55.10 W.L. Chan, J. Deibel, D.M. Mittleman: Imaging with terahertz radiation, Rep. Prog. Phys. **70**, 1325 (2007)
55.11 S. Wang, X.-C. Zhang: Pulsed terahertz tomography, J. Phys. D: Appl. Phys. **37**, R1 (2004)
55.12 E. Pickwell, V.P. Wallace: Biomedical applications of terahertz technology, J. Phys. D: Appl. Phys. **39**, R301 (2006)
55.13 H.-J. Song, T. Nagatsuma: Present and future of terahertz communications, IEEE Trans. Terahertz Sci. Technol. **1**, 256 (2011)
55.14 D. Mittleman (Ed.): *Sensing with Terahertz Radiation* (Springer, Berlin 2003)
55.15 K. Sakai (Ed.): *Terahertz Optoelectronics* (Springer, Berlin 2005)
55.16 S.D. Ganichev, W. Prettl: *Intense Terahertz Excitation of Semiconductors* (Oxford Science Publications, Oxford 2006)
55.17 S. Dexheimer (Ed.): *Terahertz Spectroscopy, Principles and Applications* (CRC Press, Boca Raton 2008)
55.18 Y.S. Lee: *Principles of Terahertz Science and Technology* (Springer, New York 2009)
55.19 X.C. Zhang, J. Xu: *Introduction to THz Wave Photonics* (Springer, New York 2010)
55.20 E. Brundermann, H.-W. Hübers, M.F. Kimmit: *Terahertz Techniques* (Springer, Heidelberg 2012)
55.21 R.A. Lewis: *Terahertz Physics* (Cambridge Univ. Press, Cambridge 2012)
55.22 C. Thacker, A. Cooray, J. Smidt, F. De Bernardis, K. Mitchell-Wynne, A. Amblard, R. Auld, M. Baes, D.L. Clements, A. Dariush, G. De Zotti, L. Dunne, S. Eales, R. Hopwood, C. Hoyos, E. Ibar, M. Jarvis, S. Maddox, M.J. Michałowski, E. Pascale, D. Scott, S. Serjeant, M.W.L. Smith, E. Valiante, P. van der Werf: H-ATLAS: The cosmic abundance of dust from the far-infrared background power spectrum, Astrophys. J. **768**, 58 (2013)
55.23 V.M. Zolotarev, R.K. Mamedov, A.N. Bekhterev, B.Z. Volchek: Spectral emissivity of a globar lamp in the 2–50-μm region, J. Opt. Technol. **74**, 378 (2007)
55.24 R.A. Friedel, A.G. Sharkey Jr.: Comparison of glower and Globar sources for infra-red spectrometry, Rev. Sci. Instrum. **18**, 928 (1947)
55.25 R.A. Friedel, A.G. Sharkey Jr.: Erratum: Comparison of glower and Globar sources for infra-red spectrometry, Rev. Sci. Instrum. **19**, 180 (1948)
55.26 K. Charrada, G. Zissis, M. Aubes: Two-temperature, two-dimensional fluid modelling of mercury plasma in high-pressure lamps, J. Phys. D: Appl. Phys. **29**, 2432 (1996)
55.27 J.H. Booske, R.J. Dobbs, C.D. Joye, C.L. Kory, G.R. Neil, G.S. Park, J. Park, R.J. Temkin: Vacuum electronic high power terahertz sources, IEEE Trans. Terahertz Sci. Technol. **1**, 54 (2011)
55.28 M. Mineo, C. Paoloni: Corrugated rectangular waveguide tunable backward wave oscillator for terahertz applications, IEEE Trans. Electron Devices **57**, 1481 (2010)
55.29 X. Xu, Y. Wei, F. Shen, H. Yin, J. Xu, Y. Gong, W. Wang: A watt-class 1-THz backward-wave oscillator based on sine waveguide, Phys. Plasmas **19**, 013113 (2012)
55.30 W. He, C.R. Donaldson, L. Zhang, K. Ronald, P. McElhinney, A.W. Cross: High power wideband gyrotron backward wave oscillator operating towards the terahertz region, Phys. Rev. Lett. **110**, 165101 (2013)
55.31 Y.M. Shin, G.S. Park, G.P. Scheitrum, G. Caryotakis: Circuit analysis of an extended interaction

klystron, J. Korean Phys. Soc. **44**, 1239 (2004)
55.32 S. Bhattacharjee, J.H. Booske, C.L. Kory, D.W. van der Weide, S. Limbach, S. Gallagher, J.D. Welter, M.R. Lopez, R.M. Gilgenbach, R.L. Ives, M.E. Read, R. Divan, D.C. Mancini: Folded waveguide traveling-wave tube sources for terahertz radiation, IEEE Trans. Plasma Sci. **32**, 1002 (2004)
55.33 V.L. Bratman, Y.K. Kalynov, V.N. Manuilov: Large-orbit gyrotron operation in the terahertz frequency range, Phys. Rev. Lett. **102**, 245101 (2009)
55.34 T. Idehara, T. Saito, I. Ogawa, S. Mitsudo, Y. Tatematsu, L. Agasu, H. Mori, S. Kobayashi: Development of terahertz FU CW Gyrotron series for DNP, Appl. Magn. Reson. **34**, 265 (2008)
55.35 B.A. Knyazev, G.N. Kulipanov, N.A. Vinokurov: Novosibirsk terahertz free electron laser: Instrumentation development and experimental achievements, Meas. Sci. Technol. **21**, 054017 (2010)
55.36 M.A. Dem'yanenko, D.G. Esaev, B.A. Knyazev, G.N. Kulipanov, N.A. Vinokurov: Imaging with a 90 frames/s microbolometer focal plane array and high-power terahertz free electron laser, Appl. Phys. Lett. **92**, 131116 (2008)
55.37 J.M. Byrd, W.P. Leemans, A. Loftsdottir, B. Marcelis, M.C. Martin, W.R. McKinney, F. Sannibale, T. Scarvie, C. Steier: Observation of broadband self-amplified spontaneous coherent terahertz synchrotron radiation in a storage ring, Phys. Rev. Lett. **89**, 224801 (2002)
55.38 S. Pérez, T. González, D. Pardo, J. Mateos: Terahertz Gunn-like oscillations in InGaAs/InAlAs planar diodes, J. Appl. Phys. **103**, 094516 (2008)
55.39 J. Lusakowski, W. Knap, N. Dyakonova, L. Varani, J. Mateos, T. Gonzalez, Y. Roelens, S. Bollaert, A. Cappy, K. Karpierz: Voltage tuneable terahertz emission from a ballistic nanometer InGaAs/InAlAs transistor, J. Appl. Phys. **97**, 064307 (2005)
55.40 A. Maestrini, J.S. Ward, J.J. Gill, C. Lee, B. Thomas, R.H. Lin, G. Chattopadhyay, I. Mehdi: A frequency-multiplied source with more than 1 mW of power across the 840–900-GHz band, IEEE Trans. Microw. Theory Tech. **58**, 1925 (2010)
55.41 L. Ozyuzer, A.E. Koshelev, C. Kurter, N. Gopalsami, Q. Li, M. Tachiki, K. Kadowaki, T. Yamamoto, H. Minami, H. Yamaguchi, T. Tachiki, K.E. Gray, W.-K. Kwok, U. Welp: Emission of coherent THz radiation from superconductors, Science **318**, 1291 (2007)
55.42 G. Dodel: On the history of far-infrared (FIR) gas lasers: Thirty-five years of research and application, Infrared Phys. Technol. **40**, 127 (1999)
55.43 H.W. Hübers, S.G. Pavlov, V.N. Shastin: Terahertz lasers based on germanium and silicon, Semicond. Sci. Technol. **20**, S211 (2005)
55.44 Y. Chassagneux, R. Colombelli, W. Maineult, S. Barbieri, H.E. Beere, D.A. Ritchie, S.P. Khanna, E.H. Linfield, A.G. Davies: Electrically pumped photonic-crystal terahertz lasers controlled by boundary conditions, Nature **457**, 174 (2009)
55.45 R. Köhler, A. Tredicucci, F. Beltram, H.E. Beere, E.H. Linfield, A.G. Davies, D.A. Ritchie, R.C. Iotti, F. Rossi: Terahertz semiconductor-heterostructure laser, Nature **417**, 156 (2002)
55.46 B.S. Williams: Terahertz quantum-cascade lasers, Nat. Photonics **1**, 517 (2007)
55.47 K.A. McIntosh, E.R. Brown, K.B. Nichols, O.B. McMahon, W.F. DiNatale, T.M. Lyszczarz: Terahertz photomixing with diode lasers in low-temperature-grown GaAs, Appl. Phys. Lett. **67**, 3844 (1995)
55.48 X.L. Wu, S.J. Xiong, Z. Liu, J. Chen, J.C. Shen, T.H. Li, P.H. Wu, P.K. Chu: Green light stimulates terahertz emission from mesocrystal microspheres, Nat. Nanotechnol. **6**, 103 (2011)
55.49 D.H. Auston: Picosecond optoelectronic switching and gating in silicon, Appl. Phys. Lett. **26**, 101 (1975)
55.50 Y.C. Shen, P.C. Upadhya, E.H. Linfield, H.E. Beere, A.G. Davies: Ultrabroadband terahertz radiation from low-temperature-grown GaAs photoconductive emitters, Appl. Phys. Lett. **83**, 3117 (2003)
55.51 K.Y. Kim, J.H. Glownia, A.J. Taylor, G. Rodriguez: Coherent control of terahertz supercontinuum generation in ultrafast laser-gas interactions, Nat. Photon. **2**, 605 (2008)
55.52 M. Kress, T. Löffler, S. Eden, M. Thomson, H.G. Roskos: Terahertz-pulse generation by photoionization of air with laser pulses composed of both fundamental and second-harmonic waves, Opt. Lett. **29**, 1120 (2004)
55.53 E. Beaurepaire, G.M. Turner, S.M. Harrel, M.C. Beard, J.Y. Bigot, C.A. Schmuttenmaer: Coherent terahertz emission from ferromagnetic films excited by femtosecond laser pulses, Appl. Phys. Lett. **84**, 3465 (2004)
55.54 K. Kawase, M. Sato, T. Taniuchi, H. Ito: Coherent tunable THz-wave generation from $LiNbO_3$ with monolithic grating coupler, Appl. Phys. Lett. **68**, 2483 (1996)
55.55 K. Radhanpura, S. Hargreaves, R.A. Lewis, L. Sirbu, I.M. Tiginyanu: Heavy noble gas (Kr, Xe) irradiated (111) InP nanoporous honeycomb membranes with enhanced ultrafast all-optical terahertz emission, Appl. Phys. Lett. **97**, 181921 (2010)
55.56 S. Hargreaves, K. Radhanpura, R.A. Lewis: Generation of terahertz radiation by bulk and surface optical rectification from crystal planes of arbitrary orientation, Phys. Rev. B **80**, 195323 (2009)
55.57 J.D. Rowley, J.K. Wahlstrand, K.T. Zawilski, P.G. Schunemann, N.C. Giles, A.D. Bristow: Terahertz generation by optical rectification in uniaxial birefringent crystals, Opt. Express **20**, 16968 (2012)
55.58 M. Reid, I.V. Cravetchi, R. Fedosejevs: Terahertz radiation and second-harmonic generation from InAs: Bulk versus surface electric-field-induced contributions, Phys. Rev. B **72**, 035201 (2005)
55.59 A. Urbanowicz, A. Krotkus, R. Adomavičius, V.L. Malevich: Terahertz emission from femtosecond laser excited Ge surfaces due to the electrical field-induced optical rectification, Physica B:

Cond. Mat. **398**, 98 (2007)
55.60 J.N. Heyman, N. Coates, A. Reinhardt, G. Strasser: Diffusion and drift in terahertz emission at GaAs surfaces, Appl. Phys. Lett. **83**, 5476 (2003)
55.61 S. Hargreaves, R.A. Lewis: Single-cycle azimuthal angle dependence of terahertz radiation from (100) *n*-type InP, Appl. Phys. Lett. **93**, 242101 (2008)
55.62 H. Dember: A photoelectrical-motive energy in copper-oxide crystals, Phys. Z. **32**, 554 (1931)
55.63 H. Dember: A crystal photocell, Phys. Z. **32**, 856 (1931)
55.64 H. Dember: Forward motion of electrons induced by light, Phys. Z. **32**, 207 (1931)
55.65 G. Klatt, F. Hilser, W. Qiao, M. Beck, R. Gebs, A. Bartels, K. Huska, U. Lemmer, G. Bastian, M.B. Johnston, M. Fischer, J. Faist, T. Dekorsy: Terahertz emission from lateral photo-Dember currents, Opt. Express **18**, 4939 (2010)
55.66 G. Klatt, B. Surrer, D. Stephan, O. Schubert, M. Fischer, J. Faist, A. Leitenstorfer, R. Huber, T. Dekorsy: Photo-Dember terahertz emitter excited with an Er:Fiber laser, Appl. Phys. Lett. **98**, 021114 (2011)
55.67 J. Lloyd-Hughes, S.K.E. Merchant, L. Fu, H.H. Tan, C. Jagadish, E. Castro-Camus, M.B. Johnston: Influence of surface passivation on ultrafast carrier dynamics and terahertz radiation generation in GaAs, Appl. Phys. Lett. **89**, 232102 (2006)
55.68 J. Horvat, R.A. Lewis: Peeling adhesive tape emits electromagnetic radiation at terahertz frequencies, Opt. Lett. **34**, 2195 (2009)
55.69 D.L. Cortie, R.A. Lewis: Terahertz surfoluminescence, Surf. Sci. **606**, 1573 (2012)
55.70 L.C. Roess, E.N. Dacus: The design and construction of rapid-response thermocouples for use as radiation detectors in infra-red spectrographs, Rev. Sci. Instrum. **16**, 164 (1945)
55.71 W.J.H. Moll: A thermopile for measuring radiation, Proc. Phys. Soc. (London) **35**, 257 (1922)
55.72 R.W. Whatmore: Pyroelectric devices and materials, Rep. Prog. Phys. **49**, 1335 (1986)
55.73 F. Alves, D. Grbovic, B. Kearney, N.V. Lavrik, G. Karunasiri: Bi-material terahertz sensors using metamaterial structures, Opt. Express **21**, 13256 (2013)
55.74 F. Alves, D. Grbovic, B. Kearney, G. Karunasiri: Microelectromechanical systems bimaterial terahertz sensor with integrated metamaterial absorber, Opt. Lett. **37**, 1886 (2012)
55.75 P.L. Richards: Bolometers for infrared and millimeter waves, J. Appl. Phys. **76**, 1 (1994)
55.76 J.R. Gao, J.N. Hovenier, Z.Q. Yang, J.J.A. Baselmans, A. Baryshev, M. Hajenius, T.M. Klapwijk, A.J.L. Adam, T.O. Klaassen, B.S. Williams, S. Kumar, Q. Hu, J.L. Reno: Terahertz heterodyne receiver based on a quantum cascade laser and a superconducting bolometer, Appl. Phys. Lett. **86**, 1 (2005)
55.77 J. Yang, S. Ruan, M. Zhang: Real-time, continuous-wave terahertz imaging by a pyroelectric camera, Chin. Opt. Lett. **6**, 29 (2008)
55.78 Q. Wu, X.-C. Zhang: Design and characterization of traveling-wave electrooptic terahertz sensors, IEEE J. Sel. Top. Quantum Electron. **2**, 693 (1996)
55.79 J. Song, G. Aizin, Y. Kawano, K. Ishibashi, N. Aoki, Y. Ochiai, J.L. Reno, J.P. Bird: Evaluating the performance of quantum point contacts as nanoscale terahertz sensors, Opt. Express **18**, 4609 (2010)
55.80 S. Pelling, R. Davis, L. Kulik, A. Tzalenchuk, S. Kubatkin, T. Ueda, S. Komiyama, V.N. Antonov: Point contact readout for a quantum dot terahertz sensor, Appl. Phys. Lett. **93**, 073501 (2008)
55.81 D. Seliuta, I. Kašalynas, V. Tamošiūnas, S. Balakauskas, Z. Martunas, S. Ašmontas, G. Valušis, A. Lisauskas, H.G. Roskos, K. Köhler: Silicon lens-coupled bow-tie InGaAs-based broadband terahertz sensor operating at room temperature, Electron. Lett. **42**, 825 (2006)
55.82 A.K. Huhn, G. Spickermann, A. Ihring, U. Schinkel, H.-G. Meyer, P. Haring Bolívar: Uncooled antenna-coupled terahertz detectors with 22 μs response time based on BiSb/Sb thermocouples, Appl. Phys. Lett. **102**, 121102 (2013)
55.83 N. Chong, T.A.S. Srinivas, H. Ahmed: Performance of GaAs microbridge thermocouple infrared detectors, J. Microelectromech. Sys. **6**, 136 (1997)
55.84 V. Milanovic, M. Gaitan, M.E. Zaghloul: Micromachined thermocouple microwave detector by commercial CMOS fabrication, IEEE Trans. Microw. Theory Tech. **46**, 550 (1998)
55.85 W.W. Coblentz: A thermopile of bismuth alloy, Phys. Rev. **3**, 59 (1914)
55.86 F. Völklein, A. Wiegand, V. Baier: High-sensitivity radiation thermopiles made of BiSbTe films, Sens. Act. A **29**, 87 (1991)
55.87 S. Ben Mbarek, S. Euphrasie, T. Baron, L. Thiery, P. Vairac, D. Briand, J.-P. Guillet, L. Chusseau: Room temperature Si-Ti thermopile THz sensor, Microsyst. Technol. **21**, 1627 (2014)
55.88 A. Graf, M. Arndt, M. Sauer, G. Gerlach: Review of micromachined thermopiles for infrared detection, Meas. Sci. Technol. **18**, R59 (2007)
55.89 A.D. Oliver, K.D. Wise: 1024-Element bulk-micromachined thermopile infrared imaging array, Sens. Actuators A: Phys. **73**, 222 (1999)
55.90 P. Muralt: Micromachined infrared detectors based on pyroelectric thin films, Rep. Prog. Phys. **64**, 1339 (2001)
55.91 S.-H. Ding, Q. Li, Y.-D. Li, Q. Wang: Continuous-wave terahertz digital holography by use of a pyroelectric array camera, Opt. Lett. **36**, 1993 (2011)
55.92 G.K.T. Conn: A thermocouple-bolometer detector, Trans. Faraday Soc. **41**, 192 (1945)
55.93 J.C. Mather: Bolometers: Ultimate sensitivity, optimization, and amplifier coupling, Appl. Opt. **23**, 584 (1984)
55.94 S. Sedky, P. Fiorini, K. Baert, L. Hermans, R. Mertens: Characterization and optimization of infrared poly SiGe bolometers, IEEE Trans. Electron Devices **46**, 675 (1999)
55.95 P.D. Mauskopf, J.J. Bock, H. Del Castillo, W.L. Holzapfel, A.E. Lange: Composite infrared bolometers with Si_3N_4 micromesh absorbers, Appl. Opt. **36**, 765 (1997)

55.96 S.-F. Lee, J.M. Gildemeister, W. Holmes, A.T. Lee, P.L. Richards: Voltage-biased superconducting transition-edge bolometer with strong electrothermal feedback operated at 370 mK, Appl. Opt. **37**, 3391 (1998)

55.97 H.B. Ye, Y.H. Zhang, W.Z. Shen: Carrier transport and optical properties in GaAs far-infrared/terahertz mirror structures, Thin Solid Films **514**, 310 (2006)

55.98 N. Krumbholz, K. Gerlach, F. Rutz, M. Koch, R. Piesiewicz, T. Kürner, D. Mittleman: Omnidirectional terahertz mirrors: A key element for future terahertz communication systems, Appl. Phys. Lett. **88**, 202905 (2006)

55.99 M. Tecimer, K. Holldack, L.R. Elias: Dynamically tunable mirrors for THz free electron laser applications, Phys. Rev. ST Accel. Beams **13**, 030703 (2010)

55.100 B. Scherger, M. Scheller, C. Jansen, M. Koch, K. Wiesauer: Terahertz lenses made by compression molding of micropowders, Appl. Opt. **50**, 2256 (2011)

55.101 B. Scherger, C. Jördens, M. Koch: Variable-focus terahertz lens, Opt. Express **19**, 4528 (2011)

55.102 H.D. Hristov, J.M. Rodriguez, W. Grote: The grooved-dielectric Fresnel zone plate: An effective terahertz lens and antenna, Microw. Opt. Technol. Lett. **54**, 1343 (2012)

55.103 X.-Y. Jiang, J.-S. Ye, J.-W. He, X.-K. Wang, D. Hu, S.-F. Feng, Q. Kan, Y. Zhang: An ultrathin terahertz lens with axial long focal depth based on metasurfaces, Opt. Express **21**, 30030 (2013)

55.104 A. Siemion, A. Siemion, M. Makowski, J. Suszek, J. Bomba, A. Czerwiński, F. Garet, J.-L. Coutaz, M. Sypek: Diffractive paper lens for terahertz optics, Opt. Lett. **37**, 4320 (2012)

55.105 L.J. Cheng, L. Liu: Optical modulation of continuous terahertz waves towards cost-effective reconfigurable quasi-optical terahertz components, Opt. Express **21**, 28657 (2013)

55.106 H. Tao, N.I. Landy, C.M. Bingham, X. Zhang, R.D. Averitt, W.J. Padilla: A metamaterial absorber for the terahertz regime: Design, fabrication and characterization, Opt. Express **16**, 7181 (2008)

55.107 H.-T. Chen, W.J. Padilla, M.J. Cich, A.K. Azad, R.D. Averitt, A.J. Taylor: A metamaterial solid-state terahertz phase modulator, Nat. Photonics **3**, 148 (2009)

55.108 H.-T. Chen, W.J. Padilla, J.M.O. Zide, A.C. Gossard, A.J. Taylor, R.D. Averitt: Active terahertz metamaterial devices, Nature **444**, 597 (2006)

55.109 W.J. Padilla, A.J. Taylor, C. Highstrete, M. Lee, R.D. Averitt: Dynamical electric and magnetic metamaterial response at terahertz frequencies, Phys. Rev. Lett. **96**, 107401 (2006)

56. Metamaterials

Vassili Fedotov

Metamaterials is a new field of interdisciplinary research, which deals with artificial material composites engineered to display physical properties that surpass (or complement) those available in nature. Originally proposed as a way of tailoring media's electromagnetic and optical properties, the metamaterial concept has recently extended its reach to also include elastic, acoustic, and thermal properties. Since 1999, with the emergence of the field, metamaterials has attracted a lot of attention from the scientific community all over the globe, offering immediate applications in antenna and waveguide engineering, imaging, microscopy, sensing, and light manipulation. Metamaterial research has brought together electrical engineers, material and optical scientists, chemists, and mathematicians; it has also advanced our understanding of electrodynamics, pushed the boundaries of nanofabrication, and stimulated the development of novel characterization techniques. Metamaterial research continues to expand quite rapidly, making it almost impossible to faithfully cover the field even with a reasonably sized book. This chapter attempts to give a brief overview of the history of electromagnetic (photonic) metamaterials, important developments, and main concepts in the field. Also, it aims to provide a basic understanding of the principles, design rules, and tricks routinely exploited by the researchers to achieve the desired material behavior. The chapter is divided into two parts, which are devoted to bulk metamaterials and their two-dimensional counterparts, the so-called planar metamaterials or metasurfaces.

56.1 **Bulk Metamaterials** 1351
56.1.1 Basic Principle of Metamaterial Operation 1352
56.1.2 The First Metamaterial 1353
56.1.3 Magnetic Metamaterials (Artificial Magnetism) 1353
56.1.4 Negative-Index Metamaterials (Left-Handed Media) 1355
56.1.5 Chiral Metamaterials 1358
56.1.6 Transformation Optics and Metamaterial Cloaks 1360
56.1.7 Hyperbolic Metamaterials 1362

56.2 **Planar Metamaterials** 1365
56.2.1 Frequency Selective Surfaces 1365
56.2.2 Magnetic Mirrors (High Impedance Surfaces) 1367
56.2.3 Perfect Metamaterial Absorbers 1368
56.2.4 Designer Metasurfaces and Flat Optics 1369
56.2.5 Polarization Manipulation with Planar Metamaterials 1370

References 1375

56.1 Bulk Metamaterials

Metamaterial is now a well-accepted term for a novel type of optical medium that is artificially engineered to interact with light in ways no natural materials can. Metamaterials gain their unusual properties from the particular design of their elementary building blocks – metamolecules (or meta-atoms), which are built on a subwavelength scale from common constituent materials such as metals and dielectrics. Because of their subwavelength structuring, metamaterials do not diffract light and hence are quite different from photonic crystals, acting as uniform media that can, in principle, be ascribed an effective refractive index.

Interest in the metamaterial research exploded in the year 2001 when scientists experimentally demonstrated a truly exotic effect of negative refraction, a phenomenon in which light rays are refracted or bent in the direction opposite to that which is normally expected [56.1]. Negative refraction became one of the signature effects of the metamaterials physics and its demonstration opened a new exciting chapter in material and electromagnetic research. While subsequent

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_56

scientific activity has mainly centered on electromagnetic properties of the metamaterials (most of which will be considered in this chapter), it would be fair to note that one of the first metamaterial definitions given in 2001 by *Valerie Browning* (DARPA, USA) encompassed many more material properties [56.2, 3]:

> *Metamaterials are a new class of ordered composites that exhibit exceptional properties not readily observed in nature. These properties arise from qualitatively new response functions that are: (1) not observed in the constituent materials; (2) result from the inclusion of artificially fabricated, extrinsic, low dimensional inhomogeneities.*

56.1.1 Basic Principle of Metamaterial Operation

The exotic and often dramatic physics exhibited by the metamaterials is in many cases underpinned by the resonant response of their metamolecules. A generic example of the metamolecule, which forms a three-dimensional metamaterial crystal when translated periodically along the x, y, and z-axes, is schematically shown in Fig. 56.1. It typically consists of metallic wires (or patches) embedded in or supported by a dielectric substrate/matrix. An electromagnetic wave incident onto the metamaterial induces oscillating electrical currents $\boldsymbol{j}$ in the conducting elements of each metamolecule, which then scatter (i. e., re-radiate) secondary electromagnetic waves forming the macroscopic response of the metamaterial. Radiation scattered backward is perceived as the metamaterial reflection, while radiation scattered forward contributes to the transmission by adding coherently to the incident wave.

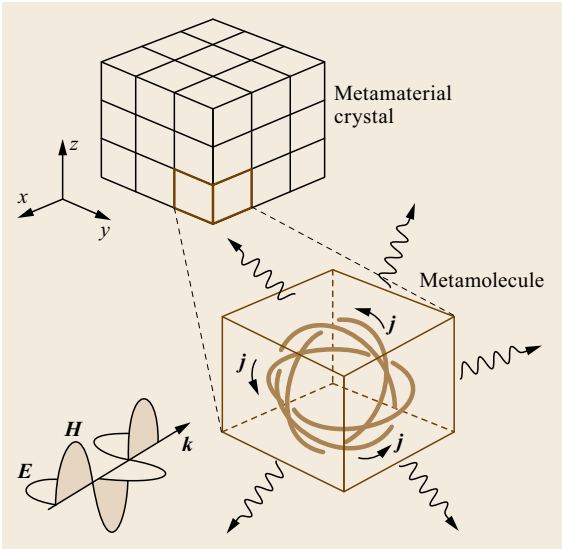


Fig. 56.1 Bulk metamaterial crystal and its elementary building block – the metamolecule (metallic wire resonators embedded into a dielectric cube)

Each metamolecule effectively acts as a very compact antenna that for a certain design and dimensions of its metallic pattern becomes resonantly coupled to the fields of the incident wave, featuring substantially amplified current oscillations and, as a result, strong electromagnetic scattering. The metamolecule can also be considered as a classical analog of an atom, where the frequency of the metamolecule's resonance ω_0 corresponds to the energy of the atom's excited state $E^* = \hbar\omega_0$, while scattering of electromagnetic waves by the induced currents corresponds to the emission of a photon $\hbar\omega_0$ by the atom upon its transition from the excited to ground state.

By tailoring the design of the metamolecules, one can selectively engineer an enhanced effective material response at virtually any wavelength (frequency) and of any origin – be it electric or magnetic, or a combination of the two (yielding optical activity or negative refraction), or even more exotic – toroidal [56.4]. Some examples of the metallic patterns commonly exploited in the metamaterial research are shown in Fig. 56.2, and some of these will be considered in detail below.

The resonant response of a metamolecule is determined by the inductance L and capacitance C of its metallic pattern, which may be viewed as an electrical circuit driven externally by the fields of the incident wave. When the overall size of the metamolecule is smaller than $\approx \lambda_0/20$, such a circuit can be described as a network of lumped electrical components with the resonance frequency given by the well-known expression $\omega_0 = 2\pi f_0 = 1/\sqrt{LC}$. The inductance and capacitance of the pattern in this case must be substantially larger than the self-inductance and self-capacitance of the constituting wire(s), although at present this can be achieved only for some radio and microwave-frequency designs [56.5, 6] due to limitations of existing micro and nanofabrication technologies. In practice, the typical size of metamolecules ranges from about $\lambda_0/6$ to $\lambda_0/2$, which one may slightly reduce (by a factor of 1.5–2) with the help of high-index dielectric substrates. Consequently, the metamolecules do not operate in the *quasi-static* regime and should be regarded as miniature transmission lines with distributed inductance and capacitance. Their resonant behavior results from the excitation of standing waves of electrical currents, which are analogs to the resonant modes of an electromagnetic cavity resonator or a vibrating string (Fig. 56.3). Although an accurate quantitative characterization of the induced current oscillations is cumbersome in this case

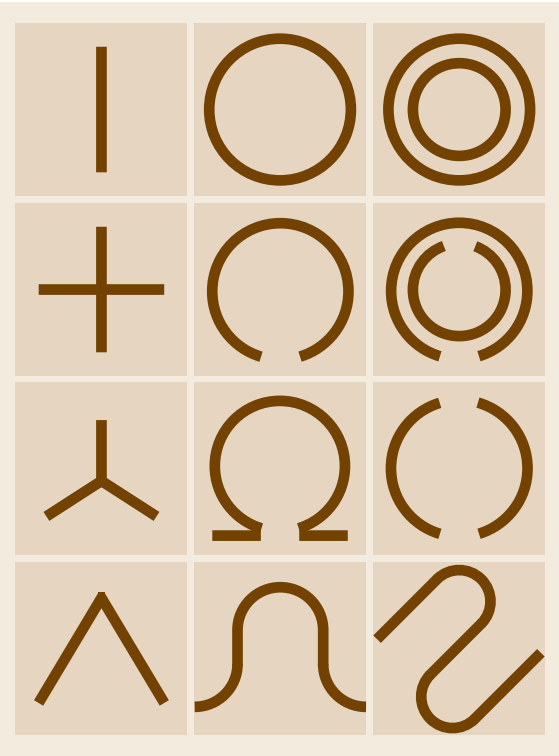


Fig. 56.2 Examples of metamaterial patterns

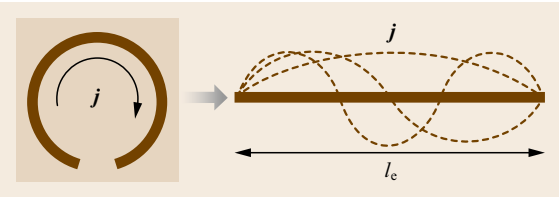


Fig. 56.3 Geometrical resonances of a metamolecule. Conduction currents $\boldsymbol{j}$ oscillating in the metamolecule form standing waves along the length of its metallic pattern (l_e)

and requires the use of the transmission line theory or numerical modeling, the wavelengths of the supported resonances can be estimated using very simple relations,

$$\lambda_0 \approx \sqrt{\varepsilon}\frac{2l_e}{m}\,, \tag{56.1}$$

where l_e is the total electrical length of the metallic pattern that forms the metamolecule, ε is the relative permittivity of the dielectric medium that the pattern is embedded in, and m is a positive integer. Resonances with $m > 4$ are usually not possible to engineer in subwavelength metamolecules (since $l_e > \lambda_0$).

Importantly, by reducing the size of the metamolecule (and hence l_e) such *geometrical* resonances can be moved to shorter wavelengths (higher frequencies) – all the way into the near-IR, where metals are still good conductors and can, therefore, support electrical currents. At visible frequencies, however, the currents are replaced with plasmons, and the linear relation between λ_0 and l_e becomes less accurate and eventually saturates due to the kinetic inductance of the electron plasma [56.7].

56.1.2 The First Metamaterial

Perhaps the first ever metamaterial was constructed and studied by *Bose* more than 100 years ago. In 1898, he wrote [56.8]:

> *In order to imitate the* [optical] *rotation by liquids like sugar solutions, I made elements of molecules of twisted jute, of two varieties, one kind being twisted to the right (positive) and the other twisted to the left (negative) . . . .*

In his experiments, Bose placed the twisted jutes between a spark generator equipped with a polarizer (which acted as a source of linearly polarized millimeter waves) and a detector, and determined the orientation of the polarization plane of the transmitted waves by rotating an analyzer in front of the detector (Fig. 56.4). He confirmed that *the twisted structure* [of the jute] *produces an optical twist of the plane of polarization* [56.8]. It is important to note here that although Bose's structure served merely to reproduce the effect of polarization rotation (i. e., optical activity) encountered in the optical part of the spectrum in sugar solutions and other media lacking reflection symmetry, it did so at millimeter-wave frequencies, where no natural materials are known to exhibit such an effect.

Since then more elaborate designs of optically active metamaterials have been proposed and realized for various operating frequencies, including the optical part of the spectrum, where they easily outperformed the naturally available media in terms of their rotary power (Sect. 56.1.5).

56.1.3 Magnetic Metamaterials (Artificial Magnetism)

Arguably, the most important development in the field of metamaterial research was the demonstration of so-called artificial magnetism, a phenomenon characterized by the appearance of a strong magnetic response (most notably at optical frequencies) in inherently nonmagnetic material composites. Indeed, as will be shown in the following sections, artificial magnetism is the key ingredient of the exotic negative-index response and

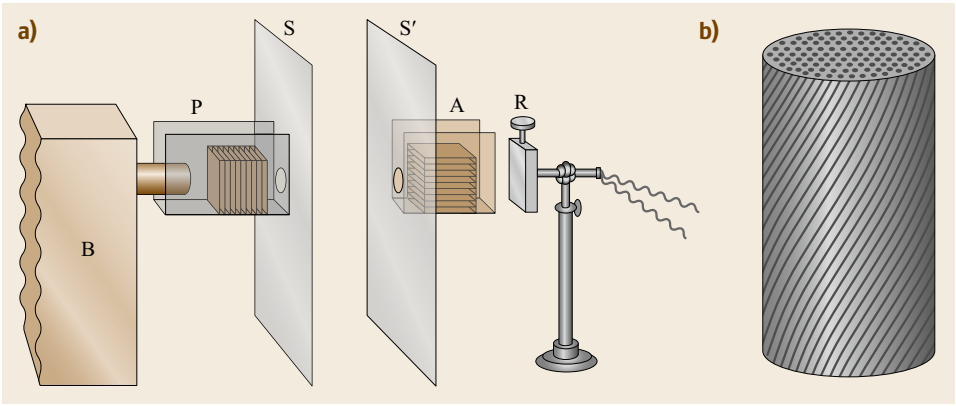


Fig. 56.4a,b Mimicking optical activity with artificially engineered structures. **(a)** Diagram of Bose's experiment: B – source of electromagnetic waves, P – polarizer, S and S′ – screens with apertures, A – analyzer, R – spiral spring receiver. **(b)** Twisted jute, the fist chiral metamolecule. (After [56.8])

a material property essential for the design of three-dimensional electromagnetic cloaks.

The idea of artificial magnetism was conceived in 1999 by *Pendry* et al., who proposed a magnetic analog of an electric conductor with controllable electromagnetic properties [56.9]. Materials exhibiting artificial magnetism were also suggested as a practical solution for manipulating the secondary radio frequency (RF) magnetic fields in an magnetic resonance imaging (MRI) environment, where the use of conventional magnetic materials would perturb the pattern of the primary quasi-static magnetic field [56.5]. Most of all, the interest in artificial magnetism was driven by the fact that strong natural magnetism (exhibited by ferromagnetic materials) is virtually nonexistent at terahertz frequencies and above, where light–matter interactions are mediated exclusively by the electrical component of light (magnetic permeability $\mu \approx 1$).

The existence of strong magnetism in nature is ensured by unpaired electron spins in atoms, which appear to be locally aligned through exchange interaction with one another, forming microscopic domains of uniform spontaneous magnetization (ferromagnetism). At the macroscopic level, the magnetic response results from the re-orientation of all such domains parallel to an externally applied magnetic field each time the field reverses its sign. The re-orientation, however, typically takes a fraction of a microsecond to complete, which means that at terahertz frequencies and above the variations of the applied field are too quick for the domains to follow, and so the ferromagnetic response vanishes. The only mechanism that could yield natural magnetism at optical frequencies is diamagnetism. Classically speaking, diamagnetism is the manifestation of Lenz's law on the atomic scale: when an external magnetic field couples to orbiting electrons, it forces an atom to precess (Larmor precession) and thus induces additional orbital currents, which generate a magnetic field that opposes the external one. Unfortunately, diamagnetic response in natural materials is α^2 times weaker than the electric one (where α is the fine-splitting constant), which is why the interaction of matter with light occurs exclusively via the electric field [56.10].

The metamaterial approach allows for the mimicking of orbital currents with loops of conduction currents that are magnetically induced in metamolecules via Faraday induction (Fig. 56.5a). Although metamolecules are much larger than natural atoms and molecules, they are still smaller than the wavelength of light and thus on the macroscopic scale (i. e., where the effective material response is formed) the above

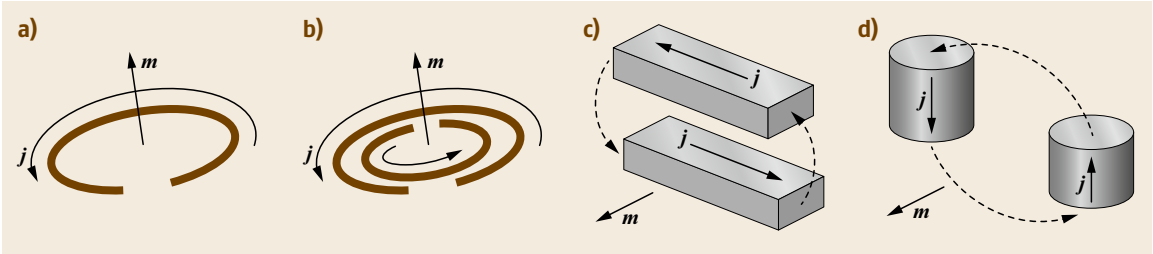


Fig. 56.5a–d Magnetic metamolecules. Orbital currents in atoms and molecules are imitated with loops of electrical currents $\boldsymbol{j}$ induced in **(a)** a split wire ring, **(b)** a double split wire ring, **(c)** a pair of metallic bars, and **(d)** a pair of metallic pillars. Current loops give rise to microscopic magnetic dipole moments $\boldsymbol{m}$

trick will remain unnoticed. Since each metamolecule operates as an *LC*-circuit, the induced conduction currents will be resonantly amplified, yielding artificial magnetic response that is several orders of magnitude stronger than found in natural materials. Note that the sign of the effective magnetic permeability μ^* can be both negative and positive, depending on the detuning from the resonance frequency f_0.

One of the most popular examples of the metamolecule that can exhibit artificial magnetism is a metallic ring with a split (Fig. 56.5a), commonly referred to as a split-ring resonator (SRR). Along with its derivative – a pair of concentric split-rings (Fig. 56.5b), also known as a double split-ring resonator (DSSR), it has been successfully exploited for demonstrating artificial magnetism experimentally within a wide spectral range, spanning from microwave to IR frequencies (see [56.10] and references therein). At near-IR and visible frequencies, however, the split-ring design was rather challenging to implement given the limitations of the fabrication techniques; it was also less effective due to large kinetic inductance of plasmons that replace conduction currents at those frequencies. Instead, magnetic metamolecules were engineered as pairs of parallel metallic bars or pillars (Fig. 56.5c,d). Current loops in those structures were formed by plasmons in the metal and displacement currents in the gaps between the ends of the bars/pillars [56.11, 12].

56.1.4 Negative-Index Metamaterials (Left-Handed Media)

The negative index of refraction ($n < 0$) signifies a peculiar regime of a medium's macroscopic electromagnetic response, when the electric permittivity ε and the magnetic permeability μ become simultaneously negative. Mathematically, negative n is given by the secondary value of the square-root function, which must be chosen to respect the negative signs of ε and μ in such a case

$$n = \sqrt{\varepsilon\mu} = \sqrt{(-1)^2|\varepsilon||\mu|} = -\sqrt{|\varepsilon||\mu|}\,. \qquad (56.2)$$

The negative sign of n is also consistent with the laws of physics and has a clear physical meaning: it indicates that the phase velocity of a wave is antiparallel to its group velocity. This was shown by several scientists in the twentieth century [56.13–15], and in particular, by the Soviet scientist *Veselago* in 1967, who investigated theoretically the electrodynamics of a medium with $\varepsilon < 0$ and $\mu < 0$ in great detail [56.16]. Such a medium (commonly referred to as the *double-negative* or *negative-index medium*) behaves as a peculiar transparent medium supporting the propagation of electromagnetic waves whose wavevectors $\boldsymbol{k}$ form left-handed triplets with the vectors of electric $\boldsymbol{E}$ and magnetic $\boldsymbol{H}$ fields, i. e., $\boldsymbol{k} \propto [\boldsymbol{H} \times \boldsymbol{E}]$. The latter, in fact, gave rise to the original term of the *left-handed medium* suggested by Veselago, which was eventually abandoned by the metamaterial community in order to avoid confusion with an artificial chiral medium (see Sect. 56.1.5). The double-negative medium also features anomalous (i. e., negative) refraction, the reversal of the Doppler shift, Cherenkov radiation, and even radiation pressure.

These exotic effects, however, could not be experimentally verified until recently due to the seeming absence of negative-index materials. Following Veselago's work, a number of attempts were made to find such materials among naturally available magnetic semiconductors, $CdCr_2Se_4$ and alike, which were known to exhibit both electric and magnetic resonances [56.17]. Unfortunately, the task of minimizing the detuning of the resonances so that they could spectrally overlap proved to be impossible. Consequently, the notion of negative-index medium and the associated electromagnetic phenomena were ignored for over 30 years.

It was the demonstration of artificial magnetism in metamaterials (considered in the previous section) that enabled the engineering of a double-negative response artificially, at virtually any frequency. The first double-negative metamaterial operated at around 5 GHz [56.18]. It was realized by structurally combining an array of 7 mm large double split-ring metamolecules with an array of continuous metallic wires (Fig. 56.6a). The former exhibited narrow-band magnetic response, while the latter produced broadband electric response characteristic to that of a bulk metal, which ensured that the bands of negative effective permeability μ^* and permittivity ε^* were spectrally overlapped. Note that while split rings and wires alone behave as opaque media (since the corresponding effective refractive indices $i\sqrt{|\mu^*|}$ and $i\sqrt{|\varepsilon^*|}$ are imaginary), their combination displays the opposite behavior, acting as an electromagnetically transparent medium, as illustrated in Fig. 56.6b. Subsequent efforts by various research groups worldwide were focused on demonstrating double-negative response at higher frequencies and eventually led to the development of negative-index metamaterials for the near-IR part of the spectrum with refractive indices ranging from -0.3 to -4.0 (see [56.10] and references therein). As in the case of magnetic metamolecules, the designs of negative-index metamolecules operating at optical frequencies were substantially modified to take into account the increasing kinetic inductance of electrons in metals and to respect the limitations of nanofabrication techniques existing at that time.

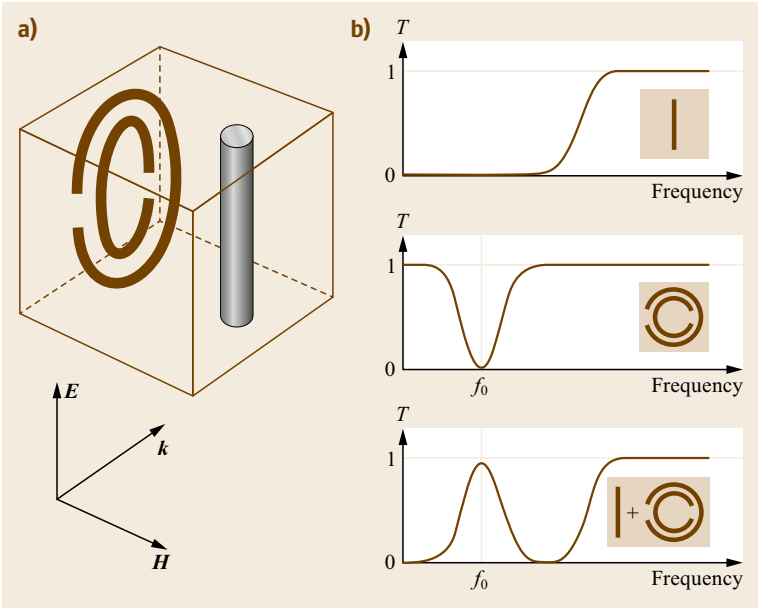


Fig. 56.6a,b Negative-index metamaterial. **(a)** Diagram of the metamolecule – a DSRR combined with a straight wire. **(b)** Transmission spectra schematically shown for metamaterials composed of metallic wires (*top*), DSRRs (*middle*), and their combination (*bottom*)

It is important to note that wave propagation in double-negative media is necessarily complicated by the implicit frequency dependence of ε and μ or, in other words, negative n is *always* strongly dispersive. The latter is dictated by the second law of thermodynamics, which introduces the following constraints valid for nearly transparent media [56.19]

$$\frac{\mathrm{d}(\omega\varepsilon)}{\mathrm{d}\omega} = \omega\frac{\mathrm{d}\varepsilon}{\mathrm{d}\omega} + \varepsilon > 0$$
$$\frac{\mathrm{d}(\omega\mu)}{\mathrm{d}\omega} = \omega\frac{\mathrm{d}\mu}{\mathrm{d}\omega} + \mu > 0\,. \qquad (56.3)$$

Since for negative-index media $\varepsilon < 0$ and $\mu < 0$, it follows that $\mathrm{d}\varepsilon/\mathrm{d}\omega > 0$ and $\mathrm{d}\mu/\mathrm{d}\omega > 0$, which means that both ε and μ cannot be constants.

Negative refraction was the most peculiar (to say the least) effect anticipated in double-negative media. It manifests itself as an anomalous change in the direction of light propagation upon refraction, when a beam of light bends the *wrong* way, while crossing the boundary of double-negative media (Fig. 56.7a). It was the demonstration of negative refraction that shook the scientific community and led to a surge of interest in metamaterial research.

The fact that the incident and refracted waves are on the same side of the normal to the boundary of the negative-index medium follows directly from Snell's law

$$n_2 \sin\theta_2 = n_1 \sin\theta_1\,. \qquad (56.4)$$

Indeed, θ_2 must be negative to satisfy (56.4) when $n_2 < 0$. Negative refraction was first described in reasonable detail by *Mandelshtam*, who concluded that *as unfamiliar as this construction may be, there is, of course, nothing surprising about it* [56.20]. Snell's law, however, does not address the refraction of actual energy carried by the wave. The initial conclusion that in the negative-index medium the energy would flow also in the direction defined by the law of refraction [56.16, 21] was actively contested [56.22–25]. It was eventually shown analytically and numerically that the directions of both phase and group velocities, $\boldsymbol{v}_{\mathrm{ph}}$ and $\boldsymbol{v}_{\mathrm{g}}$, do undergo negative refraction [56.24, 26–29], while the controversy resulted from the fact that in the negative-index medium the wave packets exhibit slanting or *crab-like* motion (Fig. 56.7b). Correspondingly, it was erroneous to interpret the normal to the fronts of the packets as the direction of the group velocity, i. e., energy flow [56.24]. Besides, in isotropic media, the phase and group velocities are always collinear. It was also established that the group velocity of the refracted wave is directed away from the boundary, while its phase velocity was directed toward the boundary (Fig. 56.7b).

Negative refraction was also unambiguously demonstrated by *Smith* et al. [56.1] in an elegant experiment illustrated in Fig. 56.8. A microwave beam was sent onto a prism that had been cut from a double-negative metamaterial (a composite of DSRR and wires), and the deflection of the beam was measured upon its transmission through the prism. The measurements also involved a reference prism made of Teflon, a low-loss dielectric with a refractive index of 1.4. The experimental data indicated that the beam was deflected by the metamaterial prism to negative angles

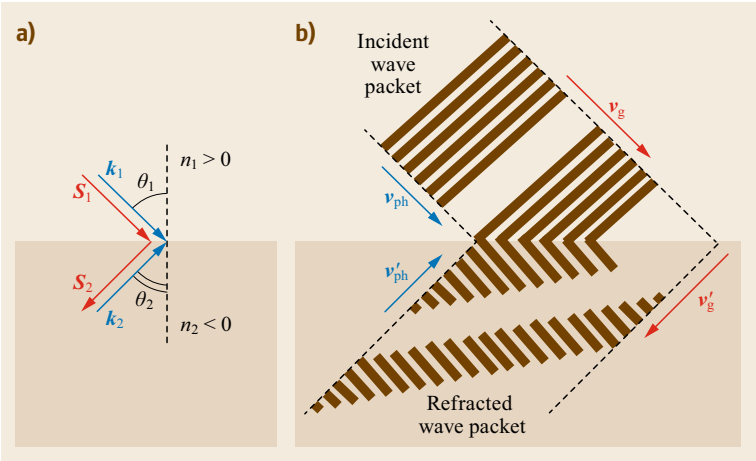


Fig. 56.7a,b Negative refraction. **(a)** Changes in the direction of the wavevector $\boldsymbol{k}$ and energy flow $\boldsymbol{S}$ upon refraction at the interface with a negative-index medium. **(b)** A series of schematically shown wave packets exhibit negative refraction

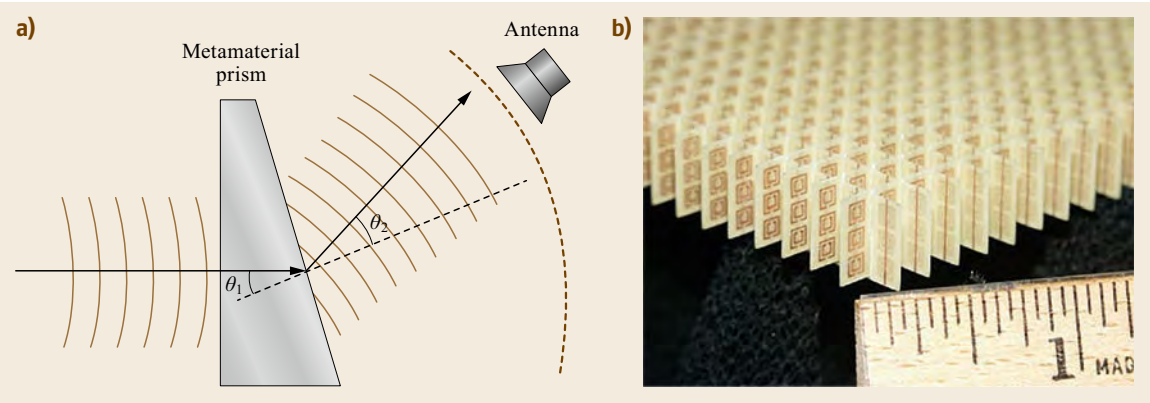


Fig. 56.8a,b Experimental demonstration of negative refraction: **(a)** schematic of the experiment, **(b)** photograph showing a fragment of a double-negative metamaterial prism. (After [56.1]. Image courtesy of NASA)

(compared to the reference case) in the frequency range 10.2 to 10.8 GHz, which had been previously identified as the left-handed propagation region. The deflection angle data were also used to calculate the corresponding index of refraction, which turned out to be negative and highly dispersive (as expected), taking values in the range -2.7 to -0.1. This and subsequent experimental evidence indicated decisively that negative refraction is a reality [56.30].

Perfect lens or superlens

One of the implications of negative refraction noted by Veselago is that a flat slab made of a double-negative material acts as a collecting lens that focuses light rather than disperses it (Fig. 56.9a). Such a lens features no focal plane and, therefore, forms a three-dimensional image of an object (much like a flat metal mirror), provided that its distance from the lens does not exceed the lens' thickness (for $n = -1$). However, unlike a metal mirror, such a *flat lens* produces a real image.

In the special case when $\varepsilon = -1$ and $\mu = -1$, the flat lens reveals additional interesting properties [56.16, 21]. It appears to be perfectly matched to free space exhibiting zero reflection, so that all the energy of the incident wave passes to the refracted wave. Also, light rays arrive at the image plane with zero phase delay, thus preserving the phase information of the object. This can be understood by noting that any ray connecting the object with the image travels half of the distance in air and the other half in the negative-index slab (Fig. 56.9a). Because the phase velocities in the two media have equal magnitudes but opposite signs, the corresponding phase delays compensate each other exactly. However, the most fascinating property of such a lens, perhaps, is that it can produce a sharper image (compared to conventional optical systems) demonstrating resolving power, which exceeds the limit set by the wave nature of light – the so-called diffraction limit. This was shown by *Pendry* in his seminal work *Negative refraction makes a perfect lens* [56.21].

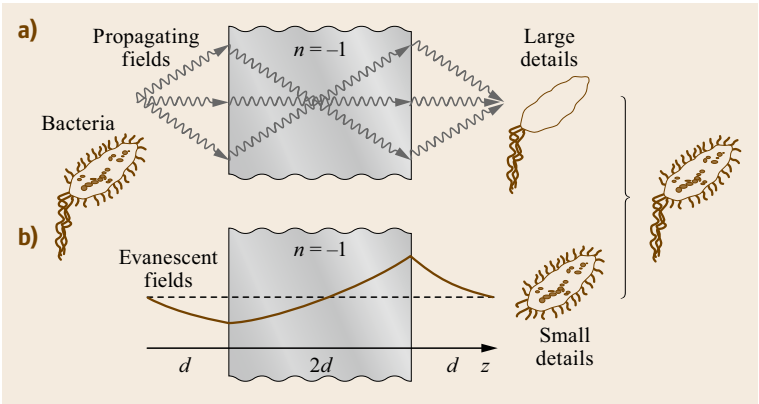


Fig. 56.9a,b Perfect lens. **(a)** The negative-index slab (thickness $2d$) acts as a collecting lens, which focuses *propagating* electromagnetic waves and restores the details of an object that are larger than the wavelength. **(b)** The negative-index slab also acts as an amplifier of *evanescent* waves and restores the details of an object that are smaller than the wavelength

Indeed, the finest details of an object (those that are smaller than the wavelength of light) are carried by the waves with purely imaginary wavevectors. Consequently, these waves decay exponentially as they propagate in air away from the object (*evanescent waves*), and thus cannot be focused by an ordinary lens. In the negative-index slab, the sign of the exponent is reversed and the amplitude of the evanescent waves is bound to increase (Fig. 56.9b). It is easy to see that the decay of the evanescent waves in air is completely negated by the *amplification* occurring inside the slab. The flat lens, therefore, behaves as a *perfect lens* (or *superlens*) – it restores not only the phase but also the amplitude of all light-field components and, as a result, preserves subwavelength features of the object (as illustrated in Fig. 56.9). On the other hand, this optical device should not be regarded as a lens in the usual sense since it cannot focus a parallel beam. The superlens acts as a matching device that transfers light fields between the object and image planes without aberrations; effectively, it *annihilates* the space between the object and the image, so that the image becomes the object.

The superlens, however, has an important drawback. The amplification of the evanescent waves inside the negative-index slab, which at first may seem to violate the conservation of energy, is in fact a transient process. The slab here acts as a resonant cavity that accumulates energy over time for each evanescent-wave component. Correspondingly, imaging (or rather transferring) finer details takes a progressively longer time since the slab must amplify evanescent components with a stronger decay, i. e., smaller input level. Furthermore, the resonant enhancement is limited by inevitable material losses, and although this limitation would not necessarily ruin the super-resolution completely [56.21], the perfect lens will operate efficiently only when the distance to the object z (as well as the thickness of the lens $2d$) is of the order of the wavelength λ_0.

56.1.5 Chiral Metamaterials

While research efforts have mainly been focused on engineering composite media with particular values of permittivity and permeability (or, alternatively, refractive index), the metamaterial approach also enables the enhancement of other, more complex optical properties, such as those governed by chirality [56.31]:

> *I call any geometrical figure, or group of points, chiral, and say it has chirality, if its image in a plane mirror, ideally realized, cannot be brought to coincide with itself.*

In short, any object that cannot be superimposed with its mirror image is chiral. A spring coil (helix), a screw, and a human's hand are the celebrated examples of chiral objects. Chirality is the most fundamental type of asymmetry, which in matter leads to the effect of *optical activity*. It is characterized by a difference between the refractive indices (i. e., phase velocities) experienced by right and left-hand circular polarizations in a chiral medium, $\Delta n = n_+ - n_-$, and leads to a progressive rotation of linear polarization as light propagates through the medium (Fig. 56.10)

$$\Delta\phi = \frac{\Delta n d\pi}{\lambda_0}\,, \tag{56.5}$$

where $\Delta\phi$ is the angle of polarization azimuth rotation, d is the thickness of the medium, and λ_0 is the wavelength of light.

In a lossy medium, the effect is also accompanied by a difference in the absorption of right and left-hand circular polarizations, known as *circular dichroism*. This effect was discovered in quartz by Arago in 1811 and linked to the structural asymmetry of matter by Pasteur in 1848. It is now widely used in analytical chemistry, biology, and crystallography for identifying the spatial structure of molecules and crystals. Chiral materials are always in demand in optoelectronics, life

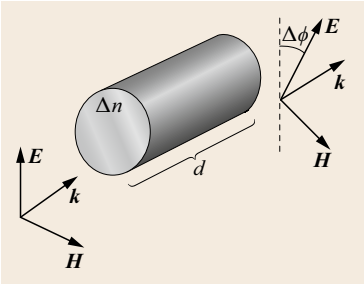


Fig. 56.10 Optical activity: polarization azimuth of linearly polarized light rotates as the latter propagates inside a chiral medium

science microscopy, and display technology providing a means of manipulating the polarization state of light; they are also widely used in photography.

In naturally available chiral media, optical activity is relatively weak and, therefore, slabs with the thickness of hundreds of wavelengths must be used to achieve measurable rotation. For example, the specific rotatory power in quartz, cinnabar, and liquid crystals corresponds to 20°/mm, 600°/mm, and 1000°/mm [56.32] in the visible part of the spectrum. Chiral metamaterials, on the other hand, can exhibit optical activity several orders of magnitude stronger [56.33–35]. The latter is particularly appealing for applications in the terahertz frequency range (and below) where the effect is virtually absent. There are many possible realizations of the chiral metamolecule depending on the application, targeted frequency range, and limits of the fabrication technique, but in all cases its design reveals a structural twist that can be related to either right or left-handed screws (Fig. 56.11).

The first chiral metamolecules, which were experimentally tested in 1898 by *Bose* using millimeter waves [56.8], were made of twisted jutes (Sect. 56.1.2). Two decades later, *Lindman* [56.36] suggested using small (a few centimeters long) randomly oriented copper helices as artificial molecules that would promote polarization azimuth rotation for radio waves. The idea was verified and eventually extended to the case of oriented helices by *Tinoco* and *Freeman* [56.37]. Recent advances in three-dimensional nanofabrication techniques have enabled the realization of such artificial chiral molecules in the optical part of the spectrum [56.38]. The demonstrated array of submicron gold helices formed a uniaxial photonic metamaterial, which operated as a compact and efficient broadband circular polarizer. A very different approach was taken by *Zheludev* et al. [56.39], who proposed the design of chiral metamolecules based on pairs of identical, spatially separated, and mutually twisted planar metallic resonators such as bars, crosses, rosettes, split rings, etc. (Fig. 56.11c–f). The advantage of this approach was that chiral metamaterials could be readily fabricated for any frequency of operation in the range from microwave [56.33] to optical [56.40, 41] frequencies using well-established planar lithography techniques. Moreover, due to the resonant electromagnetic (inductive [56.39]) coupling between the constituent metallic elements of the bi-layered metamolecule, the complete 90° rotation of the polarization can be achieved with just one layer of such metamolecules (with the thickness of the resulting metamaterial slab in the direction of light propagation being less than $\lambda_0/30$) [56.34]. Zheludev's approach also enables the engineering of strong, virtually dispersionless polarization rotation within a wide spectral range [56.42] when the twisted pair is formed by complimentary metallic patterns (Fig. 56.11g). Optical activity may also be observed for bi-layered meta-

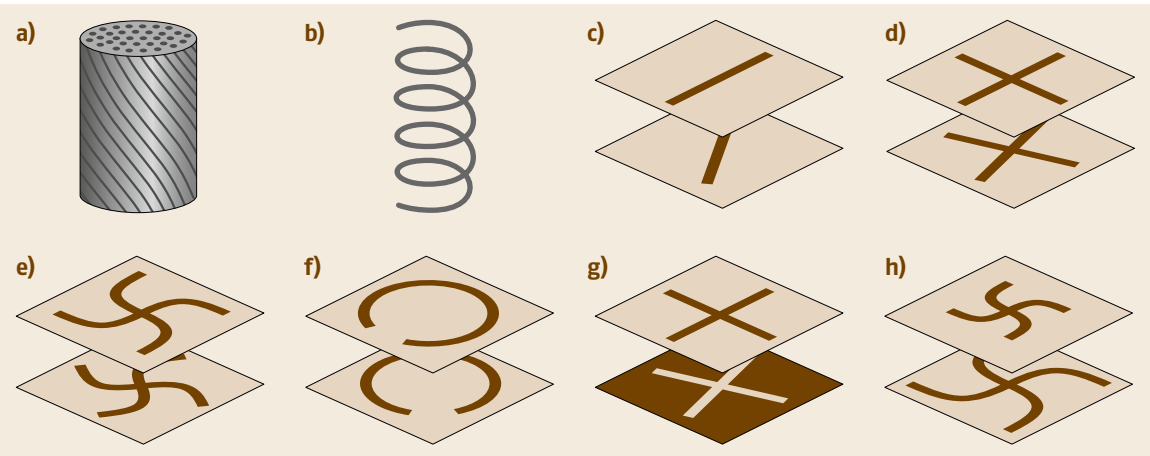


Fig. 56.11a–h Examples of chiral metamolecules: twisted jute (**a**), wire coil (**b**), pairs of mutually twisted metallic bars (**c**), crosses (**d**), rosettes (**e**), split rings (**f**), mutually twisted wire cross and cross-shaped aperture (**g**) and a pair of metallic rosettes of different sizes featuring zero twist (**h**)

molecules that feature zero twist between their layers (Fig. 56.11h). In such a case, the structural twist of the constituent resonators was shown to be sufficient provided that the resonators have different sizes [56.43].

It is easy to see that the designs of chiral metamolecules can be reduced to the combination of two basic elements, a split ring and a cut wire resonator electrically connected together, as shown in Fig. 56.12. This illustrates that chiral metamaterials exhibit simultaneously electric and magnetic responses, which appear to be mixed such that magnetic dipole moments are induced in the split rings by the electric component of light (via cut wires), and vice versa. This cross-coupling between the electric and magnetic fields is described by a dimensionless chirality parameter κ, which in the simplest case of an isotropic chiral medium is a scalar [56.44]

$$\begin{aligned} \boldsymbol{D} &= \varepsilon_0 \varepsilon \boldsymbol{E} - \mathrm{i}\kappa \sqrt{\varepsilon_0 \mu_0} \boldsymbol{H} \\ \boldsymbol{B} &= \mu_0 \mu \boldsymbol{H} + \mathrm{i}\kappa \sqrt{\varepsilon_0 \mu_0} \boldsymbol{E} \,. \end{aligned} \tag{56.6}$$

It is the chirality parameter that gives rise to the difference between the refractive indices for right- and left-hand circular polarizations

$$n_{\pm} = \sqrt{\varepsilon \mu} \pm \kappa \,. \tag{56.7}$$

While the studies of artificial chiral media found applications in many areas of microwave engineering, such as antennas, absorbers, polarizers, and waveguides [56.45, 46], the recent explosive interest in chiral metamaterials was triggered by the opportunity of achieving negative refraction in such media in a relatively simple way. It was identified independently by *Tretyakov* et al. [56.47] and *Pendry* [56.48], who noted that according to (56.7) the index of refraction of a chiral medium becomes negative for one of the circularly polarized waves if $|\kappa| > \sqrt{\varepsilon \mu}$. Also, in that case neither ε nor μ need to be negative, which was in stark contrast to the previously formulated, more complex condition for the negative refraction of linearly polarized waves (Sect. 56.1.4). Thus, the proposed chiral route to negative refraction, although limited to one circular polarization, relies on the engineering of only one resonance – the resonance of optical activity (where κ can become sufficiently large), and is, therefore, expected to improve and simplify the designs of negatively refracting composites [56.48]. The validity of this approach was confirmed experimentally using microwave [56.49] and terahertz [56.50] chiral metamaterials based on stacks of mutually twisted metallic patterns, which enables effective refractive indices as low as -2 and -5, respectively, to be attained.

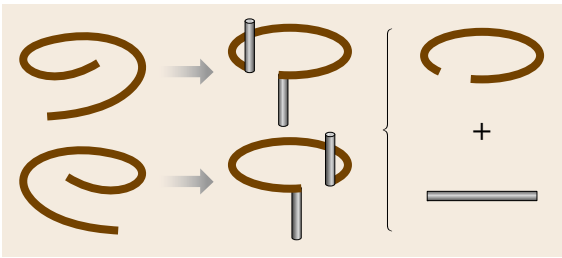


Fig. 56.12 Chiral metamolecule and its two basic structural elements – a split ring and a cut wire

56.1.6 Transformation Optics and Metamaterial Cloaks

One of the successful applications of the metamaterial concept was in the practical implementation of transformation optics, a mathematical technique aimed at designing complex electromagnetic media that can control the flow of light in a desired manner [56.51, 52]. The transformation optics approach enables the recreation of an electromagnetic phenomenon in a warping virtual space as an optical illusion in the physical Euclidean space. For example, the effect of gravitational lensing caused by a massive celestial body (which is treated by the general theory of relativity as the manifestation of a gravity-induced distortion in an abstract spacetime continuum) can be viewed as the ordinary refraction of light by a shell of the dielectric medium enclosing the body in the physical space. This appears possible because Maxwell's equations have exactly the same form in any coordinate system, while the permittivity and permeability are just scaled by a common factor, which absorbs the transformation from one coordinate system into another [56.53]

$$\varepsilon'^{\mathrm{i'j'}} = |\det(\boldsymbol{\Lambda})|^{-1} \boldsymbol{\Lambda}^{\mathrm{i'}}_{\mathrm{i}} \boldsymbol{\Lambda}^{\mathrm{j'}}_{\mathrm{j}} \varepsilon^{\mathrm{ij}} \,, \tag{56.8a}$$

$$\mu'^{\mathrm{i'j'}} = |\det(\boldsymbol{\Lambda})|^{-1} \boldsymbol{\Lambda}^{\mathrm{i'}}_{\mathrm{i}} \boldsymbol{\Lambda}^{\mathrm{j'}}_{\mathrm{j}} \mu^{\mathrm{ij}} \,, \tag{56.8b}$$

where $\boldsymbol{\Lambda}^{\alpha'}_{\alpha} = \partial x'^{\alpha'} / \partial x^{\alpha}$ denotes the Jacobian transformation matrix between the virtual space $\{x^{\alpha}\}$ and the physical space $\{x'^{\alpha'}\}$.

The correspondence between coordinate transformations and the material parameters was noted by *Dolin* more than 50 years ago [56.54], but it was perceived as being merely a mathematic trick because (56.8a) and (56.8b) do not impose physical constraints on the values of the transformed permittivity and permeability. The perception, however, has changed recently with the advent of the metamaterial approach, which in principle enables the engineering of extreme values of ε' and μ', as well as their spatial variation prescribed by the transformations in (56.8). Originally driven by the desire to realize the invisibility cloak, transformation optics has now become a powerful tool for designing a wide range

of new optical effects and devices, including electromagnetic wormholes, hidden gateways, optical black holes, field rotators and concentrators, etc. (see [56.55] and references therein).

The *shell cloak* is the most remarkable example of space distortion; it can steer light around an object making the latter invisible, when imitated in the physical space using metamaterials. The distortion squeezes the space from within a sphere with the radius R_2 to within a shell with the same outer radius and inner radius R_1 (Fig. 56.13). Electromagnetic waves will not interact with or scatter from the core of the shell because it is not a part of the resulting curvilinear space, while the shell itself will neither scatter waves nor cast a shadow in the transmitted field (either of which would allow one to detect the cloak).

In the local spherical coordinate system with the origin placed at the center of the sphere, the transformation rules give the following values of ε' and μ' in the shell ($R_1 < r' < R_2$) [56.52]

$$\begin{aligned}
\varepsilon'_{r'} &= \mu'_{r'} = \frac{R_2}{R_2 - R_1}\frac{(r' - R_1)^2}{r'} , \\
\varepsilon'_{\theta'} &= \mu'_{\theta'} = \frac{R_2}{R_2 - R_1} , \\
\varepsilon'_{\varphi'} &= \mu'_{\varphi'} = \frac{R_2}{R_2 - R_1} .
\end{aligned} \tag{56.9}$$

Correspondingly, ε' and μ' are free to take any values in the core ($r' < R_1$), while $\varepsilon' = \mu' = 1$ outside the shell ($r' > R_2$). It is easy to see that (56.9) correspond to extremely anisotropic distributions of ε' and μ' in the global Euclidean space, while some of the components of the permittivity and permeability tensors, although positive, approach zero near the core. Clearly, no natural material system can offer such complex variations of its parameters.

A proof-of-principle demonstration of a metamaterial shell cloak was performed in 2006 by *Smith* et al. [56.56]. The experiments were conducted in the microwave part of the spectrum, where the fabrication of bulk metamaterial samples and their characterization was generally easier to accomplish. Even in that case, the implementation of a true three-dimensional cloak was not practical, given the limitations of the measurement technique and the complexity of the metamaterial properties [56.56]. The cloaking principle was demonstrated for TE-waves in a parallel-plate waveguide using a two-dimensional cylindrical shell structure (Fig. 56.14). Such a configuration of the experiment enabled substantial simplification of the metamaterial design. Also, additional steps were taken to reduce the set of required material parameters; however, this was at the expense of nonzero reflectance. The metamaterial shell was formed by 10 concentric cylindrical layers of 3.3 mm large magnetic SRR metamolecules, and had the inner and outer radii $R_1 = 27.1$ mm and $R_2 = 58.9$ mm, respectively. The cloak was designed to operate at 8.5 GHz.

Figure 56.15 compares the measured and simulated fields inside and outside the cloak, which show a rather good match. As the waves propagate through the metamaterial shell, the center section of the wavefront starts to slow down when it approaches the core, exhibiting wavelength compression. The wavefront splits to pass around the core and rejoins on the opposite side. Upon exiting the shell, the wavefronts are seen to match those propagating alongside in the empty space. The

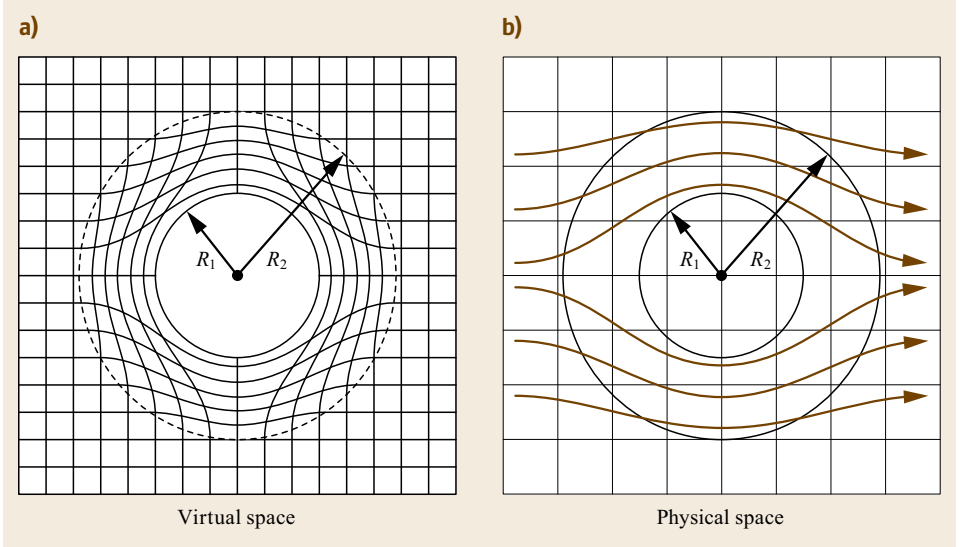


Fig. 56.13a,b Transformation optics. **(a)** Space distortion of a finite extent that results in **(b)** an electromagnetic cloaking shell, when imitated in the physical space through a distribution of material parameters

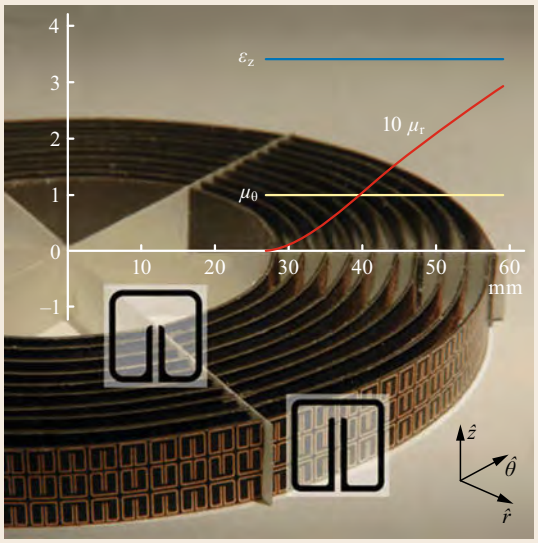


Fig. 56.14 Photograph of a two-dimensional microwave cloaking shell. (Reprinted from [56.56], with permission from AAAS)

fields outside the cloak also feature a small but non-negligible component of scattering (reflection) resulting from the implementation of the reduced parameters set. A pronounced attenuation of the transmitted fields is attributed to resistive losses present in the metamaterial structure.

Although this demonstration illustrates the feasibility of the metamaterial approach in the context of transformation optics, the practical deployment of a shell invisibility cloak is still rather challenging. The main reason is that the required values of the material parameters in this type of cloak are attained via metamaterial resonances, which renders it as an inherently narrow-band device. Also, the resonant nature of metamaterial properties makes cloaking performance highly sensitive to resistive losses in the constituent materials (especially at optical frequencies), while fairly complex design and strong electromagnetic anisotropy of the shell cloak call for the development of sophisticated computation and fabrication techniques.

The *carpet cloak* or ground-plane cloak represents an alternative, easier to implement strategy of concealing an object [56.57]. It employs a reflective bump (a perturbation on a surface of a conducting plane) covered by a layer of metamaterial with a specially tailored variation of the refractive index. The bump provides a cavity for hiding the object, while the metamaterial layer conceals the presence of the bump, creating an illusion of a flawless mirror (see Fig. 56.16a–c). In this case, given the restricted geometry of the cloak, it becomes possible to minimize the anisotropy of the transformed material parameters, while the required values of the refractive index turn out to be merely greater than unity. This allows one to substantially simplify the design of the metamaterial cover and exploit metamolecules off their resonance, which makes the carpet cloak a truly broadband and scalable optical device.

The first metamaterial carpet cloak was experimentally demonstrated in 2009 at microwave frequencies by *Smith* et al. [56.58], who built the device using over 10,000 I-shaped 2 mm large metamolecules. Its functionality was confirmed for the 13−16 GHz range of frequencies and could be extended to up to 18 GHz. Low levels of absorption due to nonresonant operation of carpet cloaks and the relative simplicity of their designs have enabled successful demonstrations of the concept also in the IR [56.59] and visible [56.60] domains (Fig. 56.16d).

56.1.7 Hyperbolic Metamaterials

Hyperbolic metamaterials are perhaps the only class of artificial electromagnetic media that does not rely on resonant interaction with the incident radiation. Instead, the unusual optical properties of the hyperbolic metamaterials result from extreme anisotropy of their *effective* permittivity ε^*, when the sign of ε^* reverses for the electric field parallel ($\|$) and orthog-

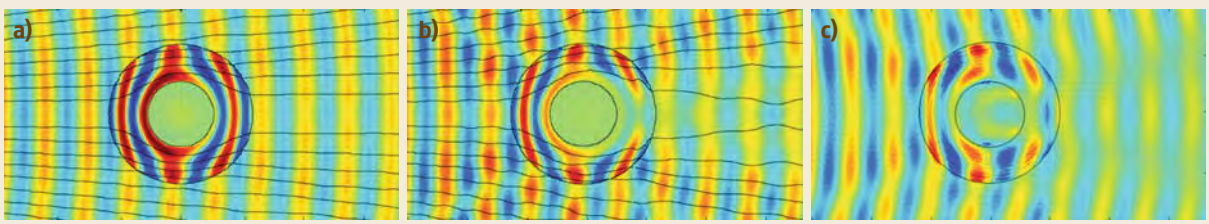


Fig. 56.15a–c Two-dimensional microwave cloak in action: propagation of an electromagnetic wave in the presence of the cloak with **(a)** exact material parameters in a simulation, **(b)** reduced material parameters in a simulation, and **(c)** reduced material parameters in the experiment. (Reprinted from [56.56], with permission from AAAS)

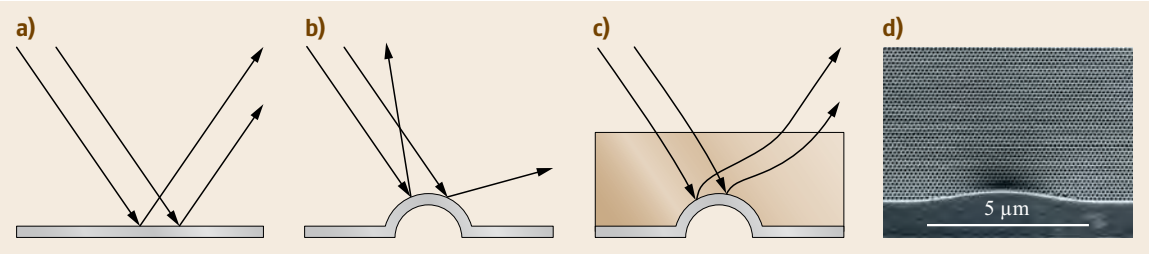


Fig. 56.16a–d Reflection of light rays shown schematically for **(a)** a flat mirror, **(b)** a deformed mirror, and **(c)** a deformed mirror covered by a carpet cloak. **(d)** Scanning electron microscope image of a carpet cloak operating in the visible spectral range. (Reprinted with permission from [56.60]. © American Chemical Society)

onal ($\perp$) to the anisotropy axis of the metamaterial crystals. In other words, depending on the direction of light propagation (and polarization) such metamaterials behave either as metals or as dielectrics (i. e., insulators). This is achieved for material composites typically formed by an array of parallel metallic wires/rods embedded into a dielectric matrix (Fig. 56.17a), or a stack of alternating thin layers of metal and dielectric (Fig. 56.17b). Normally (at visible frequencies and below), in the first case, $\varepsilon^*_{\parallel} < 0$ and $\varepsilon^*_{\perp} > 0$ (type I anisotropy), while in the second case, $\varepsilon^*_{\parallel} > 0$ and $\varepsilon^*_{\perp} < 0$ (type II anisotropy). The anisotropy of both types leads to unusual, hyperbolic dependences for the frequency ω of the allowed TM-polarized modes ($H_{\parallel} = 0$) on their wavevectors $\boldsymbol{k}^* = \{k^*_x, k^*_y, k^*_z\}$, which explains the name given to this class of metamaterials (Fig. 56.17c,d)

$$\frac{k^{*2}_z}{\varepsilon^*_{\perp}} - \frac{k^{*2}_x + k^{*2}_y}{|\varepsilon^*_{\parallel}|} = \frac{\omega^2}{c^2} \quad \left(\varepsilon^*_{\parallel} < 0, \varepsilon^*_{\perp} > 0\right) \qquad (56.10a)$$

$$\frac{k^{*2}_x + k^{*2}_y}{\varepsilon^*_{\parallel}} - \frac{k^{*2}_z}{|\varepsilon^*_{\perp}|} = \frac{\omega^2}{c^2} \quad \left(\varepsilon^*_{\parallel} > 0, \varepsilon^*_{\perp} < 0\right) . \qquad (56.10b)$$

In principle, such relations (known as the dispersion relations) allow, at any given frequency, the propagation of waves with infinitely large wavevectors, which is in stark contrast to spherical or, in the general case, elliptical dispersions characteristic of natural materials. In practice, however, the range of the permitted wavevectors k^* appears to be capped, although at rather large magnitudes, due to the small but finite structural periodicity of the composites at the microscopic level (owing to the effective medium description no longer being valid when the wavelength becomes comparable to the composite's lattice constant).

One of the immediate consequences of the hyperbolic dispersions is that there are no evanescent waves, while for certain light propagation directions, the effective index of refraction can be extremely large. The latter finds applications in miniaturizing photonic devices, subdiffraction photonic funnels, and nanoscale resonators (see [56.61] and references therein). Also, due to strong material dispersion, the hyperbolic media can act as spatial frequency filters [56.62], while the extreme anisotropy of its dielectric permittivity ε^* can be exploited in linear polarizers [56.63].

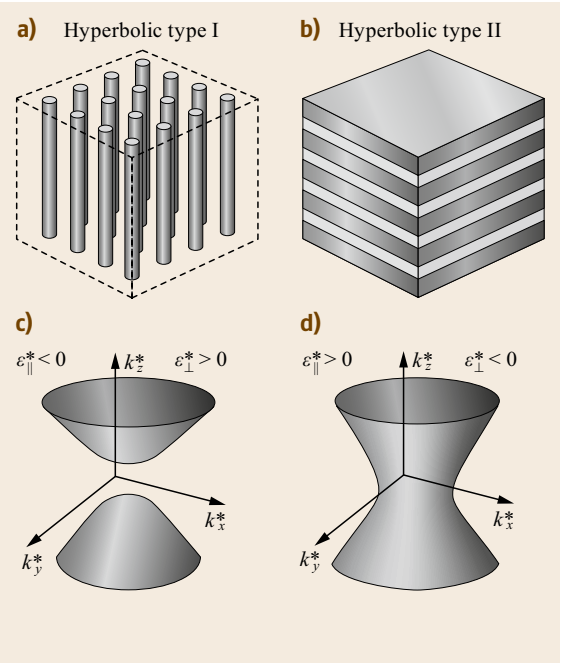


Fig. 56.17a–d Hyperbolic metamaterials of types I and II. **(a,b)** Metamaterial crystals exhibiting hyperbolic dispersion. **(c,d)** Isofrequency surfaces of extraordinary waves propagating in hyperbolic media of types I and II, correspondingly

Negative Refraction

Negative refraction was perhaps the first phenomenon demonstrated in hyperbolic media [56.65, 66]. In particular, the effect can be observed for TM-polarized waves at a facet cut perpendicular to the anisotropy axis of a type I metamaterial crystal (Fig. 56.18a). In this

case, the dielectric permittivity in the direction normal to the vacuum–metamaterial interface is negative and, due to the continuity of the normal component of the electric displacement across the interface ($E_z = \varepsilon^* E_z^*$), the normal component of the wave's electric field inside the metamaterial E_z appears to be reversed. As a result, the component of the Poynting vector propagating in the metamaterial parallel to the interface S_x^* is also reversed, rendering the angle of the beam refraction negative (Fig. 56.18a).

Nevertheless, a hyperbolic metamaterial slab operating in the negative refraction regime can exhibit only partial focusing of light, as illustrated in Fig. 56.18b. This is because the mutual orientation of the Poynting vector and the wavevector (in other words, group and phase velocities) is not always antiparallel or even collinear, as in the case of isotropic negative-index media, but depends on the propagation direction relative to the metamaterial's anisotropy axis.

Subwavelength Imaging

Subwavelength imaging is one of the most promising applications of hyperbolic media. Unlike negative-index metamaterials, which enable the complete restoration of evanescent waves at the image plane (Sect. 56.1.4), the hyperbolic media make it possible to convert evanescent waves in air/vacuum into propagating ones, which then can be collected in the far-field zone and processed by conventional optical microscopes. The device that does such a trick is called the hyperlens [56.67].

A conventional hyperlens has the form of a hollow cylinder made of a type I hyperbolic metamaterial. An example of such a lens designed to operate in the ultraviolet (UV) spectrum is shown in Fig. 56.19a. If an object is placed on the inner surface of the cylinder, the scattered evanescent waves, which carry all the subwavelength information in air/vacuum, can enter the hyperlens where they will propagate radially with no decay (k_r^* is no longer imaginary), as dictated by the hy-

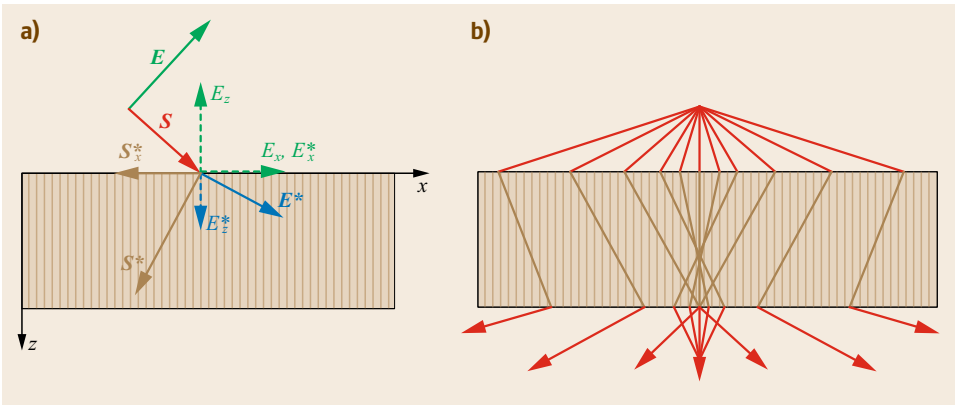


Fig. 56.18a,b Negative refraction in a hyperbolic metamaterial of type I. **(a)** Components of electric field ***E*** and pointing vector ***S*** at the surface of the metamaterial. **(b)** Partial focussing by a slab of the metamaterial

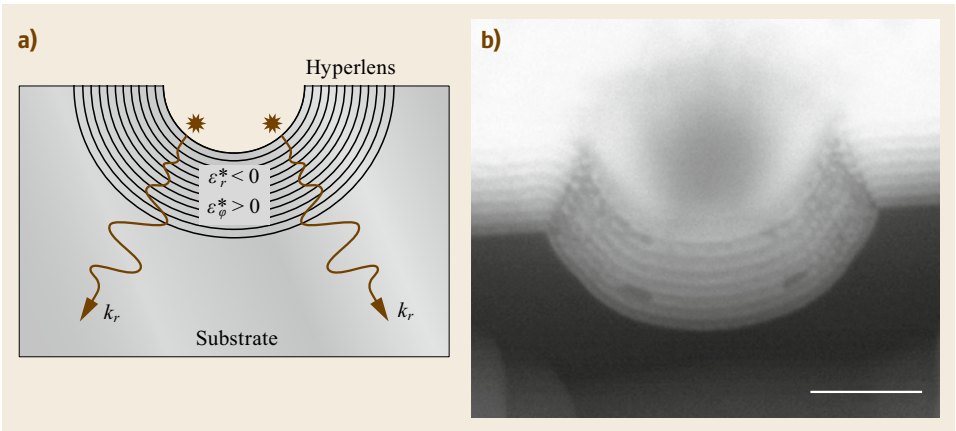


Fig. 56.19a,b Hyperlens. **(a)** The lens and principle of its operation. The lens is formed by a stack of curved alternating layers of metal and dielectric, which behaves as a type I hyperbolic metamaterial at UV frequencies. **(b)** Scanning electron microscope image of the cross-sectional view of a lens operating at the wavelength of 410 nm. The scale bar is 500 nm. (After [56.64])

perbolic dispersion. While the waves spread inside the hyperlens, the azimuthal component of their wavevectors k_φ^* gradually becomes smaller due to the conservation of the angular momentum in the cylindrical geometry, $k_\varphi^* r = \text{const}$. Correspondingly, at a certain radius r, the azimuthal component of the wavevectors may become small enough ($k_\varphi^* < \omega/c$) for the radial component to be real even outside the cylinder,

$$k_\mathrm{r} = \sqrt{\frac{\omega^2}{c^2} - k_\varphi^2}\,, \tag{56.11}$$

where $k_\varphi = k_\varphi^*$. It is easy to show that the resolution of such a hyperlens, that is, the smallest separation between two successfully resolved features of the object, is simply proportional to the ratio between the inner and outer radius of the cylinder [56.67]. Hyperlenses have been successfully demonstrated experimentally in the microwave [56.68], visible [56.69, 70], and near-UV [56.64, 71] domains (Fig. 56.19b).

Enhancement of Spontaneous Emission

Enhancement of spontaneous emission (or control of spontaneous emission) is another promising application of hyperbolic metamaterials. It is based on Fermi's golden rule, which states that the lifetime of spontaneous emission is strongly influenced by the density of available electromagnetic modes [56.72]. Given the analogy with the electronic density of states in solid state physics, the latter is proportional to the area of the isofrequency surface defining the dispersion, and is, therefore, finite in dielectrics (where the dispersion is elliptical) but diverges in hyperbolic media. Consequently, when an emitter is brought close to the surface of a hyperbolic metamaterial it starts to couple to the available high-k states [56.73], and the radiative decay rate of the emitter increases compared to that in air/vacuum.

There are, however, several factors that in practice limit the enhancement of spontaneous emission. Most notably these are the finite structural periodicity of a metamaterial composite, which introduces a cut-off for the largest wavevectors permitted by the metamaterial, and the finite separation between the emitter and the metamaterial, which makes it impossible for some of the evanescent harmonics to reach the metamaterial and hence couple to its high-k states.

Although many experiments involving hyperbolic metamaterials of types I and II demonstrated a shortening of the lifetime [56.73–75], the discrimination between the true radiative mechanism involving propagating high-k modes and the nonradiative mechanism (i. e., dissipation due to ohmic resistance of materials) has been challenging [56.75, 76].

56.2 Planar Metamaterials

Planar metamaterials, also known as metafilms or metasurfaces, represent a special class of low-dimensional artificial media. They are formed by metal films periodically patterned on a *subwavelength scale*, which are many times thinner than the wavelength of light. Despite their vanishing thickness, planar metamaterials can strongly interact with electromagnetic waves, which they transmit, absorb, or reflect without diffraction, acting as media of *zero* dimension in the direction of light propagation. As a result, planar metamaterials are able to efficiently mimic (and improve) the functionalities of the existing bulk optical components such as lenses, prisms, polarizers, filters, beam splitters, etc. More importantly, they make accessible a range of new exotic effects including a magnetic mirror, perfect absorption, asymmetric transmission, optical activity without chirality, wavefront rectification, etc.

Importantly, while practical implementation of bulk metamaterials (especially in the optical domain) still very much depends on advances in the development of three-dimensional fabrication techniques, planar photonic metamaterials are fully compatible with the existing micro and nanofabrication techniques exploited by complementary metal–oxide–semiconductor (CMOS) technology. This, in combination with the unusual electromagnetic properties of the planar metamaterials achieved through the nanoscale control of light, renders this type of artificial media an intriguing and viable alternative to various conventional photonic components.

56.2.1 Frequency Selective Surfaces

One of the most basic properties of planar metamaterials is the ability to resonantly transmit, reflect, or absorb electromagnetic radiation within narrow frequency bands that can be engineered on demand. The idea of such metamaterial filters, which were termed frequency selective surfaces (FSS) because of their vanishing thickness, originated in the RF and microwave communities [56.77, 78] and can be traced back to the work on phased antenna arrays done in the 1960s [56.79]. The simplest example of an FSS is the perforated metal screen placed behind the glass window of a microwave oven, which blocks microwave radiation escaping from

the oven but at the same time allows one to see through the window.

An FSS is typically formed by an array of complex-shaped patches of a metal foil arranged on a two-dimensional lattice with a *subwavelength* period, and mechanically supported by a thin dielectric substrate (Fig. 56.20a). The FSS is transparent everywhere apart from a specific range of frequencies, known as the stop band, where the array of patches reflects electromagnetic waves as if it were a continuous sheet of metal (Fig. 56.20b). Such a spectrally selective response is underpinned by the electromagnetic resonances of individual patches, which have the same nature as the *geometrical* resonances in bulk metamaterials (Sect. 56.1.1). They are controlled by the shape and size of the patches (also regarded as metamolecules), and the dielectric permittivity of the immediate surroundings. Consequently, planar metamaterials share most of the design features of their bulk counterparts (Fig. 56.2), and the wavelength of their resonances $\lambda_0 = c/f_0$ is given by the expressions similar to (56.1),

$$\lambda_0 \approx \sqrt{\frac{\varepsilon+1}{2}}\frac{2l_e}{m}\,, \tag{56.12}$$

where l_e is the total electrical length of the metallic pattern that forms the metamolecule, ε is the relative permittivity of the supporting substrate, and m is a positive integer. As with bulk metamaterials, resonances with $m > 4$ are usually not possible to engineer. The only difference between (56.1) and (56.12) is that factor $\sqrt{\varepsilon}$ has been replaced with $\sqrt{(\varepsilon+1)/2}$. Such a modification reflects the fact that an FSS is usually supported by a dielectric on one side, while the other side contacts with air. The effective permittivity in this case can be calculated as the average between the permittivities of the dielectric and air, which happens to be a good approximation that works well for dielectric layers as thin as $0.05\lambda_0$ [56.77]. It is easy to see that when an FSS is sandwiched between two dielectric layers with permittivity ε, (56.12) and (56.1) become identical. As in the case of bulk metamaterials, (56.12) works particularly well for metamaterials operating at near-IR frequencies and below, where metals are still good conductors and can, therefore, support electrical currents.

An FSS can also be formed by an array of complex-shaped subwavelength apertures, rather than patches, made in a thin metal screen (Fig. 56.20c). Such an *inverse* or *negative* type of planar metamaterials does not necessarily require supporting substrates, as mechanical stability here is often ensured by the metal screen itself. A negative FSS behaves as a mirror everywhere apart from in a certain range of frequencies, known as the passband, where electrical currents induced in the perforated screen are resonantly suppressed, rendering the metallic structure electromagnetically transparent (Fig. 56.20d). The spectral selectivity of the negative FSS is shaped by the resonances of individual apertures, which are different in nature from the resonances of conventional (i. e., positive) metamolecules and are usually described with the help of fictitious *magnetic currents* [56.80]. Nevertheless, the passband of a negative FSS can be predicted based on the spectral response of its complementary structure. According to the Babinet principle, the specular reflection coefficient of a perforated metal screen equals the transmission coefficient of a *free-standing* array of metallic patches that have the same shape as the apertures in the screen, and vice versa. Such an approach works particularly well in the RF and microwave domains where metals behave as perfect conductors, while the thickness of the screen can be made as $\lambda_0/10^3$ or less. Thicker screens have the

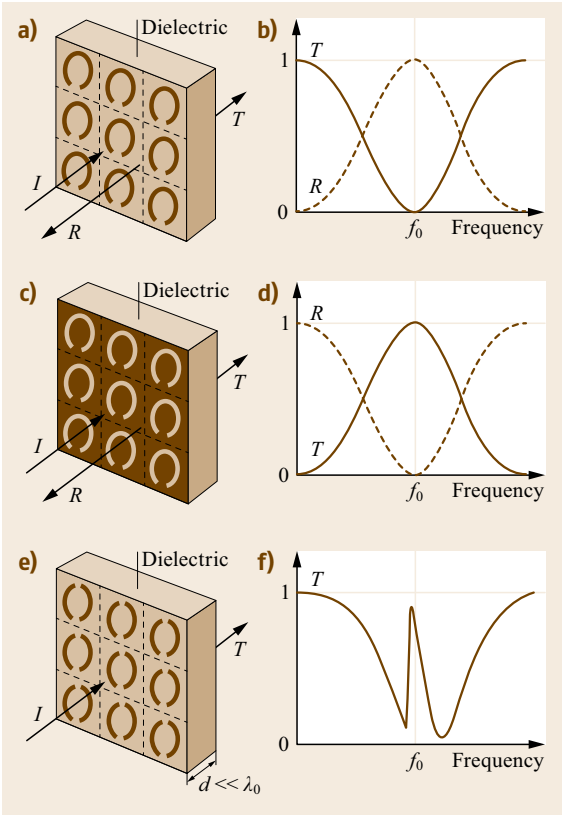


Fig. 56.20a–f Examples of frequency selective surfaces based on SRR. **(a)** An array of split-ring patches and **(b)** its transmission (*solid line*) and reflection (*dashed line*) spectra. **(c)** An array of split-ring slits made in a thin metal foil and **(d)** its transmission (*solid line*) and reflection (*dashed line*) spectra. **(e)** An array of asymmetrically-split-ring patches and **(f)** its transmission spectrum

opposite effects on the bandwidths of the complimentary metamaterial patterns, making the transmission passband narrower and the stopband wider. Also, the two types of FSS will behave differently if the thickness of the dielectric substrates d exceeds $\lambda_0/4$ [56.77].

The shape of the transmission stopband (or passband) and its width can be modified by introducing additional resonant elements in the design of the metamolecules (e.g., asymmetrically split wire rings, Fig. 56.20e) or, alternatively, by cascading several FSS [56.77, 78]. In particular, both methods can be used to engineer a very narrow window of transparency within the stopband of a planar metamaterial (Fig. 56.20f), which mimics the famous effect of electromagnetically induced transparency in cryogenically cooled gases [56.81–83]. The metamaterial transparency window is characterized by a very sharp normal dispersion of the phase of passing electromagnetic waves, which corresponds to a substantial reduction of the group velocity. For example, a $\lambda_0/35$ thick planar structure can delay electromagnetic pulses by more than $1/3$ of their initial width [56.82].

FSSs play a key role in many antenna systems for modern fixed and mobile communication systems [56.77, 78], and can be configured to operate as low-pass, high-pass, and band-pass, or dichroic filters [56.84–87]. Other applications include laser–cavity output couplers [56.88, 89] and narrow-band infrared sources [56.90–92]. Also, owing to their resonant behavior, planar metamaterials are able to enhance light–matter interaction, facilitating optical sensing [56.93, 94], energy harvesting [56.95], and generation of coherent radiation [56.96, 97].

56.2.2 Magnetic Mirrors (High Impedance Surfaces)

If an FSS is placed on top of a continuous metal screen such that the supporting substrate is sandwiched between the metallic pattern of the FSS and the screen, acting as a spacer, the electromagnetic boundary conditions at the surface of the resulting planar metamaterial can be altered in a peculiar way. In contrast to the case of regular metallic mirrors (Fig. 56.21a), the direction of the E-field component of an electromagnetic wave reflected off the structure will not reverse. Instead, the reversal will occur for the H-field component (Fig. 56.21b), and the total magnetic field tangential to the surface will vanish [56.98, 99], as if the metamaterial acts as an *artificial magnetic conductor* (hence the term *magnetic mirror*).

When the change of the phase upon reflection $\Delta\varphi$ is exactly zero and ohmic losses are absent, such a planar metamaterial can be regarded as a perfect magnetic conductor, the counterpart of a perfect electric conductor. In practice, the reflection phase of a magnetic mirror crosses zero only at a certain frequency f_0, corresponding to the resonant mode of the structure (Fig. 56.21c). The useful bandwidth of a magnetic mirror is usually defined by $|\Delta\varphi| < \pi/2$ to either side of f_0, since these changes of the phase would not cause destructive interference between the incident and reflected waves [56.98, 99].

The behavior of the magnetic mirror can be understood by examining a corrugated metal slab, where a series of narrow slots are cut vertically, or a $\lambda/4$ thick dielectric slab backed by a metal plate. In the first case, each slot can be regarded as a parallel-plate transmission line, which runs into the slab and is shorted at the bottom (Fig. 56.22a). If the slots are $\lambda_0/4$ deep, the short circuit at the bottom end is transformed by the length of the slots into an open circuit with zero currents at the top end, and hence exhibits high impedance. In the second case, the total optical path in the dielectric slab upon reflection equals $\lambda/2$, so that the first antinode of the standing wave, where the E-component is at maximum and the H-component is zero, coincides with the top surface of the slab (Fig. 56.22b). A magnetic mirror is the planar analog of the two cases that effectively have been transformed to a much more compact form by virtue of the strong resonant coupling between the top and bottom surfaces – the FSS and metal screen, respectively. Given the analogy, magnetic mirrors act as two-dimensional electric filters that block surface currents and are thus often referred to as high impedance surfaces (HIS) [56.99].

The resonant frequencies of HIS are typically lower than the resonant frequencies of the constituent FSS due to the capacitive coupling of the latter with the metal screen. For many designs of metamolecules, the wavelength of their resonances can be estimated based on the expression for the guided wavelength in a microstrip transmission line [56.100]

$$\lambda_0 \approx \sqrt{\frac{\varepsilon+1}{2} + \frac{\varepsilon-1}{2\sqrt{1+\frac{5d}{w}}}}\,\frac{2l_e}{m}\,, \tag{56.13}$$

where l_e is the total electrical length of the metallic pattern that forms the metamolecule, $2w$ is the width of the metallic strips forming the metamolecule, ε is the relative permittivity of the dielectric spacer, d is the spacer's thickness, and m is a positive integer ($m < 5$). As in the case of (56.1) and (56.12), this formula is practically useful only in the near-IR and lower frequency domains.

HIS have been realized and studied mainly in the RF and microwave domains where, due to their unusual boundary condition, they have been used as a new

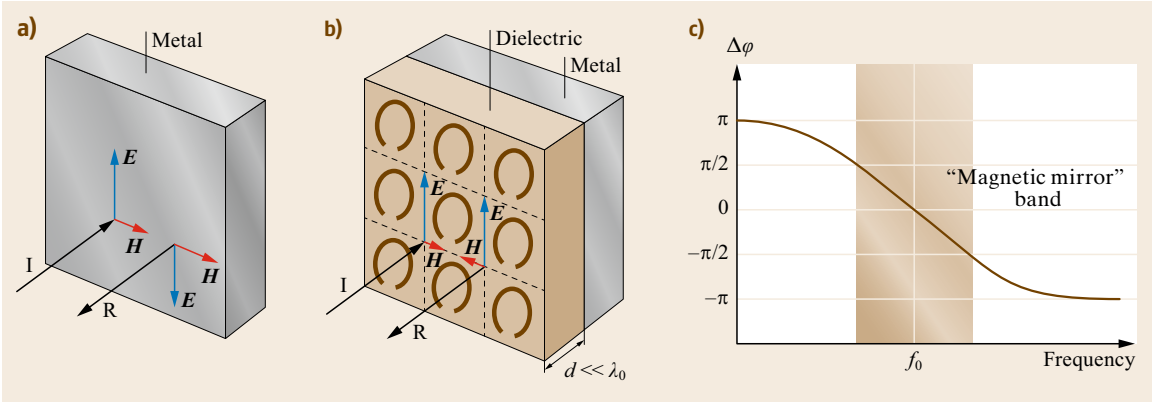


Fig. 56.21 Electromagnetic boundary conditions at the surface of **(a)** a conventional metallic mirror and **(b)** a magnetic mirror. **(c)** Dispersion of the reflection phase schematically shown for a magnetic mirror at its resonance

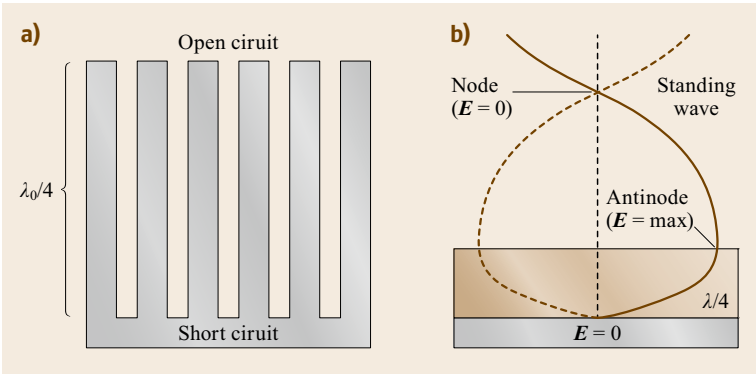


Fig. 56.22a,b Bulk analogs of a magnetic mirror: corrugated metal slab **(a)** and dielectric slab backed by a metal plate **(b)**

type of ground plane for low-profile antennas [56.99, 101, 102]. Indeed, an electric-dipole antenna lying flat against a HIS will not be shorted out (as it happens on a regular metal sheet), while its radiation will appear amplified due to the in-phase oscillations of the antenna's image created by such a mirror. Moreover, since at resonance HIS do not support propagating surface waves, the antenna will radiate efficiently, with a smooth radiation pattern that is free from the effects of multipath interference along the ground plane.

The magnetic mirror effect has been successfully demonstrated also in the visible part of the spectrum [56.103], where one faces challenges in metamaterial fabrication and has to deal with the plasmonic behavior of nanostructured metals and increased ohmic losses. The interest in optical magnetic mirrors is driven by applications in surface-enhanced spectroscopy, and the opportunity to increase the efficiency of optical antennas and nanoemitters, and engineer compact optical cavities. The first metamaterial sample was based on a periodic grid of 40 nm thick meandering aluminum nanowires forming a resonant *fish-scale* pattern, which was placed on top of an aluminum film coated with a 50 nm thick silicon dioxide layer (Fig. 56.23). The resulting nanostructured mirror exhibited high-impedance resonances in the yellow and red spectral bands. In particular, its reflection at 560 and 630 nm was accompanied by changes of the phase as small as $\pi/9$ and $\pi/6$, respectively.

56.2.3 Perfect Metamaterial Absorbers

One of the most promising practical applications of magnetic mirrors has proven to be *resonant loss amplification* (or perfect absorption [56.104]), which benefits from the presence of, otherwise detrimental, resistive losses. This metamaterial functionality exploits the high-impedance resonance where the tangential component of the total electric field on the surface of a magnetic mirror is significant. Consequently, if an electrically resistive sheet (for example, a thin slab of a lossy dielectric) is placed on top of the structure, a part of the incident energy will be dissipated in this slab and, as a result, the reflection coefficient of the system may become significantly reduced. This can be shown quantitatively by using transmission line the-

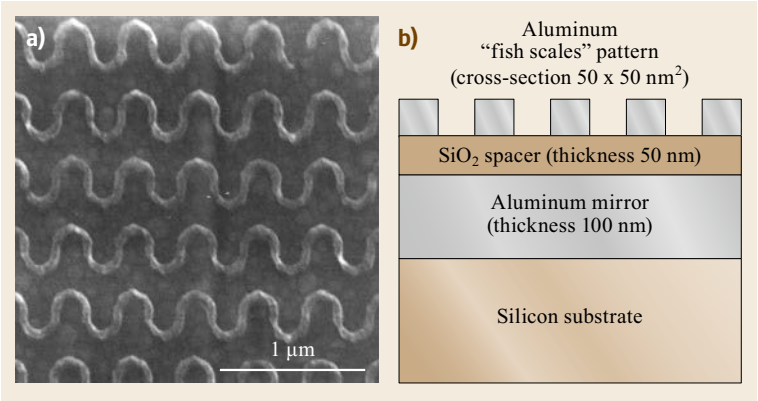


Fig. 56.23a,b First optical magnetic mirror. **(a)** Scanning electron microscope image of the aluminum *fish-scale* nanopattern. (After [56.60, 103]) **(b)** Cross-section of the mirror

ory. In fact, the reflectivity may drop as low as −30 dB ($R = 0.1\%$) [56.105], indicating that the resonant absorption of light by the metamaterial mirror is virtually complete.

The idea of the metamaterial absorbing screen was originally conceived by *Engheta* [56.106] and experimentally demonstrated by *Fedotov* et al. [56.105], who also showed that the source of losses could simply be the dielectric spacer of the magnetic mirror itself. An absorbing magnetic mirror may, therefore, be regarded as an extremely compact electromagnetic cavity, where light bounces back and forth (effectively being trapped between the constituent FSS and metal screen) while passing through the lossy dielectric many times until it is completely absorbed. Such an interpretation is also supported by the results of numerical simulations when presented in the form of field lines of the total Poynting vector, which are often used to visualize the flow of electromagnetic energy near subwavelength scatterers [56.107, 108]. At high-impedance resonance, these lines are seen to wind around the metallic fabric of the metamaterial, going in and out of the dielectric spacer where they eventually terminate (Fig. 56.24). Achieving maximum absorption, however, is possible only for a certain, optimum range of resistive losses (i. e., values of imaginary part of the substrate's permittivity, $\mathrm{Im}\{\varepsilon\}$); the absorption is naturally weak if the losses are too small, and it is also weak if the losses are too high, since large $\mathrm{Im}\{\varepsilon\}$ destroys the magnetic mirror behavior [56.105].

Resonant loss amplification may be used to improve the sensitivity of photodetectors outside their main spectral band by enhancing a small tail of interband absorption. In this case, for example, the photosensitive material could replace the dielectric spacer of the magnetic mirror, while the metamaterial structure itself would act as a light-harvesting device. This approach should be particularly efficient for semiconductor detectors and solar cells, and for increasing the quantum efficiency of photomultipliers.

56.2.4 Designer Metasurfaces and Flat Optics

The resonant response of a planar metamaterial is accompanied by a phase shift of the transmitted wave, which varies between $-\pi/2$ and $+\pi/2$, depending on the detuning from the resonance frequency. Remark-

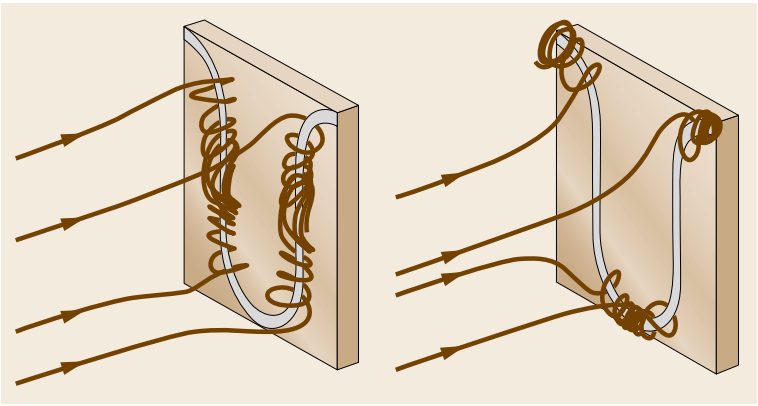


Fig. 56.24 Perfect absorption: lines of the total Poynting vector (*power flow*) calculated for two different absorption resonances of a microwave *fish-scale* magnetic mirror

ably, it occurs on the length scale of one-tenth or even one-hundredth of a wavelength, which renders the phase shift virtually discontinuous. Thus, by varying the size and geometry (and hence the detuning from the resonance frequency) of individual metamolecules along the surface of the metamaterial, one can spatially modulate the phase discontinuity and, in accordance with Huygens's principle, imprint the resulting phase profile on the transmitted beam. The same applies to the reflected beam, which is produced by the coherent superposition of the waves scattered by the individual metamolecules in the backward direction. Given the ever-expanding choice of resonant patterns, metasurfaces with gradient structuring (designer metasurfaces) should enable the shaping of transmitted and reflected wavefronts in nearly arbitrary ways, thus performing as sophisticated but ultimately flat optical devices [56.109]. This is in stark contrast to conventional bulk optical components (e.g., lenses, prisms, wave plates, and curved mirrors), where the phase shifts are accumulated gradually along the optical path, typically over the distance of several wavelengths.

Although the idea of designer metasurfaces can be traced back to earlier works on microwave FSS [56.110, 111], it attracted significant attention only recently following the work of *Capasso* et al. [56.109], who demonstrated wavefront manipulation in the IR and formulated generalized laws of refraction and reflection in the presence of phase discontinuities

$$\sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1 + \frac{\lambda_0}{2\pi\, n_2}\frac{\mathrm{d}\varphi}{\mathrm{d}x}\,, \tag{56.14a}$$

$$\sin\theta_1' = \sin\theta_1 + \frac{\lambda_0}{2\pi\, n_1}\frac{\mathrm{d}\varphi}{\mathrm{d}x}\,, \tag{56.14b}$$

where $\mathrm{d}\varphi/\mathrm{d}x$ is the gradient of the phase discontinuity along the interface (Fig. 56.25). These relations imply that the refracted and reflected beams can have arbitrary directions, provided that suitable $\mathrm{d}\varphi/\mathrm{d}x$ is introduced. In particular, refraction and reflection of light incident at angle θ_1 become negative for phase gradients smaller than $\mathrm{d}\varphi/\mathrm{d}x = -(2\pi/\lambda_0)\, n_1 \sin\theta_1$. While deriving (56.14a)–(56.14b) it was assumed that φ varies continuously along the interface, so that all the incident energy is transferred into the modified reflection and refraction. Since, in practice, a metasurface is a discrete array of scatterers, although with subwavelength separation, there will also be regularly refracted and reflected beams that obey the conventional laws of refraction and reflection [56.112].

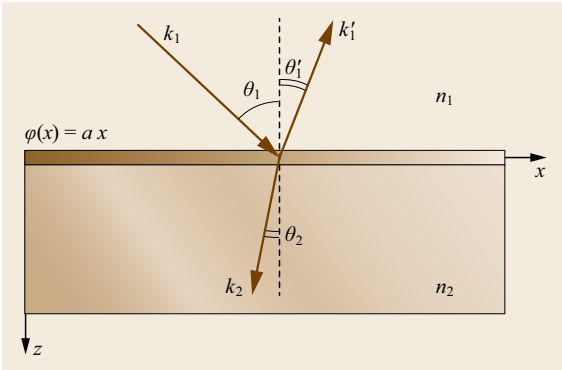


Fig. 56.25 Anomalous refraction and reflection at an interface with linearly varying phase discontinuity $\varphi(x)$

An example of the designer metasurface exhibiting negative refraction and reflection in the IR is presented in Fig. 56.26a. It features a regular array of V-shaped gold metamolecules arranged in the groups of 8, with their shape and size changing gradually in each group in the x-direction. Compared to more simple cut wires, V-shaped metamolecules produce much stronger scattering contributions to modified refraction and reflection for all their permutations and ensure phase shifts in the entire range from 0 to 2π, as required for engineering a continuous phase gradient.

A more complex scattering phenomenon, such as the generation of a vortex beam, is observed when the phase shift varies linearly in the azimuthal direction. This can be achieved with a designer metasurface, for instance, by arranging V-shaped metamolecules of different permutations in eight equal sectors, which imprint constant phase shifts with $\pi/4$ increment counterclockwise (Fig. 56.26b). By re-arranging the metamolecules such that the phase shift is constant in the azimuthal direction but increases hyperbolically in the radial direction, one obtains an aberration-free optical flat lens [56.113].

Very recently, the concept of wavefront manipulation using designer metasurfaces was taken further with the introduction of the so-called Huygens metasurfaces [56.114, 115], planar arrays of metamolecules that react to both E and H components of light. Huygens metasurfaces are expected to have major implications for integrated and transformation optics, as they allow one to control the amplitude and phase of the scattering independently, and hence generate truly arbitrary field patterns for given illumination.

56.2.5 Polarization Manipulation with Planar Metamaterials

Similar to bulk metamaterials, homogeneously patterned planar metamaterials can exhibit complex polarization effects, which lie beyond trivial polarization conversion originating from structural anisotropy [56.116]. These phenomena include, in particular, enhanced con-

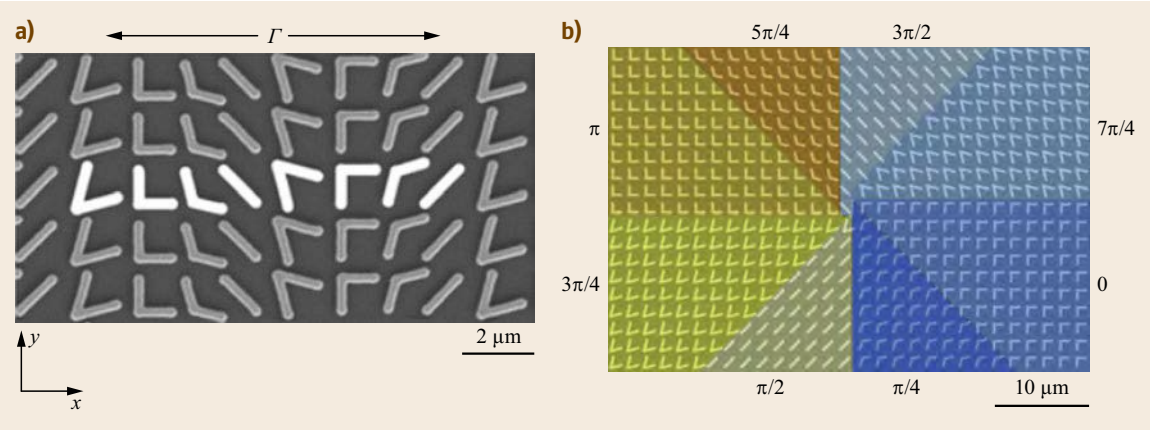


Fig. 56.26 Scanning electron microscope images of IR designer metasurfaces, which exhibit negative refraction and reflection **(a)** and generate vortex beams **(b)**. (After [56.109])

ventional optical activity, optical activity occurring in the absence of structural chirality, and asymmetric transmission [56.117], which, unlike the first two effects, does not seem to have an analog in bulk media.

Natural (Intrinsic) Optical Activity

Natural optical activity manifests itself as the polarization azimuth rotation of linearly polarized light, as well as the preferential absorption of either right or left-hand circularly polarized light in a lossy medium (Sect. 56.1.5). Conventionally, optical activity is attributed to *intrinsic chirality* of matter, a particular type of asymmetry characterized by the presence of a screw-like twist in its very structure (either on molecular or crystal lattice levels). Such a twist exists only in three dimensions and, therefore, cannot be realized in a truly planar structure. Instead, it may be introduced in a bi-layered stack, for example, by rotating metamaterial patterns in the first and second layers with respect to each other, as illustrated in Fig. 56.11c–g. Due to the strong resonant coupling between the mutually twisted patterns, the rotation of the polarization plane of up to 90° can be achieved for interlayer separation of only $\lambda_0/30$ or so [56.33, 34]. Consequently, all bi-layered chiral metamaterials considered in Sect. 56.1.5, although not technically planar, can be regarded as planar metamaterials given their vanishing thickness. Moreover, it is easy to see that even the presence of a supporting substrate makes the structure of a planar metamaterial intrinsically chiral if its pattern features a spiral-like twist. In this case, however, the manifestation of optical activity is usually rather weak [56.33, 40, 118].

Extrinsic Optical Activity

Extrinsic optical activity is a practically unknown instance of the polarization phenomenon, which is very difficult to observe using natural materials but which can be easily engineered in planar metamaterials, offering an efficient and robust polarization control mechanism. It is attributed to the chirality that is drawn extrinsically from the experimental arrangement that includes not only molecules (or metamolecules), which must be *oriented* and can be *structurally achiral*, but also the wave propagation direction. This special condition for observing optical activity was first identified by *Bunn* [56.119] and confirmed in liquid crystals by *Williams* [56.120].

In the context of metamaterial research, the effect was experimentally demonstrated by *Plum* et al. in both the microwave [56.121] and optical [56.122] parts of the spectrum by characterizing the transmission response of specially designed metasurfaces at oblique incidence. The metasurfaces were based on regular arrays of achiral asymmetrically split-ring metamolecules (Fig. 56.27a), although, in principle, any metamaterial pattern that does not possess twofold rotational symmetry could have formed an extrinsically chiral arrangement with an obliquely incident beam (and hence exhibited optical activity). The observed rotation of the polarization azimuth exceeded 80° for microwave and 1° for optical metasurfaces, which in absolute terms surpasses the strength of optical activity achieved in most intrinsically chiral planar metamaterials.

Optical activity of an extrinsically chiral metasurface has the following key features:

1. No effect can be observed at normal incidence.
2. No effect can be observed at oblique incidence if the metamolecules possess twofold rotational symmetry or if the line of their mirror symmetry is parallel to the plane of incidence.

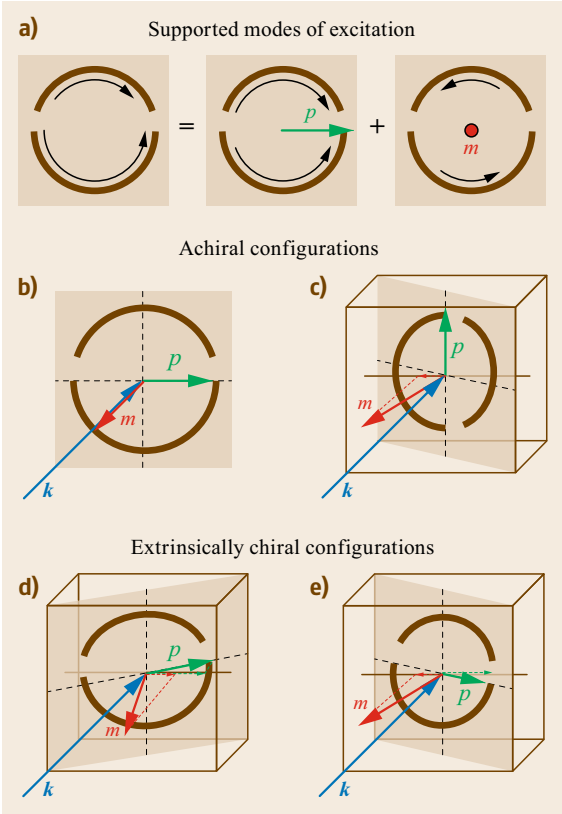


Fig. 56.27a–e Optical activity due to extrinsic chirality. **(a)** Oscillating currents in an asymmetrically split ring can be represented as a sum of symmetric and antisymmetric modes, corresponding to an electric dipole $\boldsymbol{p}$ (*green arrow*) and a magnetic dipole $\boldsymbol{m}$ (*red arrow*). Optical activity is controlled by the projections of $\boldsymbol{p}$ and $\boldsymbol{m}$ onto the plane perpendicular to the incident wavevector $\boldsymbol{k}$ (*green* and *red dashed arrows* correspondingly). The effect is absent either **(b)** at normal incidence, when one of the projections is zero, or **(c)** if the projections are mutually orthogonal. The strongest optical activity of opposite sign occurs if these projections are either **(d)** parallel or **(e)** antiparallel

3. Opposite angles of incidence yield the effects of opposite signs.
4. The magnitude of the polarization rotation and circular dichroism is controlled by the angle of incidence.

Most of the features can be readily understood based on geometrical considerations. Indeed, the first and second features expose mutual orientations of the metasurface and incident beam that result in vanishing extrinsic chirality (Fig. 56.27b,c). The third feature simply reflects the fact that reversing the angle of incidence transforms an experimental arrangement into its mirror form, which gives rise to extrinsic chirality of opposite handedness (see Fig. 56.27d,e). Understanding the fourth feature, however, requires one to revisit the physical mechanism of the effect, as detailed below.

Similar to conventional optical activity, its extrinsic counterpart results from the simultaneous presence of electric and magnetic responses in the metamolecules. The asymmetry of the metamolecules (such as, for example, asymmetrically split rings) plays a key role here: an electromagnetic wave resonantly induces unequal currents oscillating in the upper and lower arcs of the rings. Such a current configuration may be represented as a sum of symmetric and antisymmetric current modes. The symmetric mode yields an electric dipole moment, $\boldsymbol{p}$, oscillating in the plane of the ring and parallel to the split, while the antisymmetric mode gives rise to a magnetic dipole moment, $\boldsymbol{m}$, which oscillates perpendicular to the plane (Fig. 56.27a). At oblique incidence, each dipole moment acquires a component in the plane orthogonal to the wave propagation direction (Fig. 56.27d). When the split is parallel to the plane of incidence (the line of mirror symmetry is orthogonal to the plane of incidence), the projections of $\boldsymbol{m}$ and $\boldsymbol{p}$ become collinear and scatter electromagnetic waves with orthogonal polarizations, so that the polarization of the resulting transmitted wave exhibits maximal rotation. Clearly, by varying the angle of incidence, one changes the difference between the magnitudes of the dipole projections, and hence controls their relative contributions to the polarization state of the transmitted wave. Also, it can be seen that for opposite angles of incidence, the mutual phase difference between the projections of $\boldsymbol{m}$ and $\boldsymbol{p}$ reverses, which directly affects the sign of optical activity (see Fig. 56.27d,e). When the split is perpendicular to the plane of incidence (the line of mirror symmetry is parallel to the plane of incidence), the projections of $\boldsymbol{m}$ and $\boldsymbol{p}$ become orthogonal to each other (Fig. 56.27c). In this case, the electromagnetic waves scattered by the projections have the same polarizations, and so the polarization state of the transmitted wave cannot change. Finally, at normal incidence, the projection of $\boldsymbol{m}$ is zero, which makes the absence of optical activity trivial (Fig. 56.27b).

The mechanism described above equally applies to the reflection off an extrinsically chiral metasurface, where it leads to strong *specular optical activity* (the existence of this effect in bulk media is still being debated) [56.123]. Intriguingly, this reflection effect mimics the longitudinal magneto-optical Kerr effect, but obviously does not require a magnetic field for its observation.

Even though extrinsic chirality can only be ascribed to an *anisotropic* metasurface, its response features a spectral band where the manifestation of the

anisotropy (i. e., linear birefringence) is negligible. For this band the incident polarization state remains linear upon transmission/reflection, exhibiting pure azimuth rotation [56.121, 122]. Importantly, extrinsic optical activity is exhibited by truly planar metamaterials, which are typically much easier to fabricate than the bi-layered stacks. The scalability of the metamaterial designs and inherent tunability of the phenomenon, the magnitude (and sign) of which can be continuously controlled by tilting the metasurface with respect to the incident beam, enables the exploitation of extrinsic optical activity in polarization control applications both in microwave and photonic devices. Examples of the latter may include compact and efficient polarization rotators, circular polarizers, and polarization modulators, as well as vibration sensors.

Asymmetric Transmission

Asymmetric transmission is a new fundamental phenomenon of electromagnetism, which manifests itself as a difference in the overall intensity of circularly polarized waves transmitted by a medium in the opposite directions. In some ways, it resembles the famous nonreciprocity of the Faraday effect in magnetized media but requires no magnetic field for its observation and is fully compliant with the Lorentz reciprocity principle. Asymmetric transmission was discovered in 2006 by *Fedotov* et al. in planar chiral metamaterials [56.124], a new distinct class of metasurfaces that feature *two-dimensional* (2-D) *chirality* in the structure of their metamolecules.

2-D chirality implies that a *planar* object or a pattern cannot be brought into congruence with its mirror image (which is obtained by reflection across a line in the object's plane) unless it is lifted off the plane. The simplest example of a planar chiral object is the Archimedean spiral. 2-D chirality has been rarely acknowledged in the context of optics and materials research simply because natural planar materials were not known to men (at least until 2004, when graphene sheets were physically isolated and studied [56.125]). The study of optical implications of 2-D chirality was pioneered by *Zheludev*, who introduced the notion of 2-D chirality (termed as *planar chirality*) to the field of metamaterials research in 2003 [56.126].

While conventional chirality reveals a structural twist in three dimensions, planar chirality does so in two dimensions. It is convenient to associate the twist of a planar chiral pattern with a special unitary vector, $\boldsymbol{W}$, known as the *twist vector*. It points along the normal to the plane of the pattern, and its direction is governed by the corkscrew law: if the screw rotates in the direction of the structural twist, then it moves along the twist vector. Correspondingly, if $\boldsymbol{W}$ points along the direction of observation, the pattern is perceived as being twisted clockwise and, conversely, if $\boldsymbol{W}$ points toward the observer, the pattern is twisted counter-clockwise (Fig. 56.28). Importantly, the sense of the twist (and hence the direction of $\boldsymbol{W}$) reverses when a planar chiral pattern is observed from opposite sides, and this is exactly what makes the engineering of the asymmetric transmission effect possible.

Phenomenologically, the origin of the effect may be traced to a particular structure of the complex matrix that describes the transmission of circularly polarized waves through a planar chiral metamaterial

$$\boldsymbol{\chi}^{2\mathrm{D}} = \begin{pmatrix} t_{++} & t_{+-} \\ t_{-+} & t_{--} \end{pmatrix} , \tag{56.15}$$

which is a non-Hermitian matrix with equal diagonal elements ($t_{++} = t_{--}$). When expressed in terms of transmission coefficients in the linear (Cartesian) basis, this matrix can be transformed into a sum [56.124]

$$\boldsymbol{\chi}^{2\mathrm{D}} = \boldsymbol{\chi}_0^{2\mathrm{D}} + \mathrm{i}\, t_{xy}\, \mathrm{sgn}\,(\boldsymbol{k}\cdot\boldsymbol{W})\,\hat{\boldsymbol{g}} , \tag{56.16}$$

where $\boldsymbol{k}$ is the wavevector of the incident wave, t_{xy} is the coefficient characterizing the conversion of a lin-

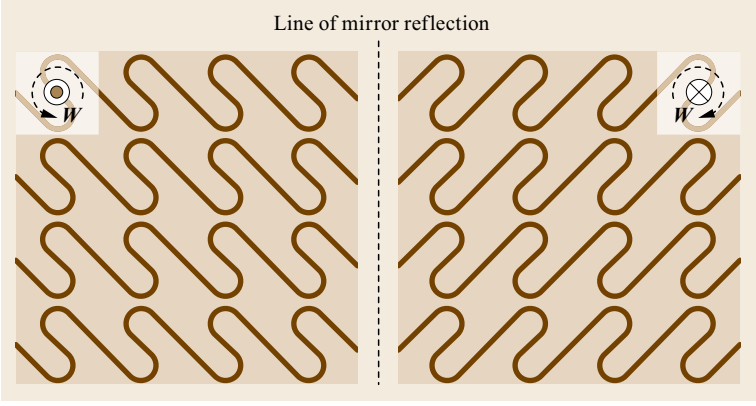


Fig. 56.28 Examples of planar chiral metamaterial patterns featuring clockwise and counter-clockwise twists

early polarized state into its orthogonal counterpart upon transmission,

$$\chi_0^{2D} = \frac{1}{2}\begin{pmatrix} t_{xx}+t_{yy} & t_{xx}-t_{yy} \\ t_{xx}-t_{yy} & t_{xx}+t_{yy} \end{pmatrix} \quad (56.17a)$$

is a symmetric matrix that describes the anisotropy of the chiral pattern, and

$$\hat{\mathbf{g}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (56.17b)$$

is a unitary antisymmetric matrix. The antisymmetric part of χ^{2D} is proportional to the pseudoscalar combination $(\boldsymbol{k}\cdot\boldsymbol{W})$ and is, therefore, responsible for direction-dependent transmission, as the latter changes sign upon the reversal of the wave propagation direction. The antisymmetric part is also proportional to t_{xy}, which can only exist for anisotropic metamolecules of low-structural symmetry and vanishes for patterns possessing a fourfold symmetry axis. The polarization conversion term t_{xy} cannot be eliminated by the choice of appropriate coordinate systems only in the presence of dissipation. In other words, *the asymmetric transmission effect can be observed in a planar chiral metamaterial if, and only if, it is simultaneously anisotropic and lossy*. It also follows form (56.16) that the asymmetry is reversed for the opposite handedness of (i) incident polarization and (ii) structural twist.

The transmission asymmetry exhibited by planar chiral metamaterials is different from that of the Faraday effect in magnetized media. Indeed, in the case of the Faraday effect, the asymmetry applies to the transmission amplitude and phase of an incident circularly polarized wave itself, whereas planar chiral metamaterials, due to their anisotropy, partially convert the incident wave into one of the opposite handedness, and it is the efficiency of this conversion that is asymmetric with respect to the direction of propagation. The fundamental novelty of the asymmetric transmission effect may also be appreciated by comparing the polarization eigenstates (i. e., the polarizations that are preserved upon transmission) of planar chiral metamaterials with that of the Faraday media. The eigenstates of the transparent Faraday media, in general, are two elliptical polarizations of opposite handedness, while the eigenstates of planar chiral metamaterials are two elliptical polarizations of the same handedness, which differ only in the azimuths of their main axes – they are orthogonal (Fig. 56.29). The eigenstates of planar chiral metama-

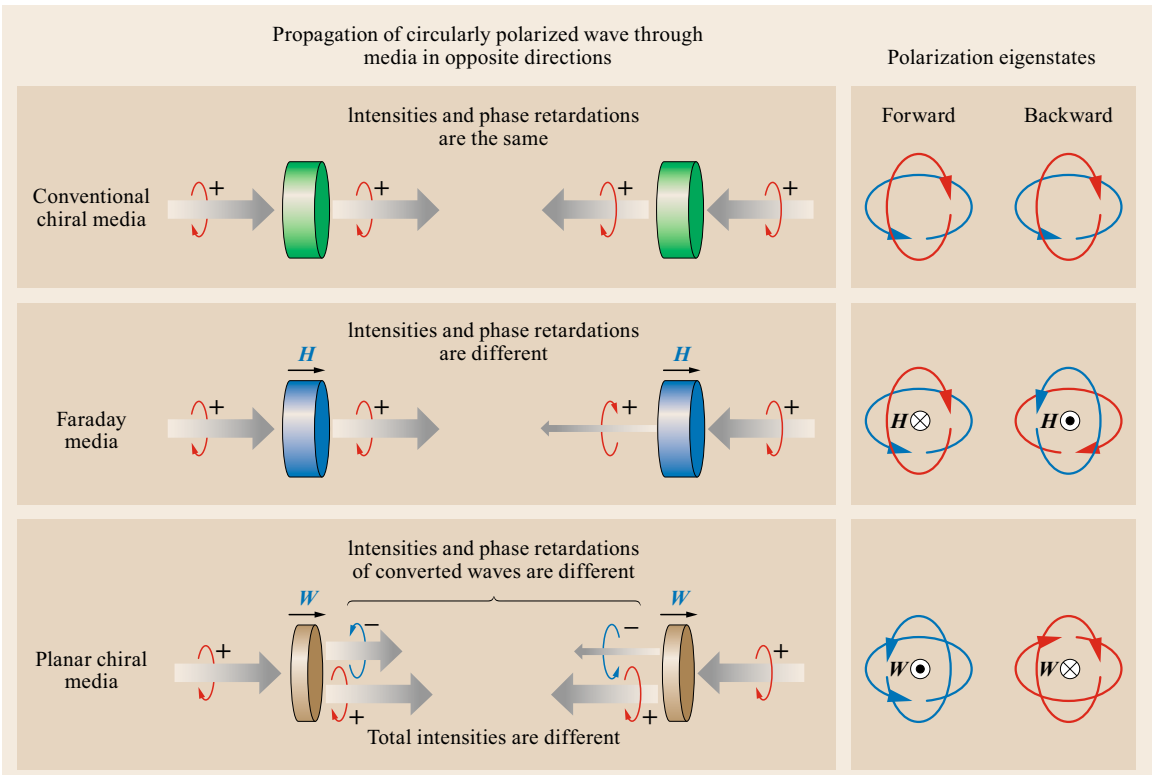


Fig. 56.29 Manifestations of polarization effects and corresponding polarization eigenstates in conventional chiral media, Faraday media, and planar chiral media. Vectors $\boldsymbol{H}$ and $\boldsymbol{W}$ denote magnetization and planar chiral twist, respectively

terials are also different from those of conventionally chiral media (and chiral metamaterials) exhibiting optical activity. The eigenstates of the latter are completely symmetric with respect to the propagation direction, whereas the eigenstates of planar chiral metamaterials reverse their handedness (Fig. 56.29).

At an abstract crystallographic level, it has been shown that a combination of birefringence and circular dichroism (termed *elliptical dichroism*) provides a basis for the optical manifestation of asymmetric transmission and that a hypothetical bi-axial elliptically dichroic bulk medium should exhibit optical properties characteristic to a planar chiral metamaterial [56.127]. More rigorous, quantitative descriptions of asymmetric transmission with a detailed analysis of the effect on a microscopic level and systematic study of planar chiral metamolecules have also been developed [56.128–131].

In practical terms, the asymmetric transmission is a strong and scalable phenomenon, which has been experimentally demonstrated in the microwave [56.124, 132], terahertz [56.133], and optical [56.134, 135] parts of the spectrum, with the observed absolute transmission asymmetry exceeding 15%, 6%, and 4%, respectively. This invites the application of planar chiral metamaterials as novel polarization sensitive components of electromagnetic devices, such as polarization and direction sensitive beam splitters, circulators, and sensor components. For example, when placed into a ring resonator, a planar chiral metamaterial will enforce generation of left and right-handed elliptically polarized waves in the opposite directions of circulation.

References

56.1 R.A. Shelby, D.R. Smith, S. Schultz: Science **292**, 77 (2001)
56.2 V. Browning, S. Wolf: DARPA Multi-University Research Initiative "Scalable and Reconfigurable Electromagnetic Metamaterials and Devices" (2001) http://xlab.me.berkeley.edu/MURI/muri.html
56.3 R.W. Ziolkowski: EPJ Appl. Metamater. **1**, 5 (2014)
56.4 T. Kaelberer, V.A. Fedotov, N. Papasimakis, D.P. Tsai, N.I. Zheludev: Science **330**, 1510 (2010)
56.5 M.C.K. Wiltshire, J.B. Pendry, I.R. Young, D.J. Larkman, D.J. Gilderdale, J.V. Hajnal: Science **291**, 849 (2001)
56.6 C. Kurter, J. Abrahams, S.M. Anlage: Appl. Phys. Lett. **96**, 253504 (2010)
56.7 J. Zhou, Th. Koschny, M. Kafesaki, E.N. Economou, J.B. Pendry, C.M. Soukoulis: Phys. Rev. Lett. **95**, 223902 (2005)
56.8 J.C. Bose: Proc. R. Soc. London **63**, 146 (1898)
56.9 J.B. Pendry, A.J. Holden, D.J. Robbins, W.J. Stewart: IEEE Trans. Microw. Theory Tech. **47**, 2075 (1999)
56.10 V.M. Shalaev: Nature Photon. **1**, 41 (2007)
56.11 V.M. Shalaev, W. Cai, U.K. Chettiar, H.-K. Yuan, A.K. Sarychev, V.P. Drachev, A.V. Kildishev: Opt. Lett. **30**, 3356 (2005)
56.12 A.N. Grigorenko, A.K. Geim, H.F. Gleeson, Y. Zhang, A.A. Firsov, I.Y. Khrushchev, J. Petrovic: Nature **438**, 17 (2005)
56.13 A. Schuster: *An Introduction to the Theory of Optics* (Arnold, London 1904)
56.14 D.V. Sivukhin: Opt. Spektrosk. **3**, 308 (1957)
56.15 V.E. Pafomov: Sov. Phys. JETP **9**, 1321 (1959)
56.16 V.G. Veselago: Sov. Phys. USPEKHI **10**, 509 (1968)
56.17 V.G. Veselago, E.E. Narimanov: Nat. Mat. **5**, 759 (2006)
56.18 D.R. Smith, W.J. Padilla, D.C. Vier, S.C. Nemat-Nasser, S. Schultz: Phys. Rev. Lett. **84**, 4184 (2000)
56.19 L.D. Landau, E.M. Lifschitz, L.P. Pitaevskii: *Electrodynamics of Continuous Media – Course of Theoretical Physics*, Vol. 8, 2nd edn. (Pergamon, Oxford 1984)
56.20 L.I. Mandelshtam: *Complete Collected Works* (Akad. Nauk SSSR, Moscow 1947)
56.21 J.B. Pendry: Phys. Rev. Lett. **85**, 3966 (2000)
56.22 M. Valanju, R.M. Walser, A.P. Valanju: Phys. Rev. Lett. **88**, 187401 (2002)
56.23 N. Garcia, M. Nieto-Vesperinas: Opt. Lett. **27**, 885 (2002)
56.24 J.B. Pendry, D.R. Smith: Phys. Rev. Lett. **90**, 029703 (2003)
56.25 P.M. Valanju, R.M. Walser, A.P. Valanju: Phys. Rev. Lett. **90**, 029704 (2003)
56.26 J. Pacheco, T.M. Gregorczyk, B.-I. Wu, Y. Zhang, J.A. Kong: Phys. Rev. Lett. **89**, 257401 (2002)
56.27 S. Foteinopoulou, E.N. Economou, C.M. Soukoulis: Phys. Rev. Lett. **90**, 107402 (2003)
56.28 P.F. Loschialpo, D.L. Smith, D.W. Forester, F.J. Rachford: Phys. Rev. E **67**, 025602 (2003)
56.29 K.Y. Bliokh, Y.P. Bliokh: Phys.-USPEKHI **47**, 393 (2004)
56.30 D.R. Smith, J.B. Pendry, M.C.K. Wiltshire: Science **305**, 788 (2004)
56.31 Lord Kelvin: *The Molecular Tactics of a Crystal* (Clarendon, Oxford 1894) p. 27
56.32 P.G. de Gennes: *The Physics of Liquid Crystals* (Clarendon, Oxford 1974)
56.33 A.V. Rogacheva, V.A. Fedotov, A.S. Schwanecke, N.I. Zheludev: Phys. Rev. Lett. **97**, 177401 (2006)
56.34 Y. Ye, S. He: Appl. Phys. Lett. **96**, 203501 (2010)
56.35 G. Kenanakis, R. Zhao, A. Stavrinidis, G. Konstantinidis, N. Katsarakis, M. Kafesaki, C.M. Soukoulis, E.N. Economou: Opt. Mater. Express **2**, 1702 (2012)
56.36 K.F. Lindman: Ann. Phys. **63**, 621 (1920)
56.37 I. Tinoco, M.P. Freeman: J. Phys. Chem. **61**, 1196 (1957)

56.38 J.K. Gansel, M. Thiel, M.S. Rill, M. Decker, K. Bade, V. Saile, G. von Freymann, S. Linden, M. Wegener: Science **325**, 1513 (2009)
56.39 Y. Svirko, N. Zheludev, M. Osipov: Appl. Phys. Lett. **78**, 498 (2001)
56.40 E. Plum, V.A. Fedotov, A.S. Schwanecke, N.I. Zheludev, Y. Chen: Appl. Phys. Lett. **90**, 223113 (2007)
56.41 M. Decker, M. Ruther, C.E. Kriegler, J. Zhou, C.M. Soukoulis, S. Linden, M. Wegener: Opt. Lett. **34**, 2501 (2009)
56.42 K. Hannam, D.A. Powell, I.V. Shadrivov, Y.S. Kivshar: Appl. Phys. Lett. **102**, 201121 (2013)
56.43 M. Decker, M.W. Klein, M. Wegener, S. Linden: Opt. Lett. **32**, 856 (2007)
56.44 I.V. Lindell, A.J. Viitanen: IEEE Trans. Antennas Propagat. **40**, 91 (1992)
56.45 D.L. Jaggard, N. Engheta: Electron. Lett. **25**, 173 (1989)
56.46 S. Tretyakov, A. Serdyukov, I. Semchenko, A. Sihvola: *Electromagnetics of Bi-anisotropic Materials: Theory and Applications* (Gordon Breach Science, London 2001)
56.47 S. Tretyakov, I. Nefedov, A. Sihvola, S. Maslovski, C. Simovski: J. Electromagn. Waves Appl. **17**, 695 (2003)
56.48 J.B. Pendry: Science **306**, 1353 (2004)
56.49 E. Plum, J. Zhou, J. Dong, V.A. Fedotov, T. Koschny, C.M. Soukoulis, N.I. Zheludev: Phys. Rev. B **79**, 035407 (2009)
56.50 S. Zhang, Y.-S. Park, J. Li, X. Lu, W. Zhang, X. Zhang: Phys. Rev. Lett. **102**, 023901 (2009)
56.51 U. Leonhardt: Science **312**, 1777 (2006)
56.52 J.B. Pendry, D. Schurig, D.R. Smith: Science **312**, 1780 (2006)
56.53 G.W. Milton, M. Briane, J.R. Willis: New J. Phys. **8**, 248 (2006)
56.54 L.S. Dolin: Izv. Vuz. Radiofiz. **4**, 964 (1961)
56.55 H. Chen, C.T. Chan, P. Sheng: Nat. Mater. **9**, 387 (2010)
56.56 D. Schurig, J.J. Mock, B.J. Justice, S.A. Cummer, J.B. Pendry, A.F. Starr, D.R. Smith: Science **314**, 977 (2006)
56.57 J. Li, J.B. Pendry: Phys. Rev. Lett. **101**, 203901 (2008)
56.58 R. Liu, C. Ji, J.J. Mock, J.Y. Chin, T.J. Cui, D.R. Smith: Science **323**, 366 (2009)
56.59 T. Ergin, N. Stenger, P. Brenner, J.B. Pendry, M. Wegener: Science **328**, 337 (2010)
56.60 M. Gharghi, C. Gladden, T. Zentgraf, Y. Liu, X. Yin, J. Valentine, X. Zhang: Nano Lett. **11**, 2825 (2011)
56.61 A. Poddubny, I. Iorsh, P. Belov, Y. Kivshar: Nat. Photon. **7**, 958 (2013)
56.62 D. Schurig, D.R. Smith: Appl. Phys. Lett. **82**, 2215 (2003)
56.63 Z. Ma, P. Wang, Y. Cao, H. Tang, H. Ming: Appl. Phys. B **84**, 261 (2006)
56.64 J. Rho, Z. Ye, Y. Xiong, X. Yin, Z. Liu, H. Choi, G. Bartal, X. Zhang: Nat. Commun. **1**, 143 (2010)
56.65 P.A. Belov: Microw. Opt. Technol. Lett. **37**, 259 (2003)
56.66 D.R. Smith, P. Kolinko, D. Schurig: J. Opt. Soc. Am. B **21**, 1032 (2004)
56.67 Z. Jacob, L.V. Alekseyev, E. Narimanov: Opt. Express **14**, 8247 (2006)
56.68 P. Ikonen, C. Simovski, S. Tretyakov, P. Belov, Y. Hao: Appl. Phys. Lett. **91**, 104102 (2007)
56.69 I.I. Smolyaninov, Y.-J. Hung, C.C. Davis: Science **315**, 1699 (2007)
56.70 J. Sun, M.I. Shalaev, N.M. Litchinitser: Nat. Commun. **6**, 7201 (2015)
56.71 Z. Liu, H. Lee, Y. Xiong, C. Sun, X. Zhang: Science **315**, 1686 (2007)
56.72 B. Lounis, M. Orrit: Rep. Prog. Phys. **68**, 1129 (2005)
56.73 H.N.S. Krishnamoorthy, Z. Jacob, E. Narimanov, I. Kretzschmar, V.M. Menon: Science **336**, 205 (2012)
56.74 M.A. Noginov, H. Li, Y.A. Barnakov, D. Dryden, G. Nataraj, G. Zhu, C.E. Bonner, M. Mayy, Z. Jacob, E.E. Narimanov: Opt. Lett. **35**, 1863 (2010)
56.75 J. Kim, V.P. Drachev, Z. Jacob, G.V. Naik, A. Boltasseva, E.E. Narimanov, V.M. Shalaev: Opt. Express **20**, 8100 (2012)
56.76 C.L. Cortes, W. Newman, S. Molesky, Z. Jacob: J. Opt. **14**, 063001 (2012)
56.77 B.A. Munk: *Frequency Selective Surfaces: Theory and Design* (Wiley, New York 2000)
56.78 J.C. Vardaxoglou: *Frequency Selective Surfaces: Analysis and Design* (Research Studies Press, Baldock 1997)
56.79 N. Amitay, V. Galindo, C.P. Wu: *Theory and Analysis of Phased Array Antennas* (Interscience, New York 1972)
56.80 H.A. Bethe: Phys. Rev. **66**, 163 (1944)
56.81 V.A. Fedotov, M. Rose, S.L. Prosvirnin, N. Papasimakis, N.I. Zheludev: Phys. Rev. Lett. **99**, 147401 (2007)
56.82 N. Papasimakis, V.A. Fedotov, N.I. Zheludev: Phys. Rev. Lett. **101**, 253903 (2008)
56.83 N. Papasimakis, N.I. Zheludev: Opt. Photon. News **20**, 22 (2009)
56.84 J.E. Davis: Infrared Phys. **20**, 287 (1980)
56.85 C.M. Rhoads, E.K. Damon, B.A. Munk: Appl. Opt. **21**, 2814 (1982)
56.86 K.J. Kogler, R.G. Pastor: Appl. Opt. **27**, 18 (1988)
56.87 K.D. Möller, J.B. Warren, J.B. Heaney, C. Kotecki: Appl. Opt. **35**, 6210 (1996)
56.88 R. Ulrich, T.J. Bridges, M.A. Pollack: Appl. Opt. **9**, 2511 (1970)
56.89 D.A. Weitz, W.J. Skocpol, M. Tinkham: Opt. Lett. **3**, 13 (1978)
56.90 J. LeGall, M. Olivier, J.J. Greffet: Phys. Rev. B **55**, 10105 (1997)
56.91 I. Pusçasu, W.L. Schaich, G. Boreman: Infrared Phys. Technol. **43**, 101 (2002)
56.92 A.P. Raman, M.A. Anoma, L. Zhu, E. Rephaeli, S. Fan: Nature **515**, 540 (2014)
56.93 J. Meléndez, A.J. de Castro, F. López: J. Meneses: Sens. Actuators A **47**, 417 (1995)
56.94 J. Zhao, C. Zhang, P.V. Braun, H. Giessen: Adv. Mater. **24**, OP247 (2012)
56.95 V. Savinov, V.A. Fedotov, P.A.J. de Groot, N.I. Zheludev: Supercond. Sci. Technol. **26**, 084001 (2013)
56.96 N.I. Zheludev, S.L. Prosvirnin, N. Papasimakis, V.A. Fedotov: Nat. Photon. **2**, 351 (2008)

56.97 G. Adamo, J.Y. Ou, J.K. So, S.D. Jenkins, F. de Angelis, K.F. MacDonald, E. Di Fabrizio, J. Ruostekoski, N.I. Zheludev: Phys. Rev. Lett. **109**, 217401 (2012)
56.98 K.P. Ma, K. Hirose, F.R. Yang, Y. Qian, T. Itoh: Electron. Lett. **34**, 2041 (1998)
56.99 D. Sievenpiper, L. Zhang, R.F.J. Broas, N.G. Alexopolous, E. Yablonovitch: IEEE Trans. Microw. Theory Tech. **47**, 2059 (1999)
56.100 M.V. Schneider: Bell Syst. Tech. J. **5**, 1421 (1969)
56.101 R.C. Hansen: IEEE Antennas Wirel. Propagat. Lett. **1**, 46 (2002)
56.102 Y. Zhang, J. von Hagen, W. Wiesbeck: Microw. Opt. Technol. Lett. **35**, 172 (2002)
56.103 A.S. Schwanecke, V.A. Fedotov, V.V. Khardikov, S.L. Prosvirnin, Y. Chen, N.I. Zheludev: J. Opt. A **9**, L1 (2007)
56.104 C.M. Watts, X. Liu, W.J. Padilla: Adv. Mater. **24**, OP98 (2012)
56.105 V.A. Fedotov, P.L. Mladyonov, S.L. Prosvirnin, N.I. Zheludev: Phys. Rev. E **72**, 056613 (2005)
56.106 N. Engheta: Thin absorbing screens using metamaterial surfaces. In: *Dig. 2002 IEEE Ant. Prop. Soc. Symp., San Antonio*, Vol. 2 (2002) p. 392
56.107 C.F. Bohren, D.R. Huffman: *Absorption and Scattering of Light by Small Particles* (Wiley, New York 1983)
56.108 M.V. Bashevoy, V.A. Fedotov, N.I. Zheludev: Opt. Express **13**, 8372 (2005)
56.109 N. Yu, F. Capasso: Nat. Mater. **13**, 139 (2014)
56.110 S.C. Chandran, B. Little, J.C. Vardaxoglou: Phasing effects of cascaded and spatially varying planar arrays. In: *Proc. IEE Int. Conf. Antennas and Propagat. (ICAP), Heriot-Watt* (1993) p. 265
56.111 S. Chadran, J.C. Vardaxoglou: Microw. Opt. Tech. Lett. **6**, 339 (1993)
56.112 N. Yu, P. Genevet, M.A. Kats, F. Aieta, J.-P. Tetienne, F. Capasso, Z. Gaburro: Science **334**, 333 (2011)
56.113 F. Aieta, P. Genevet, M.A. Kats, N. Yu, R. Blanchard, Z. Gaburro, F. Capasso: Nano Lett. **12**, 4932 (2012)
56.114 C. Pfeiffer, A. Grbic: Phys. Rev. Lett. **110**, 197401 (2013)
56.115 M. Kim, A.M. Wong, G.V. Eleftheriades: Phys. Rev. X **4**, 041042 (2014)
56.116 J. Hao, Y. Yuan, L. Ran, T. Jiang, J.A. Kong, C.T. Chan, L. Zhou: Phys. Rev. Lett. **99**, 063908 (2007)
56.117 A.A. Maradudin: *Structured Surfaces as Optical Metamaterials* (Cambridge Univ. Press, Cambridge 2011) p. 94
56.118 M. Decker, M. Klein, M. Wegener, S. Linden: Opt. Lett. **32**, 856 (2007)
56.119 C.W. Bunn: *Chemical Crystallography* (Oxford Univ. Press, New York 1945) p. 88
56.120 R. Williams: Phys. Rev. Lett. **21**, 342 (1968)
56.121 E. Plum, V.A. Fedotov, N.I. Zheludev: Appl. Phys. Lett. **93**, 191911 (2008)
56.122 E. Plum, X.-X. Liu, V.A. Fedotov, Y. Chen, D.P. Tsai, N.I. Zheludev: Phys. Rev. Lett. **102**, 113902 (2009)
56.123 E. Plum, V.A. Fedotov, N.I. Zheludev: Appl. Phys. Lett. **108**, 141905 (2016)
56.124 V.A. Fedotov, P.L. Mladyonov, S.L. Prosvirnin, A.V. Rogacheva, Y. Chen, N.I. Zheludev: Phys. Rev. Lett. **97**, 167401 (2006)
56.125 K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, Y. Zhang, S.V. Dubonos, I.V. Grigorieva, A.A. Firsov: Science **306**, 666 (2004)
56.126 A. Papakostas, A. Potts, D.M. Bagnall, S.I. Prosvirnin, H.J. Coles, N.I. Zheludev: Phys. Rev. Lett. **90**, 107404 (2003)
56.127 S.V. Zhukovsky, A.V. Novitsky, V.M. Galynsky: Opt. Lett. **34**, 1988 (2009)
56.128 J. Petschulat, A. Chipouline, A. Tünnermann, T. Pertsch: Phys. Rev. B **82**, 075102 (2010)
56.129 D.N. Chigrin, C. Kremers, S.V. Zhukovsky: Appl. Phys. B **105**, 81 (2011)
56.130 E. Plum, V.A. Fedotov, N.I. Zheludev: J. Opt. **13**, 024006 (2011)
56.131 A.V. Novitsky, V.M. Galynsky, S.V. Zhukovsky: Phys. Rev. B **86**, 075138 (2012)
56.132 E. Plum, V.A. Fedotov, N.I. Zheludev: Appl. Phys. Lett. **94**, 131901 (2009)
56.133 R. Singh, E. Plum, C. Menzel, C. Rockstuhl, A.K. Azad, R.A. Cheville, F. Lederer, W. Zhang, N.I. Zheludev: Phys. Rev. B **80**, 153104 (2009)
56.134 A.S. Schwanecke, V.A. Fedotov, V.V. Khardikov, S.L. Prosvirnin, Y. Chen, N.I. Zheludev: Nano Lett. **8**, 2940 (2008)
56.135 A. Drezet, C. Genet, J.-Y. Laluet, T.W. Ebbesen: Opt. Express **16**, 12559 (2008)

57. Thermoelectric Materials

Donald T. Morelli

Thermoelectricity is one of the oldest phenomena to be observed in semiconductors, with discovery of the various thermoelectric effects dating back to the early part of the 19th century. These effects manifest themselves as the appearance of a voltage in a circuit comprised of two different conductors due to a temperature difference (Seebeck effect), or as the absorption and evolution of heat at the junction of two different materials under electrical current excitation (Peltier effect). These effects can be utilized in devices to generate electrical power from waste heat or to provide solid state cooling, respectively.

This chapter reviews the main factors governing thermoelectric effects in solids, and how these factors may be manipulated to produce materials with high thermoelectric figure of merit. The first portion of the chapter covers the main features that determine electrical and thermal transport in crystalline semiconductors, while the latter portion discusses several new approaches to this old problem that hold promise for highly efficient thermoelectric materials in the future.

57.1 **Overview of the Field** 1380
57.2 **Semiconductors as Thermoelectric Materials** 1381
57.2.1 Electronic Properties of Charge Carriers in Semiconductors 1381
57.2.2 Semiconductors with Low Intrinsic Thermal Conductivity 1383
57.2.3 Semiconductor Solid Solutions 1384
57.3 **New Concepts in Thermoelectric Materials Design** 1385
57.3.1 Phonon-Glass Electron Crystal: Skutterudites 1385
57.3.2 Bulk Nanostructured Thermoelectrics: Pb-Chalcogenide Nanocomposites 1387
57.3.3 Crystals with Large Anharmonicity: Tetrahedrites 1388
57.4 **Summary and Future Outlook** 1389
References 1390

Thermoelectricity encompasses the collective effects that involve the conversion of a temperature difference to electricity and vice versa. The thermoelectric process is fundamentally a microscopic one involving the transport and exchange of energy by and between electrons and lattice vibrations, or phonons, in a solid. This process can be described in terms of the Seebeck, Peltier, and Thomson coefficients. As such these effects may be implemented to generate electrical power from primary or waste sources of heat, or inversely to provide all-solid-state heat pumping, i. e., climate control, under electrical activation. The extent to which thermoelectricity might impact our utilization of energy in the United States is provided by an examination of the flow and usage of energy in the transportation, industrial, commercial, and residential sectors in the United States, as provided by *Lawrence Livermore Laboratory*, *United States Department of Energy* [57.1]. These data indicate that almost two-thirds of the energy produced is rejected mainly in the form of waste heat. This is not really surprising to us, as we all learn at an early age to keep well away from the exhaust pipe of even an idling automobile and to never open the cap on the radiator of a vehicle whose engine is running; and we all have witnessed the great clouds of bellowing steam emerging from the stack of a nuclear- or coal-burning power plant. This rejected energy represents a tremendous opportunity – in some sense, it is like a new energy source, prodigious in size, available almost for *free* and waiting to be used. To appreciate the size of this energy source, the energy content available in this pathway is equivalent to that in over 400 billion gallons of gasoline [57.2]. It is here, in the recovery of wasted thermal energy, where thermoelectricity can play a pivotal role. Conversion of this heat to electricity, even at modest fractions, can improve significantly the overall efficiency of energy utilization in the US and worldwide, and thereby both reduce dependence on foreign sources of fossil fuels and significantly decrease carbon emissions. It is prin-

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_57

cipally for this reason that the field of thermoelectricity, which has always been of interest scientifically if not technologically, has grown substantially in the last ten years.

57.1 Overview of the Field

Thermoelectric energy conversion systems use the thermoelectric effect to convert heat, either from a primary source such as a burner, or waste heat, such as that in the exhaust stream of an automobile or power plant, to useful electricity. The efficiency for conversion of heat to electricity is determined by the dimensionless figure of merit $zT = \alpha^2\sigma T/\kappa$, where α is the Seebeck coefficient, σ the electrical conductivity, κ the thermal conductivity of the materials comprising the thermoelectric device, and T the absolute temperature. For a thermoelectric device operating between a heat source at temperature T_H and a heat sink at temperature T_C, the efficiency is given by [57.3]

$$\varepsilon = \frac{\Delta T}{T_H}\left(\frac{\sqrt{1+zT_{ave}}-1}{\sqrt{1+zT_{ave}}+\frac{T_C}{T_H}}\right), \qquad (57.1)$$

where $T_{ave} = (T_H + T_C)/2$. Figure 57.1 is a surface plot of this relationship for $T_H = 700\,^\circ\mathrm{C}$ (973 K) as a function of T_C and zT. We see that a conversion efficiency exceeding 25% under these conditions can be achieved over a wide range of figures of merit and cold-side temperatures.

Thermoelectric materials research dates back to the time of *Seebeck* himself, who studied the effect that bears his name in a large number of metals and minerals [57.4]. Subsequent studies were carried out by *Haken* [57.5] and *Altenkirch* [57.6] in Germany in the early part of the 20th century. The modern era of thermoelectricity dates back to the work of *Ioffe* (Institute of Semiconductors, USSR) [57.7] and *Goldsmid* (General Electric Research Laboratories, Great Britain) [57.3]. Through combined experimental and theoretical work it was well established by 1970 that the best thermoelectric materials, such as PbTe and Bi_2Te_3 with $zT \approx 0.5-1.0$, were narrow gap semiconductors with large charge carrier effective masses and comprised of heavy-atom elements arranged in high-symmetry crystal structures.

In spite of a plethora of studies, both theoretical and experimental, on new thermoelectric materials, for the next quarter of a century the field remained somewhat quiescent, with no substantial improvements in the figure of merit. It was recognized that the thermoelectric problem was one in which the figure of merit requires a combination of *contraindicated* physical properties: a set of properties which nature is not disposed to provide in a single material, for instance high electrical conductivity and large Seebeck coefficient, or high electrical conductivity and low thermal conductivity.

The situation changed rapidly in the mid-1990s for a number of different reasons. Firstly, due to a greater awareness of the insecurity of our energy future and the impact of energy usage on the precarious state of the environment, the scientific and technical communities began to focus more heavily on the science and technology of energy processes and energy-related materials; secondly, the field began to benefit from the use of more sophisticated synthesis techniques and sensitive characterization methods, especially at the nanoscale; and finally, and perhaps most importantly, the scientific community developed an increased understanding, at a fundamental level, of the nature of electrical and thermal transport processes in solids, two examples being the theoretical treatment of *Hicks* and *Dresselhaus* [57.8, 9] of thermoelectricity in quantum-confined systems and the origination by *Slack* [57.10] of the concept of *phonon-glass-electron-crystal*.

From (57.1), it can been seen that commonly available thermoelectric materials like Bi_2Te_3 and PbTe, with zT values on the order of unity or less, have a conversion efficiency of under 10%. Clearly the most direct means of increasing the efficiency is to improve upon the transport properties that comprise the figure of merit. We now look at this relation in somewhat greater detail

$$z = \frac{\alpha^2\sigma}{\kappa} = \frac{P}{\kappa}\,. \qquad (57.2)$$

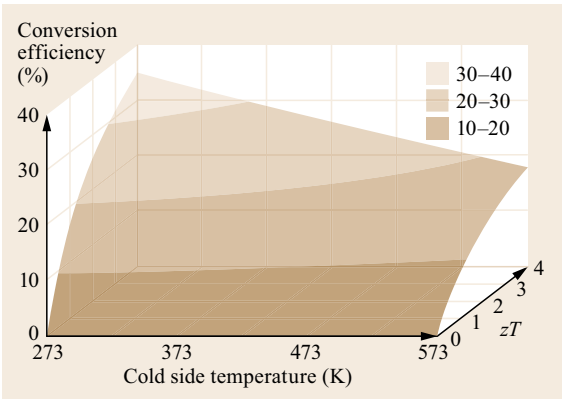


Fig. 57.1 Efficiency of a thermoelectric generator with a hot-side temperature of 700 °C (973 K) as a function of cold-side temperature and figure of merit zT

Here κ is the total thermal conductivity and consists of the sum of the electronic (κ_e) and lattice, or phonon (κ_p) contributions. The numerator term, the *electrical* part of the figure of merit, is called the power factor (P).

One of the oldest [57.7] and most effective ways of increasing z is by alloying to form solid solutions such as $Pb_{1-x}Sn_xTe_{1-y}Se_y$ or $Bi_{2(1-x)}Sb_{2(1-x)}Te_{3(1-y)}Se_{3y}$. The basis for this approach is the strong phonon scattering that occurs due to the mass difference on one or both of the binary compound lattice sites. In this approach the long-range periodicity of the lattice is preserved, but there are intense short-range perturbations. Electrons, with wavelengths many times the interatomic distance, are essentially uninfluenced by these perturbations, while phonons, the characteristic wavelength of which is on the order of the interatomic distance, can be strongly scattered by these local distortions in atom mass and size. The net result is that the ratio σ/κ_p, or more particularly μ/κ_p, where μ is the carrier mobility, is increased substantially, leading to an enhancement of z.

Within the last decade there have been many novel approaches to designing new thermoelectric materials with values of figure of merit beyond those possible using the solid-solution method. These approaches can be broadly grouped into two categories: 1. those that seek to increase the Seebeck coefficient α; and 2. those that seek to increase the ratio of σ/κ. An example of the former includes the use of low-dimensional systems [57.8, 9, 11] in which the altered density of electron states leads to enhanced α; examples of the latter include *phonon-glass electron-crystal* structures such as filled skutterudite [57.12] and clathrate [57.13] compounds with glass-like lattice thermal conductivity, and superlattice structures [57.14] with increased phonon scattering by interfaces.

The main challenge in any methodology seeking to reduce the lattice thermal conductivity is maintaining adequate carrier mobility, and hence electrical conductivity. Magnetic scattering, for instance, using rare earth ions or transition metal ions with partially filled d-shells, can dramatically reduce the lattice conductivity, but is inevitably accompanied by a strong reduction in carrier mobility [57.15]. Even the filled skutterudites, with their notable reduction in κ_p, suffer a strong reduction in mobility upon filling and reach high z only by an alteration of band structure that causes an increase in Seebeck coefficient relative to their unfilled counterparts.

57.2 Semiconductors as Thermoelectric Materials

57.2.1 Electronic Properties of Charge Carriers in Semiconductors

An intrinsic semiconductor is defined as a material in which, at $T = 0$, completely filled electron energy levels (known collectively as the valence band) are separated in energy from energy levels that are empty (known collectively as the conduction band); thus, at zero temperature, a semiconductor is insulating, because there are no empty states nearby in energy into which an electron may move under application of an electric field. As temperature is increased, electrons may be thermally excited across the energy gap to occupy states in the conduction band. Electrons so promoted in energy are then free to move to other nearby states in the conduction band, and thus can contribute to electrical conductivity. At the same time, the empty states left behind in the valence band, called holes (equal in number to electrons promoted), can also contribute to conduction as a positive charge carrier. The two types of charge carriers in an intrinsic semiconductor, electrons (n) and holes (p), move through the lattice with distinct mobilities, μ_e and μ_h respectively, and thus in an intrinsic semiconductor, the total electrical conductivity, σ, is given by the sum of the partial conductivities of electrons (σ_e) and holes (σ_h)

$$\sigma = \sigma_e + \sigma_h = ne\mu_e + pe\mu_h \,. \tag{57.3}$$

Because electrons and holes are fermions, the probability of their occupation of a given energy state is determined by Fermi–Dirac statistics. The Seebeck coefficient of an intrinsic semiconductor, on the other hand, is determined by the partial Seebeck coefficient of the electrons and holes, weighted by their conductivities, according to the relation

$$\alpha = \frac{\alpha_e\sigma_e + \alpha_h\sigma_h}{\sigma_e + \sigma_h} \,. \tag{57.4}$$

In an intrinsic semiconductor, since by definition $n = p$, this relation simplifies to one in which the partial Seebeck coefficients are weighted by the carrier mobilities

$$\alpha = \frac{\alpha_e\mu_e + \alpha_h\mu_h}{\mu_e + \mu_h} \,. \tag{57.5}$$

Once again, the partial Seebeck coefficients of electrons and holes are, as we shall see in a moment, determined

by Fermi–Dirac statistics. Here we note, however, that the *sign* of the Seebeck coefficient is the same as that of the sign of the charge carrier, that is, negative for electrons and positive for holes. Thus in an intrinsic semiconductor, the total Seebeck coefficient is always less than the partial Seebeck coefficient of either the electrons or the holes. For this reason, virtually all useful thermoelectrics are designed using extrinsic doping to produce n- and p-type materials respectively.

Going a step further, the mobility of charge carriers may be written as

$$\mu = \frac{e\tau}{m^*} \,. \tag{57.6}$$

Here τ is the scattering time of the charge carriers and m^* is the carrier effective mass. We make a further assumption that the charge carrier scattering time depends on the carrier energy E as

$$\tau(E) = \tau_o E^{\lambda} \,. \tag{57.7}$$

Returning now to the equation for the figure of merit, using the definition of the conductivity this equation may now be written, for an n-type material

$$z = \frac{\alpha^2\sigma}{\kappa} = \frac{\alpha^2(ne\mu)}{\kappa} = (\alpha^2 ne)\frac{\mu}{\kappa} \,. \tag{57.8}$$

In this form, z depends on one quantity, $\alpha^2 ne$, which is determined by the electronic band structure, and a second quantity μ/κ, which depends on the nature of electron and phonon scattering. To analyze the figure of merit more deeply, we break it into these two pieces and discuss each in turn.

Fermi–Dirac statistics describe the probability of occupation of electron (and hole) states in semiconductors. At temperature $T > 0$, there is finite probability that an electron can occupy the conduction band or a hole can occupy the valence band. In this situation, the electron concentration and Seebeck coefficient are given by the expressions [57.3]

$$\begin{aligned} n &= \frac{4}{\sqrt{\pi}} \left(\frac{2\pi m^* k_B T}{h^2} \right)^{\frac{3}{2}} F_{\frac{1}{2}} \\ \alpha &= \mp \frac{k_B}{e} \left[\frac{\left(\frac{5}{2} + \lambda\right) F_{\frac{3}{4}+\lambda}}{\left(\frac{3}{2} + \lambda\right) F_{\frac{1}{2}+\lambda}} - \eta \right] . \end{aligned} \tag{57.9}$$

Here the Fs are the so-called Fermi integrals of order r

$$\begin{aligned} F_r &= \int_0^{\infty} x_r f(\eta) \mathrm{d}x \quad \text{where} \\ x &= \frac{E}{k_B T} \quad \text{and} \quad \eta = \frac{E_f}{k_B T} \,. \end{aligned} \tag{57.10}$$

The Fermi–Dirac distribution f is given by

$$f = \frac{1}{1 + \exp(x - \eta)} \,. \tag{57.11}$$

As usual h and k_B designate the Planck and Boltzmann constants respectively, and E_f is the charge carrier Fermi energy. The reduced Fermi energy η is the Fermi energy as measured from the edge of the band, i. e., $\eta = 0$ at the bottom of the conduction for electrons and is negative for Fermi energy in the gap, and similarly $\eta = 0$ for holes at the top of the valence band and is negative again for Fermi energy in the gap. The Fermi energy, and thus η, takes on positive values as it moves either into the conduction band for electrons or into the valence band for holes. Using these equations, for a given electron scattering coefficient λ and carrier effective mass m^*, one can calculate the Seebeck coefficient α and the quantity $\alpha^2 n$ that enters directly into the thermoelectric figure of merit.

Figure 57.2 shows the results of such calculations and compares them to the results obtained using classical statistics, i. e., treating the electrons in a solid as if they were like molecules in a gas. Some interesting and important differences emerge. For $\eta < -2$, i. e., for Fermi energy well in the gap, the classical and Fermi–Dirac statistics give identical results. This is the nondegenerate case of a lightly doped semiconductor. For classical statistics the Seebeck coefficient falls more rapidly and actually reaches zero at $\eta = 2$. For Fermi–Dirac statistics, the Seebeck coefficient is diminished as the Fermi level moves into the conduction or valence band, but it never reaches zero. Similarly, the important quantity $\alpha^2 n$ is maximized for Fermi energy right at the band edge for classical statistics, whereas for the more general Fermi–Dirac case, the maximum occurs for Fermi energy one unit of $k_B T$ into the band.

Of course, experimentally, one does not directly control the Fermi energy; rather, one controls the carrier concentration through doping. To see this more clearly, Fig. 57.3 plots the quantity $\alpha^2 n$ versus carrier concentration n, for Fermi–Dirac statistics. Here we clearly see the origin in the peak in the power factor $\alpha^2\sigma$, predominantly arising due to the inverse relationship between α and n.

Thus Fermi–Dirac statistics give a neat interpretation of the quantity $\alpha^2 n$ and in particular what value should be chosen for the Fermi energy and can be used as a guide to approximate what level of doping is required in order to optimize the thermoelectric figure of merit. In practice, because others factors such as the dependence of effective mass on doping level and mixed carrier scattering mechanisms, usually the

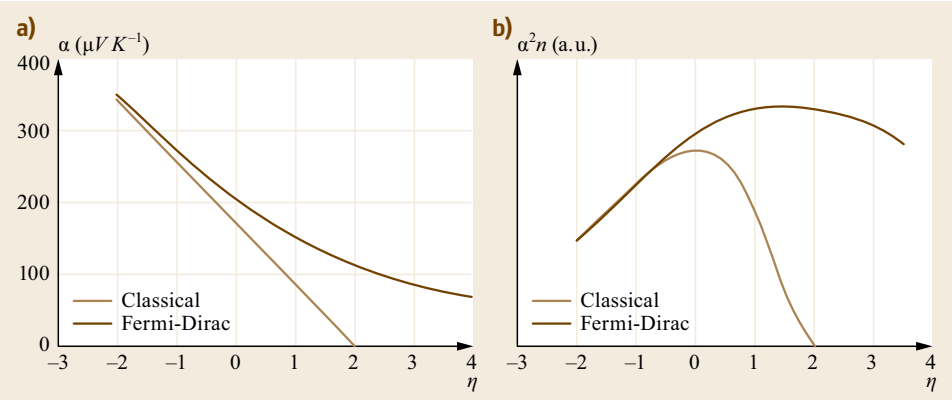


Fig. 57.2 Seebeck coefficient **(a)** α and **(b)** $\alpha^2 n$ as a function of reduced Fermi energy for holes with effective mass $m^* = 0.3$

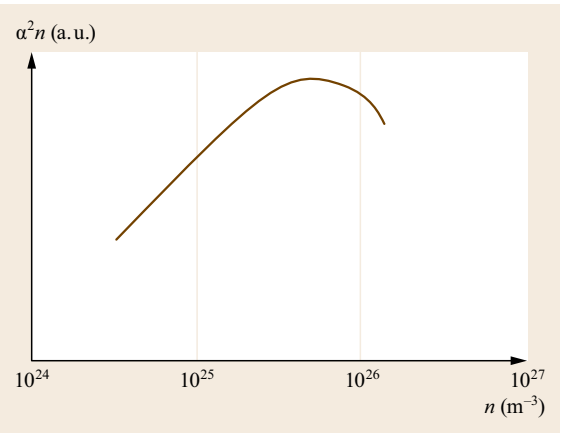


Fig. 57.3 The quantity $\alpha^2 ne$ as a function of carrier concentration at 300 K for a carrier effective mass $m^* = 0.3$

optimum doping level for maximum power factor is determined by trial and error.

57.2.2 Semiconductors with Low Intrinsic Thermal Conductivity

While a reasonably clear understanding of the thermoelectric power factor and the elements that influence it can be obtained with knowledge of the band structure of a material, the dominant carrier scattering mechanism, and the use of carrier statistics, that of course is only part of the thermoelectric problem – we have not yet dealt with the denominator of the figure of merit; the thermal conductivity. While to first order there is much to be learned from *decoupling* the treatment of the power factor from that of the thermal conductivity, in reality all of the transport properties comprising the figure of merit are intertwined and interdependent at some level.

Serious study of the thermal conductivity of crystalline solids was begun before World War II, including the work of *Akhiezer* [57.16], *Leibfried* and *Schloemann* [57.17], and *Lawson* [57.18]. A basic conclusion of all of these studies was that the lattice thermal conductivity of crystals near and above room temperature obeyed the general relation [57.19]

$$\kappa_L \propto \frac{\theta^3}{\gamma^2 T} , \qquad (57.12)$$

where θ is the Debye temperature, γ the Grüneisen constant, and T the absolute temperature. On the basis of this simple relation, one can thus predict that materials with high Debye temperature and/or small Grüneisen parameter will have high thermal conductivity, and vice versa. Since the Debye temperature depends inversely on the average atom mass, this further implies that low thermal conductivity will be favored in materials comprised of heavy mass atoms. Subsequently, *Slack* [57.20] extended this relationship to include a dependence of crystal complexity, i. e., number of atoms in the unit cell, and showed that crystal structures with large unit cells generally possess lower thermal conductivity due to a more complex phonon spectrum. Taken together, this information suggests that the place to look for good thermoelectric materials is amongst semiconductors with low Debye temperature, large numbers of atoms in the unit cell, heavy atom masses, and large Grüneisen coefficients. At the same time, of course, these compounds must possess high thermoelectric power factor, as determined by the equations outlined in Sect. 57.2.1 above.

In practice, there are thousands of possible compounds and dozens of different crystal structure types that must be considered for possible choices for thermoelectric materials, but the power of classifying materials according to their crystal structure and bonding can be illustrated by comparing some results for diamond/zincblende and rock salt structure compounds. It

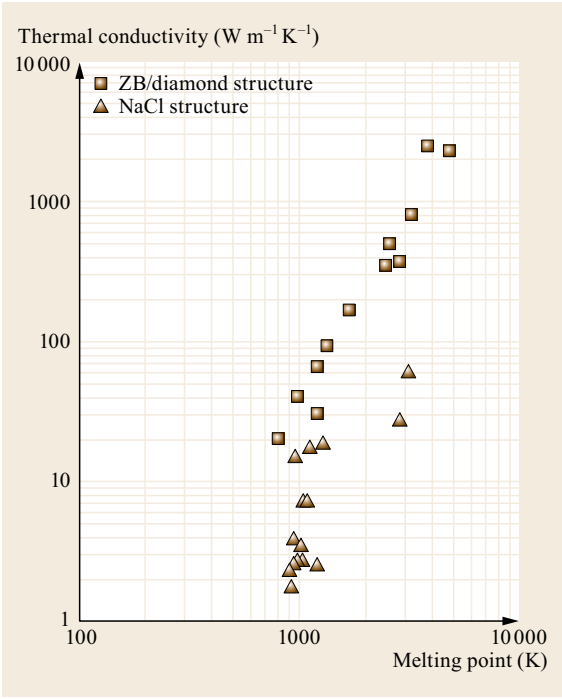


Fig. 57.4 Lattice thermal conductivity versus melting point for a host of diamond, zincblende and rock salt structure compounds

may be noted that the same properties that favor low thermal conductivity (heavy atom masses and low Debye temperature) are also typical of lower melting point compounds, and indeed (Fig. 57.4) there is a strong correlation between lattice thermal conductivity and melting point, at least within families of compounds such as the diamond/zincblende structure and the rock salt structure. The magnitude of the conductivity, however, does depend on the nature of bonding between atoms, as clearly the covalently bonded diamond/zincblende structure compounds and elements have higher thermal conductivity than the ionically bonded rock salt structures. On this basis alone one would naturally be led to choose the rock salt structure compounds over the zincblendes due to the lower thermal conductivity. On the other hand, as mentioned above, we must ultimately find materials that not only have low lattice thermal conductivity, but also high charge carrier mobility, in order that the electrical conductivity may remain high, and this tends to be favored in covalently bonded crystals. It is the hallmark of research on thermoelectric materials that one is constantly confronted, in different forms and guises, an inherent *contraindication* in properties that can frustrate the search for high z.

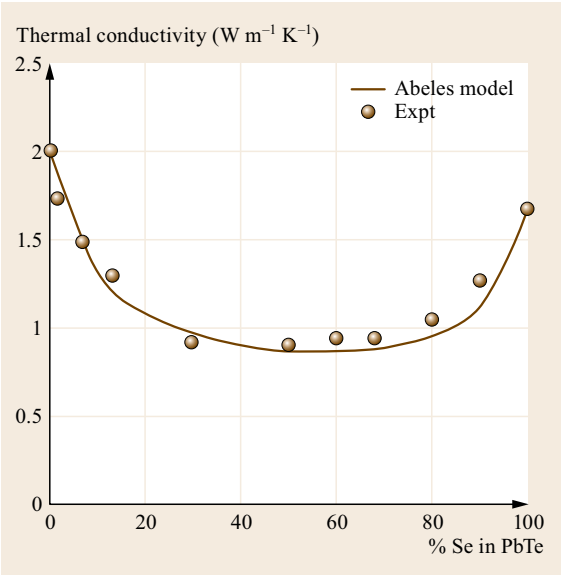


Fig. 57.5 Lattice thermal conductivity of the $PbTe_{1-x}Se_x$ solid solution. The reduction in thermal conductivity is maximized near 50% solute concentration, the point at which mass disorder in the lattice is largest

57.2.3 Semiconductor Solid Solutions

While the above approach for finding compounds with low intrinsic thermal conductivity was very successful, it resulted in very few materials being identified with high z. Researchers quickly noted that z could be increased further, however, by lowering the thermal conductivity using defects or impurities, as long as the electrical properties could be largely maintained. By far the most successful method of achieving this result was by realizing isoelectronic solid solutions of elemental and compound semiconductors. The use of solid solutions to reduce thermal conductivity is rooted in the concept that impurities, when doped into a host lattice, will scatter phonons due to differences in their mass and size with respect to the host atoms. Such behavior was predicted originally by *Pomeranchuk* [57.21], was explored theoretically first by *Klemens* [57.22], and was experimentally demonstrated on a host of materials systems. In the case of thermoelectrics, as we have discussed, it is imperative that when such impurities are introduced into the lattice, they disturb the transport of electrons as little as possible. For example, dopants introduced into a semiconductor to increase the charge carrier concentration can at high concentrations reduce charge carrier mobility due to ionized impurity scattering. Solute atoms that have the same valence electron count as the host atoms in a lattice, on the other hand,

are neutral impurities and interact with charge carriers much less strongly; the resulting reduction in mobility can be quite small. By choosing atoms in the same column of the periodic table (but obviously from a different row), reduction in lattice thermal conductivity can be achieved while largely maintaining the good electrical properties. Moreover, because of the same valence electron count, these atoms can be substituted at large fractions, and in many cases will form a complete solid solution.

The isoelectronic substitution approach was applied very successfully to both PbTe [57.7] ($PbTe_{1-x}Se_x$ and $PbTe_{1-x}S_x$) and Bi_2Te_3 [57.8] ($Bi_2Te_{3-x}Se_x$, $Bi_2Te_{3-x}S_x$, and $Bi_{2-x}Sb_xTe_3$), and was also successfully implemented in the group IV semiconductor system Si-Ge [57.23]. Several theoretical models have been developed to account for the lattice thermal conductivity of alloy solid solutions. Figure 57.5 shows an example of the experimental data for $PbTe_{1-x}Se_x$ in comparison to the model predictions of *Abeles* [57.24]. As expected, the reduction in thermal conductivity, over a factor of two in this case, is maximum near 50% substitution, when the mass disorder in the lattice is at its largest.

57.3 New Concepts in Thermoelectric Materials Design

By the mid-20th century, most of the ideas discussed above on how to optimize the thermoelectric properties of materials were well known and had been successfully implemented to identify PbTe and Bi_2Te_3 solid solutions as good thermoelectric materials. As part of the electronics revolution, thermoelectricity seemed to be on the cusp of explosive discovery and development as a new technology for generating electrical power and providing solid-state heating and cooling.

It seemed that researchers simply had to apply the new concepts and knowledge to find materials even better than the lead and bismuth tellurides. In spite of that promise, however, very little progress was made over the next several decades, and even by 1990 essentially no new thermoelectric materials with properties superior to the heritage materials were discovered. On top of this, the very factors that make these heritage materials such good thermoelectrics, for example heavy masses and weak atomic bonding, also mean that they tend to be mechanically poor, and comprised of elements that are not in great abundance in the Earth's crust. Thus, even these good thermoelectric materials found limited applications and thermoelectricity was a field on the verge of becoming relegated to the backwaters of scientific inquiry.

By the late 20th century, however, a number of factors and influences came together to lead to a refocusing of efforts on discovery of new thermoelectric materials. While indeed over the past several decades few if any new thermoelectric materials of note were discovered, steady progress was made at the level of advancing our fundamental understanding of thermal and electronic transport in solids. At the same time, new ideas, such as quantum confinement and nanoscience, were introduced and explored in a variety of materials systems, and new methods of synthesis of materials were perfected. Thus several tools and fresh approaches could be brought to bear on the field – one just needed the impetus for doing so. That impetus was provided by a new sensitivity toward the utilization of energy and its potential impact on the climate and environment of the Earth. Suddenly, this old idea of converting heat to electricity, mostly viewed upon as a laboratory curiosity, became a new way to generate power, increase efficiency, and decrease emissions of potentially harmful pollutants. When large sources of federal funding became available for research, scientists were quick to bring to bear a host of these new ideas and new approaches to the field of thermoelectricity. Here, in the interest of space and to keep the discussion focused, we consider three new concepts that have emerged over the last two decades and that continue to drive thermoelectric materials research today, and we describe each of these new approaches in terms of model materials systems:

- Phonon-glass-electron crystal (PGEC)
- Bulk nanostructured materials
- Crystals with large anharmonicity.

57.3.1 Phonon-Glass Electron Crystal: Skutterudites

While the solid-solution approach can reduce lattice thermal conductivity strongly, the magnitude of the thermal conductivity typically is still several times larger than that observed in materials with very low thermal conductivity, such as glasses and other amorphous structures. Study of the thermal conductivity of the latter materials has its own long and storied history, dating back to at least the studies of *Zeller* and *Pohl* [57.25], who noted the vast difference in thermal conductivity between crystalline and amorphous SiO_2. A significant amount of work was sub-

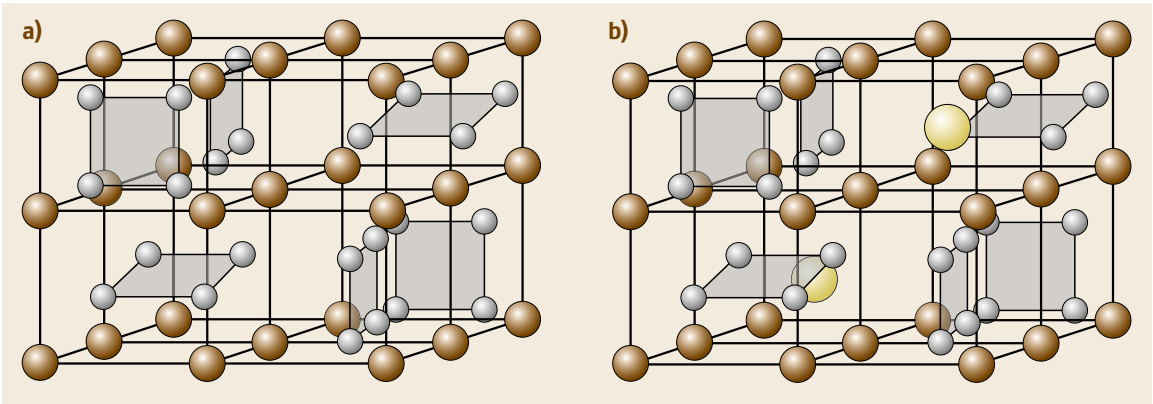


Fig. 57.6a,b The skutterudite crystal structure. A prototypical compound is $CoSb_3$; Co atoms are *brown* and Sb are *gray*. The Sb atoms reside in square planar rings occupying six of the eight quadrants in the unit cell. In the *unfilled* compound the remaining two quadrants are empty **(a)**; in the *filled* version they are occupied by a rare earth or alkaline earth atom **(b)**

sequently performed on such materials that described heat transport by a phonon-tunneling-type mechanism. Eventually, the concept of *minimal thermal conductivity (MTC)* was introduced that described how in such amorphous and highly disordered solids thermal conductivity reaches a theoretically minimum value corresponding to a phonon mean free path on the order of the lattice spacing [57.26].

While the thermal transport properties of glassy systems is an interesting topic in its own right, it did not seem to bear importance relative to the thermoelectric problem; most amorphous solids are poor electrical conductors and are not good candidates for thermoelectrics. Slack, however, noted that there were a few crystalline solids that exhibited glass-like thermal conductivity – a notable example being YB_{66}. This compound displays galleries of boron atoms, between which Y ions can tunnel. In spite of its crystalline character, the thermal conductivity reaches theoretically minimum values. Slack pointed out that perhaps crystalline semiconductors may exist that have not only good thermoelectric properties, but at the same time also exhibit MTC behavior, and suggested that skutterudite compounds could be such a material.

Like many thermoelectric materials, skutterudites were first studied in the Soviet Union by *Dudkin* et al. [57.27–30]. They found that this family of compounds, typified by the composition $CoSb_3$, could be doped both n- and p-type and displayed a good combination of high mobility, large Seebeck coefficient, and large thermoelectric power factor. However, in spite of the rather complex crystal structure (Fig. 57.6) consisting of a cubic lattice Co atoms containing Sb four-membered square rings in six of the eight quadrants of the unit cell, the lattice thermal conductivity of skutterudites was found to be too large to yield high z. This early work did not explore in detail any mechanisms to reduce the thermal conductivity.

More than a decade later, *Jeitschko* [57.31] showed the skutterudite structure could be *filled* by placing lanthanide ions in the two vacant quadrants in the unit cell, and that these ions exhibited large x-ray displacement parameters, indicating that they were loosely bound or *rattled* about their ionic positions. Some 15 years later, with tremendous insight, Slack suggested that these *filled* skutterudites might exhibit minimal thermal conductivity due to the *rattling* phonon mode of the lanthanide ion. A strong reduction in lattice thermal conductivity of the filled skutterudite $CeFe_4Sb_{12}$ relative to its unfilled parent compound, $CoSb_3$, was experimentally demonstrated by *Morelli* and *Meisner* [57.12] shortly thereafter. Skutterudites were the subject of intense scrutiny, both from the point of view of their interesting physics and chemistry, and for their potential thermoelectric applications, and soon it was shown that these compounds could exhibit a figure of merit in excess of unity at high temperature.

Crucial to the performance of skutterudites is the combination of low lattice thermal conductivity in the presence of good semiconducting electrical properties. Because the filling ion (which in addition to an element from the lanthanide series, may also be an alkaline earth ion) donates electrons to the conduction band, its presence can also modify the electrical properties. By simultaneously substituting Fe for Co, the maximum filling fraction can be adjusted and the electrical properties controlled to produce either n- or p-type materials of desired carrier concentration [57.32]. Recognizing that the *rattling* behavior exhibited by the filler atom was a form of resonant scattering of phonons with

a characteristic frequency determined by the mass of the filler atom and the size of the cage, researchers used combinations of fillers (double or even triple filling) to provide different characteristic resonant frequencies, thus producing the ability to scatter wider ranges of the phonon spectrum and reduce the thermal conductivity even further [57.33, 34].

Research and development on skutterudites has continued up to the present day, and impressive zT values in excess of 1.5 for n-type materials and in excess of unity for p-type materials have been achieved. Prototype generators based on skutterudites for exhaust gas-driven power generation on vehicles have been built.

57.3.2 Bulk Nanostructured Thermoelectrics: Pb–Chalcogenide Nanocomposites

About the same time that Slack introduced the PGEC concept, researchers began to contemplate the role that quantum confinement, low-dimensionality, and nanostructuring may play in thermoelectricity. The papers by *Hicks* and *Dresselhaus* [57.8, 9] considered the influence of the modification of the electron density of states that occurs in one-dimensionally (quantum wire) and two-dimensionally (quantum well) confined systems, and these theoretical studies launched intense investigation of these topics. While the promising early results of *Harman* et al. [57.35] and *Venkatasubramanian* et al. [57.14] have not been reproduced, the subsequent studies on nanostructured materials that grew out of these efforts were to bear fruit in the coming years.

An illustration of the positive effects of nanostructuring is provided by the work on so-called LAST materials and hierarchically architectured nanocomposite thermoelectrics. The former is an acronym for compounds based on the PbTe-$AgSbTe_2$ (lead–antimony-silver-telluride) family and was extensively studied by *Kanatzidis* et al. [57.36]. The unusual thermoelectric properties of these materials were noted much earlier by both *Fleischmann* [57.37] and *Irie* et al. [57.38]. As $AgSbTe_2$ is isostructural with PbTe (face-centered cubic rock salt structure) the PbTe-$AgSbTe_2$ compounds were originally thought to form a complete solid solution. The work of the Kanatzidis group showed, however, that small Ag-Sb rich nanoclusters form in these materials, and in fact this morphology is responsible for an enhancement of the thermoelectric properties, primarily through a reduction in thermal conductivity by scattering of phonons at the interface between the host matrix and the secondary phase. Phonon scattering by inclusions on the nanoscale has a long history of study, dating back to at least the work of *Slack* on

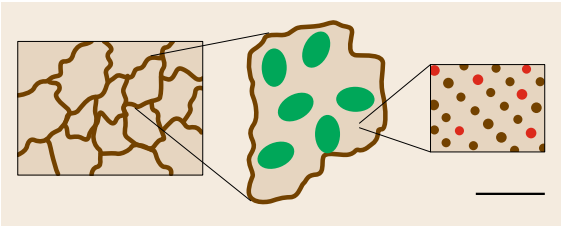


Fig. 57.7 *Hierarchical* structuring of thermoelectric materials. Defects and impurities on different length scales scatter phonons of different wavelengths. Grain boundaries on the mesoscale (*left*; scale bar $\approx 10^3-10^4$ nm) scatter long-wavelength phonons; nanoscale precipitated second phases (*center*; scale bar $\approx 10^1-10^3$ nm) scatter medium-wavelength phonons; and atomic scale impurities (*right*; scale bar $\approx 10^0-10^1$ nm) scatter short-wavelength phonons

$CaCl_2$ precipitates in KCl [57.39], and in fact the use of nanoscale inclusions to reduce the thermal conductivity of Si-Ge thermoelectric alloys was studied as early as 1993 [57.40]. The nanostructured LAST compounds exhibited a thermoelectric figure of merit of $zT = 1.6$.

While nanoscale inclusions can be effective scatterers of medium-wavelength phonons in solids, phonons with frequencies both higher and lower than those that interact strongly with inclusions can still transport heat. On this basis, workers began conceiving of materials systems that could scatter a broader range of phonon frequencies, and thus reduce the thermal conductivity further. It was in this spirit that the concept of *hierarchical structuring* of defects in thermoelectric materials was born (Fig. 57.7). Here the idea is that by combining different types of defects/impurities, each type of which scatter phonons in different frequency ranges, almost all heat transport by phonons can be blocked, just as different optical filters can block different wavelengths of light and when combined can completely block photon transmission. This concept was realized experimentally in the work of *Biswas* et al. [57.41] in which atomic-scale defects, nanostructured second phases, and grain boundaries were combined to scatter high-, medium-, and long-wavelength phonons respectively. The base materials system used to achieve this was once again PbTe, into which was doped SrTe, another rock salt structure compound. The solubility limit of Sr for Pb in PbTe, however, is rather small, on the order of 1%. At first some Sr substitutes for Pb (atomic scale defects), but as the concentration is increased, SrTe nanoclusters form in the material (nanoscale defects). By processing these materials using spark plasma sintering, a fine grain size can be maintained (mesoscale defects). The resulting hierarchy in defect size reduces the thermal conductivity in these materials down to close to

minimal thermal conductivity values. As a result, the thermoelectric figure of merit of optimally doped samples reaches values as high as 2.2.

57.3.3 Crystals with Large Anharmonicity: Tetrahedrites

While impressive results have been obtained using nanostructuring, there are some potential drawbacks with this approach. The composition, distribution of second phases, and grain sizes must be carefully controlled in order to achieve the maximum reduction in thermal conductivity. The processing of these materials is, while straightforward, time consuming. Finally, the stability of these materials with thermal and mechanical cycling is a still-open question.

For these reasons, workers have been looking for other approaches to finding solids with low thermal conductivity. One such approach is to not use any defects or impurities at all – instead, to try to find materials with built-in, intrinsically low thermal conductivity. Of course, as discussed earlier, many of the main drivers of low thermal conductivity, such as complex crystal structures and heavy atom masses, have been known for many decades. Somewhat overlooked, however, has been the nature of phonon scattering, and how it might be maximized intrinsically based on crystal structure and bonding in a solid.

In a simple model of a crystal lattice, atoms are bonded together like masses connected by springs. In a quantum-mechanical picture of the lattice, the vibrations of atoms are quantized and are collective modes known as phonons. For a perfectly harmonic lattice, i. e., one in which the potential energy of each ion is a quadratic function of its displacement, there are no interactions between phonons, and the intrinsic thermal resistance is zero. Lattice thermal resistivity in solids arises due to anharmonicity in the lattice vibrational spectrum – i. e., vibrations for which the potential energy has additional terms of higher order than quadratic. For instance, a cubic term in the potential energy gives rise to anharmonic three-phonon processes, two examples of which are given in Fig. 57.8. Four-phonon and higher-order processes will arise from terms quartic and higher.

A measure of the anharmonicity in a crystal lattice is given by the Grüneisen coefficient γ. The ideally harmonic crystal would have $\gamma = 0$ for all phonon modes and frequencies; this is an ideality that does not exist in nature. The larger the value of γ, the stronger the anharmonicity in the lattice. Thus, a good place to look for crystalline solids with low intrinsic thermal conductivity is in those materials with highly anharmonic phonon spectra.

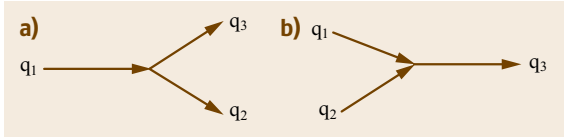


Fig. 57.8a,b Two types of three-phonon anharmonic processes that give rise to intrinsic lattice thermal resistivity in solid. **(a)** A phonon of wavevector q_1 is destroyed and two phonons of wavevectors q_2 and q_3 are created (phonon creation); **(b)** two phonons of wavevector q_1 and q_2 are destroyed and a phonon of wavevector q_3 is created (phonon annihilation)

But what gives rise to anharmonicity? This very fundamental question has been of intense interest to researchers in recent years. One very interesting and fruitful approach is the concept of the effect of lone pair electrons on vibrational modes in solids. The seed of this idea goes back to the work of *Zhuze* [57.42] in the Soviet Union, who noticed the large difference in thermal conductivity between certain compounds depending on their composition and the valence state of the atoms comprising them, and in particular those containing Sb ions. The electronic configuration of Sb is $[Kr]4d^{10}5s^25p^3$. In certain coordinations of Sb in the lattice, all five of the valence electrons participate in bonding; an example is the zincblende-like compounds Cu_3SbSe_4, in which each Sb atom is tetrahedrally coordinated. On the other hand, in other compounds, for instance Cu_3SbSe_3, the Sb is trigonally coordinated, and only three of the valence electrons participate in bonding; the remaining two electrons reside close to the Sb ion in an unbonded configuration. This configuration of unbonded electrons around Sb (or indeed any other ion) is known as the lone pair. Zhuze suggested that the electron clouds of overlapping Sb lone pair electrons gives rise to nonlinear forces in the lattice vibrations, i. e., give rise to anharmonicity. This effect was recently invoked by to explain the low intrinsic thermal conductivity of $AgSbTe_2$ [57.43]. *Skoug* and *Morelli* [57.44] showed that the thermal conductivity of large families of compounds exhibiting pnicogen atom lone pairs could be understood in this context, and were able to relate the degree of thermal resistivity with the distance of the lone pair electrons from the nucleus of the atom, with those configurations in which the lone pair was furthest from the nucleus having the strongest reduction in thermal conductivity. *Ozolins* et al. [57.45] formalized the connection between the presence of the lone pair and the degree of anharmonicity using calculations of the lattice dynamics for several compounds.

A very interesting example of how this effect can manifest itself in enhanced thermoelectric behavior

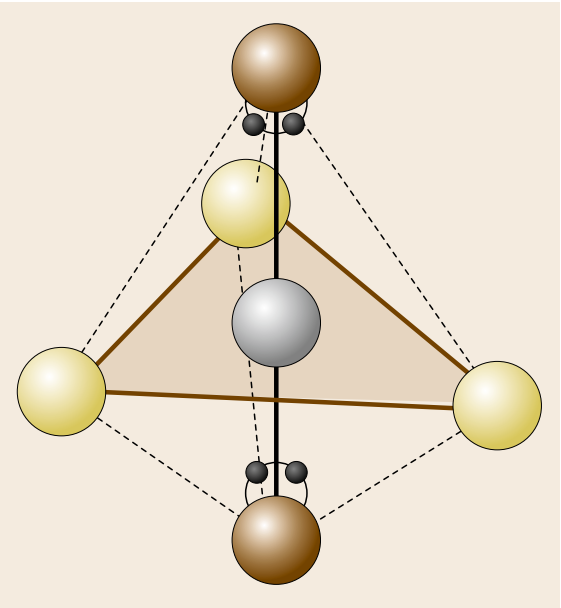


Fig. 57.9 The trigonally coordinated Cu site in tetrahedrite. A copper atom (*gray*) is trigonally surrounded by three sulfur atoms (*yellow*). Two antimony atoms (*brown*) possessing unbonded lone pair electrons (*black dots*) reside above and below the trigonal plane. The Sb lone pairs act to *pull* the copper atom of the trigonal plane. The trigonal bipyramid arrangement is a cage in which the Cu atom rattles, giving rise to PGEC-like behavior

is provided by the family of compounds known as tetrahedrites. The base composition of tetrahedrite is $Cu_{12}Sb_4S_{13}$. Its crystal structure is cubic with 29 atoms in the unit cell. Tetrahedrite possesses several key structural features that have strong implications for its thermoelectric properties. First, the Sb atoms are trigonally coordinated and do indeed display the classic lone pair as described above. Additionally, Cu occupies two distinct crystallographic sites in the lattice; one in which the Cu atom is tetrahedrally coordinated by sulfur, and a second site in which the Cu ion resides in a trigonal plane of sulfur atoms; six Cu atoms are supposed to reside in each site in the lattice. Guided by electronic band structure calculations, *Lu* et al. [57.46] showed that while the base composition $Cu_{12}Sb_4S_{13}$ is metallic, the substitution of divalent Zn for Cu can yield semiconducting behavior. At the same time, the lattice thermal conductivity exhibited values below $0.5\,Wm^{-1}K^{-1}$, i. e., close to the minimal thermal conductivity of a solid. By optimizing the Zn concentration they demonstrated that tetrahedrites can reach zT values close to 0.9. Since then, several other studies [57.47–49] investigating different atoms substitutions and different methods of processing these materials has resulted in zT values in the range of 1.1−1.2. Recently, *Lai* et al. [57.50] combined density functional theory (DFT) calculations with high resolution x-ray scattering studies to show that Sb lone pairs residing above and below the trigonal Cu–S plane (Fig. 57.9) act to displace the Cu away from the plane, causing a buckling-type behavior. In this picture, the three sulfur atoms surrounding the copper ion together with the lone-pair-containing antimony above and below the plane comprise a bipyramidal *cage* in which the copper atom *rattles*. This PGEC-like behavior is thought to be the source of low thermal conductivity in these compounds.

With the growing ability to calculate the properties of materials from first principles, approaches such as that used for tetrahedrite offer the opportunity to use powerful materials-by-design principles to conceive of new structures and compositions with the unusual combination of transport properties required for thermoelectricity directly *built-in* to them. This is a new paradigm in thermoelectric materials research that offers great promise from thermoelectric in the future.

57.4 Summary and Future Outlook

Thermoelectrics are very unique material systems in which the fundamental behavior associated with electrons and phonons in crystalline solids can give rise to interesting thermal and electronic effects, which can be important technologically. The stringent requirements on Seebeck coefficient, electrical conductivity, and thermal conductivity that are necessary to produce good thermoelectric behavior continue to offer challenges and inspiration to myriad researchers. Thermoelectricity has become a proving ground for numerous new and innovative concepts in physics, chemistry, and materials science, including quantum size effects, nanostructuring, phonon-glass-electron crystal behavior, and large anharmonicity. New knowledge uncovered in these and other areas is driving improvements in material performance, and the future holds promise that highly efficient thermoelectric devices for both power generation from waste heat and solid-state climate control will become a reality.

References

57.1 US Department of Energy: Lawrence Livermore National Laboratory, Estimated Energy Flow Chart, https://flowcharts.llnl.gov/
57.2 US Department of Energy: Alternative Fuels Data Center, Fuel properties Comparison, http://www.afdc.energy.gov/fuels/fuel_comparison_chart.pdf
57.3 H.J. Goldsmid: *Applications of Thermoelectricity* (Methuen, London 1960)
57.4 T.J. Seebeck: Abh. Akad. Wiss. Berlin **21**, 289 (1822)
57.5 W. Haken: Ann. Phys. **32**, 291 (1910)
57.6 E. Altenkrich: Phys. Z. **12**, 920 (1911)
57.7 A.F. Ioffe: *Semiconductor Thermoelements and Thermoelectric Cooling* (Infosearch, London 1957)
57.8 L.D. Hicks, M.S. Dresselhaus: Phys. Rev. B **47**, 12727 (1993)
57.9 L.D. Hicks, M.S. Dresselhaus: Phys. Rev. B **47**, 16631 (1993)
57.10 G.A. Slack: New materials and performance limits for thermoelectric cooling. In: *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC, Boca Raton 1995) p. 407
57.11 J.P. Heremans, M.S. Dresselhaus, L.E. Bell, D.T. Morelli: Nat. Nanotechnol. **8**, 471 (2013)
57.12 D.T. Morelli, G.P. Meisner: J. Appl. Phys. **77**, 3777 (1995)
57.13 G.S. Nolas, G.A. Slack: Am. Sci. **89**, 136 (2001)
57.14 R. Venkatasubramanian, E. Siivola, T. Colpitts, B. O'Quinn: Nature **413**, 597 (2001)
57.15 D.T. Morelli, J.P. Heremans, C.M. Thrush: Phys. Rev. B **67**, 035206 (2003)
57.16 A. Akhiezer: Fiz. Zh. **1**, 277 (1939)
57.17 G. Leibfried, E. Schloemann: Akad. Wiss. Göttingen, Math. Phys. Kl. A **4**, 71 (1954)
57.18 A.W. Lawson: J. Phys. Chem. Solids **3**, 155 (1957)
57.19 R. Berman: *Thermal Conduction in Solids* (Clarendon, Oxford 1976)
57.20 G.A. Slack: In: *Solid State Physics*, , ed. by F. Seitz, D. Turnbull, Vol. 34 (Academic, New York 1979) p. 1
57.21 I. Pomeranchuk: J. Phys. USSR **4**, 259 (1941)
57.22 P.G. Klemens: Proc. Phys. Soc. Sect. A **68**, 1113 (1955)
57.23 A.V. Ioffe, A.F. Ioffe: Izv. Akad. Nauk SSSR Ser. Fiz. **20**, 65 (1956)
57.24 B. Abeles: Phys. Rev. **131**, 1906 (1963)
57.25 R.C. Zeller, R.O. Pohl: Phys. Rev. B **4**, 2029 (1971)
57.26 D.G. Cahill, R.O. Pohl: Phys. Rev. B **35**, 4067 (1987)
57.27 L.D. Dudkin, N.K. Abrikosov: Sov. J. Inorg. Chem. **2**, 212 (1957)
57.28 L.D. Dudkin: Sov. Phys. Tech. Phys. **3**, 216 (1958)
57.29 L.D. Dudkin, N.K. Abrikosov: Sov. Phys. Solid State **1**, 126 (1959)
57.30 B.N. Zobrina, L.D. Dudkin: Sov. Phys. Solid State **1**, 1688 (1960)
57.31 W. Jeitschko, D. Braun: Acta Cryst. B **33**, 3401 (1977)
57.32 B.X. Chen, J.H. Xu, C. Uher, D.T. Morelli, G.P. Meisner, J.-P. Fleurial, T. Caillat, A. Borshchevsky: Phys. Rev. B **55**, 1476 (1997)
57.33 X. Shi, H. Kong, C.-P. Li, C. Uher, J. Yang, J.R. Salvador, H. Wang, L. Chen, W. Zhang: Appl. Phys. Lett. **92**, 182101 (2008)
57.34 L. Zhang, A. Grytsiv, P. Rogl, M. Zehetbauer: J. Phys. D Appl. Phys. **42**, 225405 (2009)
57.35 T.C. Harman, P.J. Taylor, M.P. Walsh, B.E. LaForge: Science **297**, 2229 (2002)
57.36 K.F. Hsu, S. Loo, F. Guo, W. Chen, J.S. Dyck, C. Uher, T. Hogan, E.K. Polychroniadis, M.G. Kanatzidis: Science **303**, 818 (2004)
57.37 H. Fleischmann: Z. Naturforsch. **16a**, 765 (1961)
57.38 T. Irie, T. Takahama, T. Ono: Jap. J. Appl. Phys. **2**, 72 (1963)
57.39 G.A. Slack: Phys. Rev. **105**, 832 (1957)
57.40 G.A. Slack, M.A. Hussain: J. Appl. Phys. **70**, 2694 (1991)
57.41 K. Biswas, J. He, I.D. Blum, C.-I. Wu, T.P. Hogan, D.N. Seidman, V.P. Dravid, M.G. Kanatzidis: Nature **489**, 414 (2012)
57.42 V.P. Zhuze, V.M. Sergeeva, E.L. Shtrum: Sov. Phys. Tech. Phys. **3**, 1925 (1958)
57.43 D.T. Morelli, V. Jovovic, J.P. Heremans: Phys. Rev. Lett. **101**, 035901 (2008)
57.44 E.J. Skoug, D.T. Morelli: Phys. Rev. Lett. **107**, 235901 (2011)
57.45 M.D. Nielsen, V. Ozolins, J.P. Heremans: Energy Environ. Sci. **6**, 570 (2013)
57.46 X. Lu, D.T. Morelli, Y. Xia, F. Zhou, V. Ozolins, H. Chi, X. Zhou, C. Uher: Adv. Energy Mater. **3**, 342 (2013)
57.47 X. Lu, D.T. Morelli: Phys. Chem. Chem. Phys. **1**(5), 5762 (2013)
57.48 X. Lu, D.T. Morelli: MRS Communications **3**, 129 (2013)
57.49 J. Heo, G. Laurita, S. Muir, M.A. Subramanian, D.A. Keszler: Chem. Mat. **26**, 2047 (2014)
57.50 W. Lai, Y. Wang, D.T. Morelli, X. Lu: Adv. Func. Mater. **25**, 3648 (2015)

58. Transparent Conductive Oxides

Hideo Hosono, Kazushige Ueda

Transparent conducting oxides (TCOs) such as doped ZnO, In_2O_3 and SnO_2 play important roles as transparent electrodes in commercial applications such as display and lighting devices. Although transparency and electrical conductivity are inherently conflicting, TCOs possess both properties simultaneously. To understand the fundamentals of TCOs, the essetials of the transparency and electrical conductivity are reviewed simply. Comprehension of the essentials enables us to develop novel TCOs following the principles of the materials design that carrier conduction paths should be constructed in wide-gap oxides. Because the electronic structure of oxides is considerably different between the conduction band and the valence band, procedures to form the conduction paths for electrons are in contrast to those for holes. In n-type TCOs, only isotropically spread ns^0 orbitals of heavy cations such as Zn^{2+}, In^{3+} and Sn^{4+} are necessary to form conduction paths for electrons at the bottom of the conduction band. In p-type TCOs, however, hybridization of orbitals between oxygen $2p^6$ orbitals and other orbitals such as Cu $3d^{10}$, Sn $5s^2$ and S $3p^6$ orbitals is essential to shape conduction paths for holes at the top of the valence band.

58.1 **Overview** 1391
58.1.1 Electrical Conduction 1391
58.1.2 Transparency 1393

58.2 **Materials Design for TCOs** 1395
58.2.1 Electronic Structures of Oxides 1396
58.2.2 Materials Design for n-Type TCOs 1396
58.2.3 Materials Design for p-Type TCOs 1398

58.3 **New Approach to Explore Candidate Materials: Materials Genome Approach** 1402

References 1403

58.1 Overview

Transparent conducting oxides (TCOs) are known as materials that show both transparent and conducting properties, and their films are utilized in flat panel displays, solar cells and electroluminescent devices as transparent electrodes as reviewed in [58.1–7]. The transparent properties mean TCOs show high optical transmission in the visible range, which requires an energy gap larger than 3.3 eV. The conducting properties signify that TCOs show high electrical conductivity in the range from 1 to $10^4\,S\,cm^{-1}$.

Although TCOs are common in industry nowadays, transparency and electrical conductivity are basically incompatible. Transparent materials are usually insulators such as SiO_2 glasses. If they are used in powder or polycrystalline bodies, they look white. On the other hand, highly conductive materials usually show metallic luster or deep dark color. For instance, Ag and Al metals show metallic color with high reflectivity, which are utilized in mirrors. Oxide high-temperature superconductors such as $YBa_2Cu_3O_7$ show black color. Therefore, to obtain TCOs, it is necessary to convert transparent insulators to highly conducting semiconductors by increasing electrical conductivity without coloration.

58.1.1 Electrical Conduction

Electrical current I in materials is expressed by Ohm's law as $I = V/R$, where V and R are potential difference and resistance respectively. When the equation is normalized by the sizes of materials, Ohm's law can be expressed as $\boldsymbol{J} = \sigma \boldsymbol{E}$, where $\boldsymbol{J}$, σ, $\boldsymbol{E}$ are current density, conductivity, and electric field strength. The conductivity σ is generally dependent on materials themselves and a useful parameter, which varies by a factor of over 10^{20}, to categorize materials into metals, semiconductors and insulators. Figure 58.1 shows the electrical conductivities of some materials. Metals

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_58

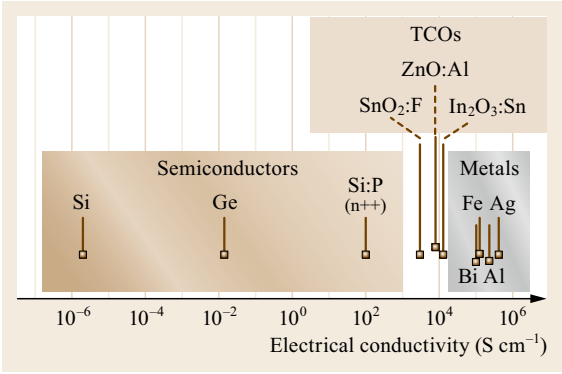


Fig. 58.1 Electrical conductivities of some conductive materials

Table 58.1 Electrical properties of Ag, Si:P and ITO

Materials	σ [a] (S cm^{-1})	n [b] (cm^{-3})	μ [c] (cm^2 V^{-1} s^{-1})
Ag	6.80×10^5	5.76×10^{22}	72
Si:P n	10^{-2}	10^{14}	1500
Si:P n++	10^2	10^{19}	100
ITO	10^4	10^{20}–10^{21}	20–80

[a] Conductivity, [b] Carrier concentration, [c] Mobility

such as Ag and Al possess high electrical conductivities of the order of 10^5 S cm^{-1}. Semiconductors such as Si and Ge have medium conductivities. Although pure semiconductors are almost insulating, impurity doping markedly increases the conductivities up to the order of 10^2 S cm^{-1}. The typical conductivities of TCOs are located in the middle of metals and doped semiconductors. This would be natural because TCOs are originally doped transparent oxide semiconductors (TOSs).

To understand the nature of the conductivities of TCOs, it is necessary to compare carrier concentration and mobility between some materials. The electrical conductivity is a product of carrier concentration n, elementary electric charge e, and mobility μ as shown in Fig. 58.2 and in (58.1).

$$\sigma = ne\mu \tag{58.1}$$

These parameters straightforwardly characterize conducting materials of metal, semiconductor and TCO. The parameters of typical materials are listed in Table 58.1 [58.17, 18]. Ag metal is found to show high electrical conductivity because of high carrier concentration in the order of approximately 10^{23} cm^{-3}. The mobility of Ag metal is not markedly high among these conductors. P-doped Si shows the lowest electrical conductivity among them because the carrier concentration is fairly low. It can become an electrical conductor simply because the carrier mobility of Si is much higher than that of Ag metal. Even if Si is heavily doped with the carrier concentration up to 10^{19} cm^{-3}, the mobility decreases by carrier scattering and the conductivity does not increase as high as Ag metal. Sn-doped In_2O_3 (ITO) shows the conductivity that is approxmately one-tenth of the Ag value. Because the carrier mobility of ITO is not largely different from that of Ag metal, the carrier concentration of ITO, which is one-tenth of the Ag value, is responsible for its smaller conductivity. These features can be extended to other conducting materials and their carrier

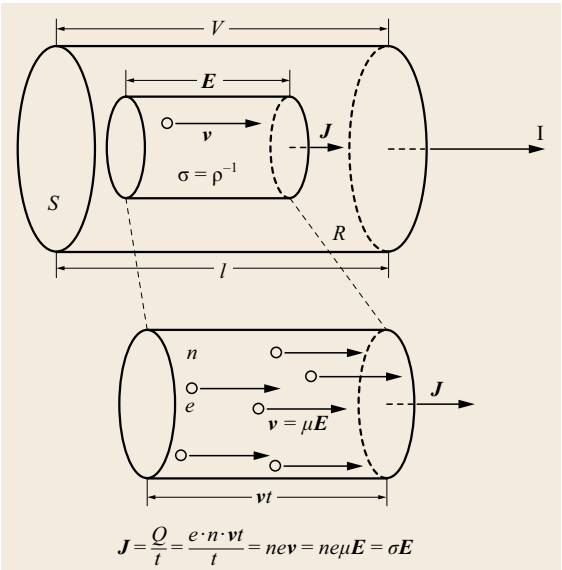


Fig. 58.2 Schematic illustration for Ohm's law

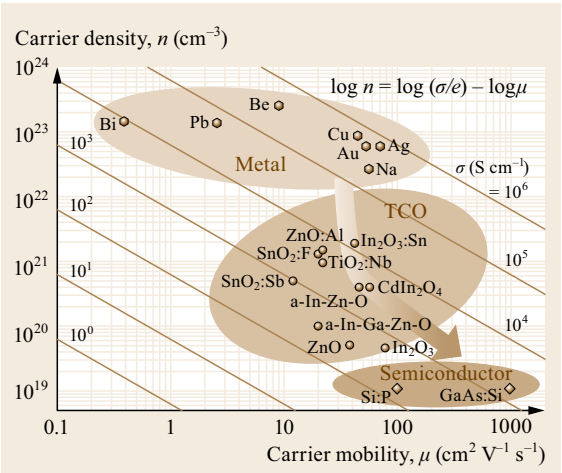


Fig. 58.3 Carrier density and mobility for some metals, semiconductors and TCOs. (After [58.8–16])

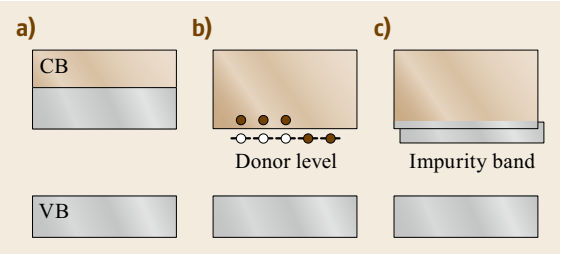


Fig. 58.4a–c Electronic structures of (a) metal; (b) semiconductors; (c) degenerate semiconductors

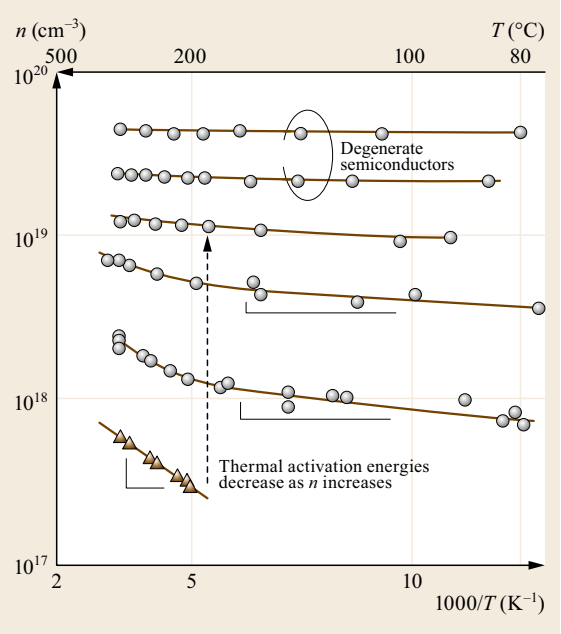


Fig. 58.5 Carrier concentration as a function of inversed temperature in ITO

concentrations and mobilities are plotted in Fig. 58.3. Metals are located in the *top-left* region in Fig. 58.3, which indicates the metals have extremely high carrier concentration but their mobilities are not so high. In contrast, doped semiconductors are located in the *bottom-right* region because of low carrier concentration and high mobility. Figure 58.3 clearly illustrates TCOs are located between metals and doped semiconductors. Because the mobilities of TCOs are not largely different from those of typical metals such as Ag, each TCO can be regarded as a metal with small carrier concentration, which is called as a degenerate semiconductor. The schematic electronic structure of a degenerate semiconductor is shown in Fig. 58.4. Although the thermal activation energy is necessary for semiconductors to generate carriers, it is unnecessary for metals and degenerate semiconductors. This

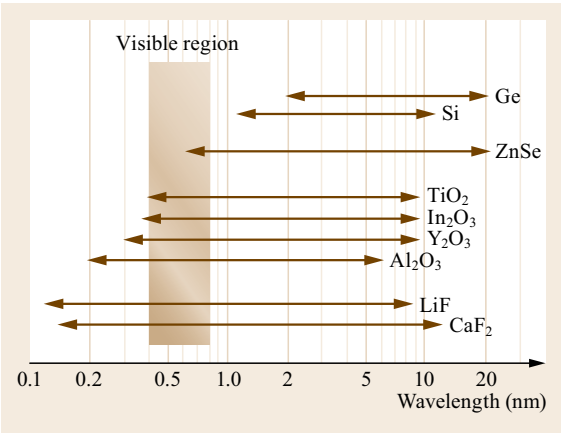


Fig. 58.6 Optically transparent regions in some materials

feature is experimentally observed as absence of thermal activation energy as shown in Fig. 58.5. Figure 58.5 shows the carrier concentration n as a function of reciprocal temperatures [58.19]. At increasing doping concentration, namely carrier concentration, the slope of the lines, which is derived from the activation energy for thermal ionization of dopant, becomes shallow and the conductivity finally becomes almost independent of temperatures.

58.1.2 Transparency

The main group metal oxides such as MgO and Al_2O_3 are electrically insulating but optically transparent in the wavelength range from 380 to 780 nm, namely the visible region. Therefore, oxide single crystals or glasses are frequently used as windows in optical applications. Figure 58.6 shows optical transparent ranges of several materials [58.20]. Typical oxides are transparent in the range approximately from 300 to 10 μm including the visible region. Typical fluorides such as LiF and CaF_2 show a much wider window range than oxides, especially in deep ultraviolet or vacuum ultraviolet regions [58.21]. On the other hand, conventional semiconductors such as Si and Ge have a window range only in the infrared region. Therefore, it is not surprising that TCOs are transparent in the visible region. Figure 58.7 shows a typical transmission spectrum of an oxide single crystal. As shown in Fig. 58.7, the absorption edge is controlled by the lattice vibration in the infrared region, and by the fundamental absorption caused by the electron transition from the valence band to the conduction band in the ultraviolet region.

When carriers are introduced into the parent material (undoped) of TCOs, optical absorption and reflection by the carriers newly appears. The collective motion of the conducting carriers behave as a kind

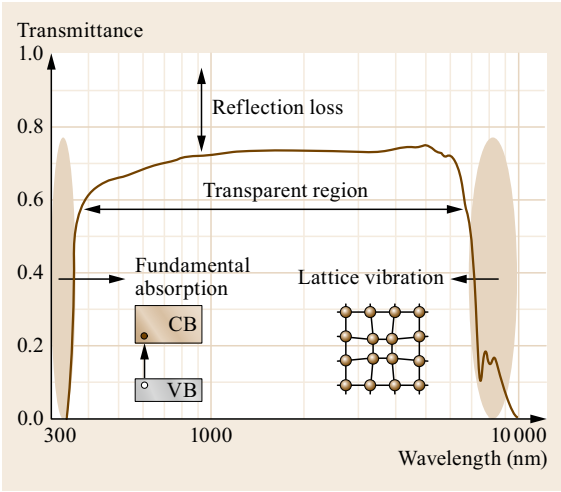


Fig. 58.7 Transmission spectrum of a typical oxide

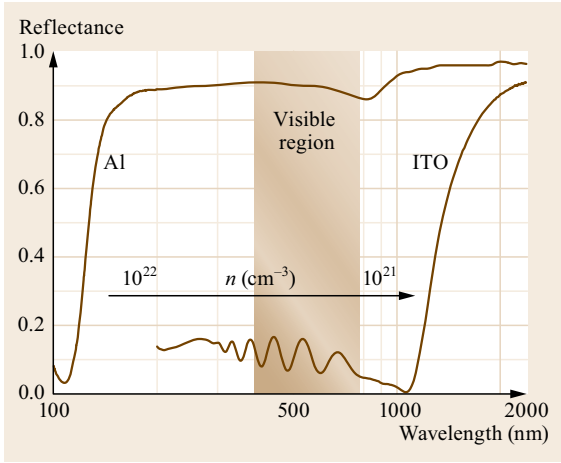


Fig. 58.8 Reflection spectra of Al and ITO

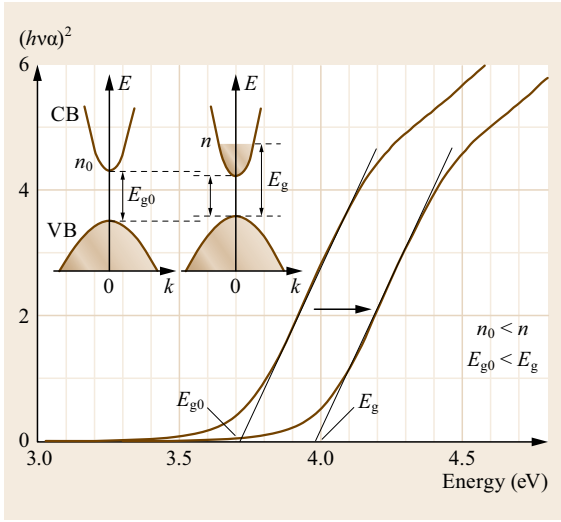


Fig. 58.9 Blue shift of fundamental absorption by band filling

of plasma in conductors. When light, namely an electromagnetic wave, is irradiated to the conductors, the carriers oscillate at the frequency of the light, which is called plasma oscillation. This results in the reflection of the light at the surface of the conductors. However, against light with higher frequency, the carriers cannot catch up with the fast electric field oscillation of the light. The light is transmitted through the conductors without causing the plasma oscillation or reflection at the surface. The maximum threshold frequency that enables the plasma oscillation is called the plasma frequency ω_p and can be expressed by (58.2) [58.17]. The plasma frequency ω_p can be converted to the wavelength λ_p as shown in

$$\omega_p = \frac{ne^2}{\varepsilon_0 m}\,, \tag{58.2}$$

$$\lambda_p = \frac{2\pi c}{e}\sqrt{\frac{\varepsilon_0 m}{n}}\,, \tag{58.3}$$

where n, m, ε_0 and c denote carrier concentration, electron rest mass, permittivity and speed of light in vacuum respectively. In Al metal with high carrier concentration, over $10^{22}\,\mathrm{cm}^{-3}$, the threshold λ_p is located in the vacuum ultraviolet region as shown in Fig. 58.8, and visible light is totally reflected at the surface [58.22]. This is the origin of the metallic luster of the Al metal, and why the wide-ranged reflection of Al is often used as a mirror. On the other hand, since ITO has a carrier concentration one order of magnitude smaller than Al metal, the λ_p moves to the infrared region, which results in the transparency in the visible region because ITO has a sufficient bandgap ($> 3.5\,\mathrm{eV}$) to transmit visible light. Consequently, the maximum carrier concentration of TCOs should be less than $10^{21}\,\mathrm{cm}^{-3}$ because the carriers make the optical windows of oxides narrower in the infrared region.

The carriers influence the optical windows of oxides not only in the infrared side but also in the ultraviolet side. Energy gaps are usually estimated by the $(h\nu\alpha)^2$–E plot, where h, ν and α are Planck constant, frequency and absorption coefficient respectively, as shown in Fig. 58.9. As the carrier concentration increases, absorption edges move to the higher energy side, enlarging apparent energy gaps and making optical windows wider in the ultraviolet region. This variation is attributed to the occupation of the conduction band by carrier electrons as shown in the *inset* in Fig. 58.9. In n-type TCOs, because carrier electrons

occupy the bottom of the conduction band, the optical bandgap energy E_g needed for the optical transition from the valence band to the unoccupied conduction band becomes larger than the original bandgap E_{g0}. This energy shift is called band filling or Bernstein moss (BM) shift.

58.2 Materials Design for TCOs

The essential factor of materials design for TCOs is determining how to dope carriers into oxides, because the metal oxides (except transition metals with an open shell electronic configuration) are basically transparent due to the large bandgap. As shown in (58.1), to increase the electrical conductivity is to increase carrier concentration and mobility. Because the carrier concentration can be controlled by users depending on purposes, the carrier concentration is generally independent from materials themselves. Therefore, the mobility is the material-dependent parameter to be designed. The mobility is a velocity of carriers under unit electric field strength and inversely proportional to the effective mass m^* of carriers as in (58.4).

$$\mu = \frac{v}{E} = \frac{e\tau}{m^*} \tag{58.4}$$

The effective mass m^* is theoretically deduced so that m^* is inversely proportional to the curvature of an energy band in the energy diagram as a function of wavenumber k as in (58.5)

$$\frac{1}{m^*} = \left(\frac{2\pi}{h}\right)^2 \frac{\partial^2 E}{\partial k^2} \tag{58.5}$$

In a simple one-dimensional crystal with lattice constant a, a simple energy band is formed and composed of a one-dimensionally aligned atomic orbital ϕ_m as shown in Fig. 58.10 [58.23]. The energy E of the band is expressed in (58.6) and (58.7) by using the orbital interaction or resonant integral H_{mn} between neighboring orbitals.

$$H_{mn} = \int \phi_m^* \hat{H} \phi_n \mathrm{d}x \,, \tag{58.6}$$

$$\begin{aligned} E &= H_{nn} + 2H_{mn}\cos(ka) \\ &\approx H_{nn} + 2H_{mn} - H_{mn}(ka)^2 \quad (x \approx 0) \end{aligned} \tag{58.7}$$

Equations (58.4)–(58.7) lead to a simple conclusion that the mobility is proportional to the orbital interaction H_{mn}

$$\mu \propto \frac{\partial^2 \mathrm{E}}{\partial k^2} = -2H_{mn}a^2 \,. \tag{58.8}$$

The value of H_{mn} is simply estimated so that H_{mn} becomes large when the overlap of atomic orbitals between ϕ_m and ϕ_n is large, because H_{mn} is calculated from a product of the atomic orbitals as seen in (58.6). Because making large overlap of the orbitals is just equal to increasing a probability of carrier transport or forming a conduction path in the crystal, carriers move between the orbitals through their overlapped areas. Therefore, a smooth conduction path for carriers must be constructed in wide-gap oxides by making sufficient overlap of component orbitals to provide them with large carrier mobility.

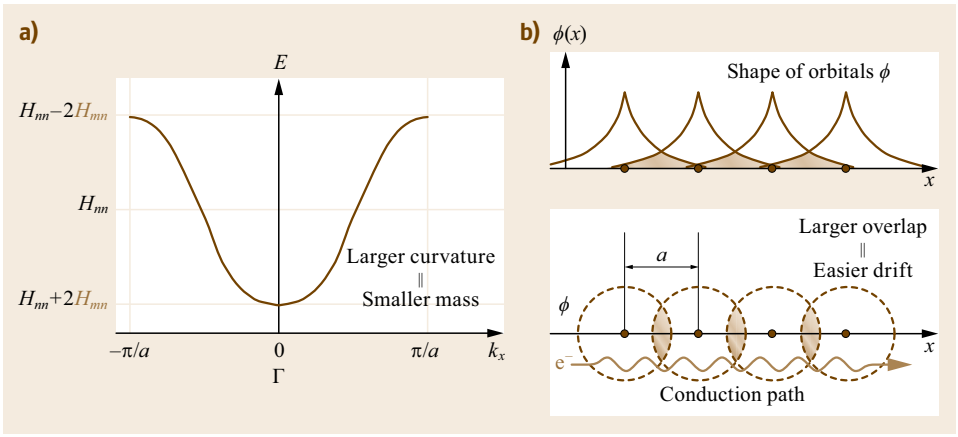


Fig. 58.10a,b Band dispersion in k space **(a)** and overlap of orbitals in real space **(b)**

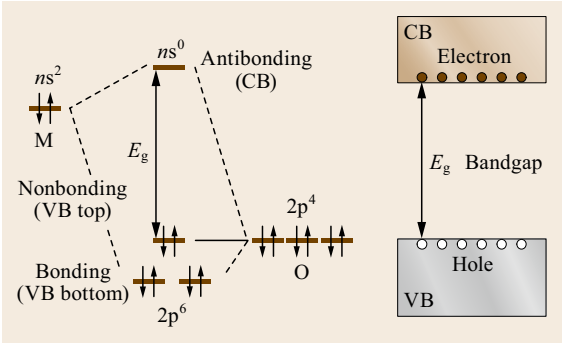


Fig. 58.11 Schematic molecular orbital diagram and band structure of oxides

	10	11	12	13	14	15
2				5 B	6 C	7 N
3				13 Al	14 Si	15 P
4	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As
5	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb
6	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi

Fig. 58.12 Key elements for TCOs in the periodic table

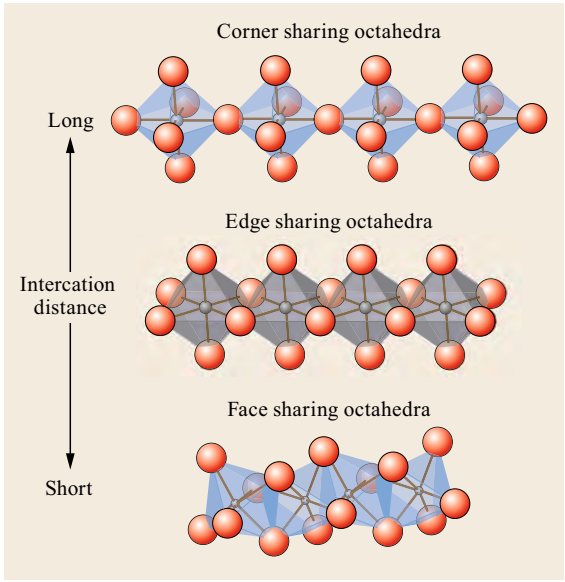


Fig. 58.13 Configuration of octahedra in crystal structures

58.2.1 Electronic Structures of Oxides

Understanding of the electronic structures of simple oxides is inevitable for the materials design for TCOs. Figure 58.11 shows the typical electronic structure of metal oxide (MO) illustrated in the form of molecular orbitals and energy bands. The conduction band is composed of empty metal ns^0 orbitals and the valence band is composed of occupied oxygen 2p orbitals. The conduction band has antibonding features, and the bottom of the valence band has bonding features as a counterpart. The top of the valence band has nonbonding features because nonbonding oxygen 2p orbitals remain without changing the energies.

58.2.2 Materials Design for n-Type TCOs

For n-type TCOs, a conduction band consisting of empty orbitals becomes a conduction path for electrons. To obtain the sufficient overlap of orbitals for the conduction path, cations' orbitals should be extended as widely as possible and intercation distances should be as short as possible. Therefore, the essentials for developing n-type TCOs are to choose proper cations for widely extended orbitals, and to choose a proper crystal structure for shorter intercation distances [58.24–26].

In the choice of cations, p-block atoms should be chosen for widely extended orbitals. As known in a simple hydrogen atomic model, radii of orbitals are proportional to principal quantum number n, namely, the period in the periodic table in Fig. 58.12. Although heavy atoms with large n are preferable, atoms with $n = 6$ such as Tl and Pb are inadequate because their orbitals are too widely extended to keep the wide bandgap for transparency. Transition metal ions with d^n open shells are also improper due to the coloration. In addition to the radii of orbitals, the shape of orbitals is also important to maximize the orbital overlap. Isotropic s orbitals, which are free from bonding direction, are more preferable than p- or d orbitals. Therefore, cations with $d^{10}ns^0$ $(n = 4, 5)$ electronic configuration are favorable to form the conduction path for electrons in the conduction band. Actually, the selected cations includes cations such as Zn^{2+}, In^{3+} and Sn^{4+} found in major TCOs.

In the choice of crystal structures, the configuration of polyhedra consisting of a cation at the center and oxygen ions at the apexes should be examined. In the case of the cations with $d^{10}ns^0$ $(n = 4, 5)$ such as In^{3+} and Sn^{4+}, they usually form oxygen octahedra due to the size of the ions. Various configurations such as isolated octahedra, corner-sharing octahedra, edge-sharing and face-sharing octahedra are seen in crystal structures. Among them, intercation distances are the shortest in face-sharing structures as shown in Fig. 58.13. It is known, however, that such struc-

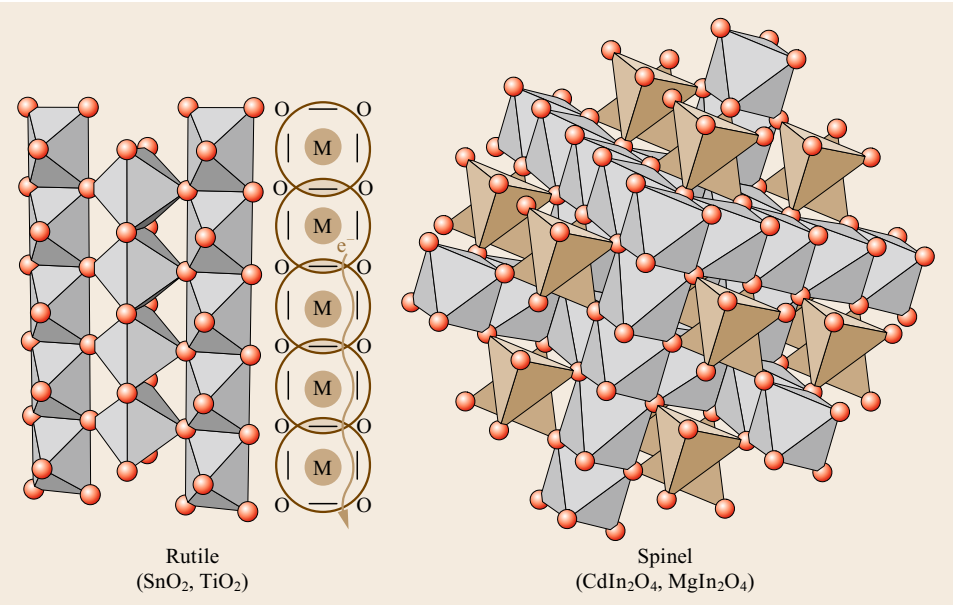


Fig. 58.14
Crystal structures of typical n-type TCOs

tures are usually unstable in oxides due to the intercation coulomb repulsion [58.27] and hardly seen in many crystal structures. Therefore, the edge-sharing structure substantially gives the shortest intercation distances. Accordingly, with widely extended orbitals, edge-sharing structures can form conduction paths in n-type TCOs.

Typical n-Type TCOs and Their Electronic Structure

From the concept of the materials design for n-type TCOs, the oxides, which include the cations with $d^{10}ns^0$ $(n = 4, 5)$ configuration and the partial crystal structures with edge-sharing octahedra, can be leading candidates. Actually, such materials often show transparent n-type conductive properties and some typical n-type TCOs are shown in Fig. 58.14. SnO_2 with a rutile structure composed of edge-sharing octahedra along the *c*-axis has been known as an n-type TCO for a long time. $CdIn_2O_4$ and $MgIn_2O_4$ with spinel structure have three-dimensionally connected edge-sharing octahedra [58.28, 29]. Although the effective ionic radii of In^{3+} and Sn^{4+} are 0.8 and 0.7 Å respectively, such ionic sizes neglect $5s^0$ orbitals. Interatomic distances are 3.2 Å in In metal and 3.0 Å in Sn metal, while intercationic distances are 3.3 Å in In_2O_3 and 3.2 Å in SnO_2. Because the intercationic distances in the oxides are very close to the interatomic distances in the metals, it would be easily supposed that $5s^0$ orbitals between In^{3+} or Sn^{4+} ions in the oxides can largely overlap as in the metals.

The large overlap of the orbitals in In_2O_3 and SnO_2 appears as a large energy dispersion of the conduction band in their energy diagrams as shown in Fig. 58.15 [58.30, 31], where the valence band maximum is set to zero in energy. The large energy dispersion results in the large curvature and large mobility at the bottom of the conduction band. These characteristic conduction band bottoms can be experimentally observed in the photoemission spectra of Fig. 58.16. Because the Fermi energy E_F is taken as zero in energy in Fig. 58.16, E_F is located at the bottom of the conduction band indicating these materials are n-type TCOs. The large dispersion of the conduction band causes a gentle increase of the density of states, which appears as a gradual increase in the photoemission intensity above E_F. The gradual increase at the bottom of the conduction band is in contrast to the steep increase at the top of the valence band, especially in In_2O_3.

Other n-Type TCOs

The characteristics of the crystal and electronic structures preferable to n-type TCOs are commonly seen in other n-type TCOs such as ZnO, Ga_2O_3, Cd_2SnO_4 and $SrGeO_3$ [58.32–36]. The edge-sharing octahedra in crystal structures are not always necessary if the overlap of the orbitals for the conduction band is large enough to form a conduction path. Therefore, even amorphous oxides such as a-In_2O_3-ZnO and a-In_2O_3-Ga_2O_3-ZnO can be n-type TCOs because the conduction paths are formed by the orbitals' overlap in these materials irrespective of the crystal structures [58.37–39]. Since

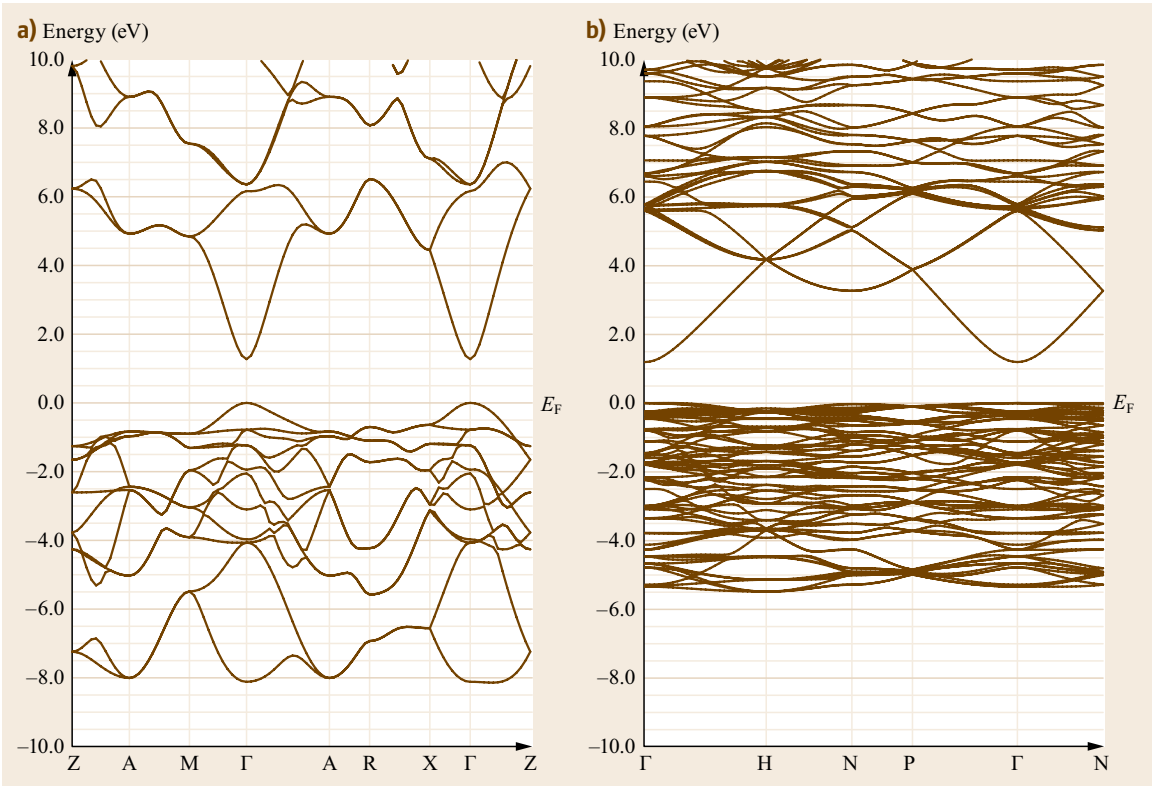


Fig. 58.15a,b Energy band diagrams of SnO_2 **(a)** and In_2O_3 **(b)**

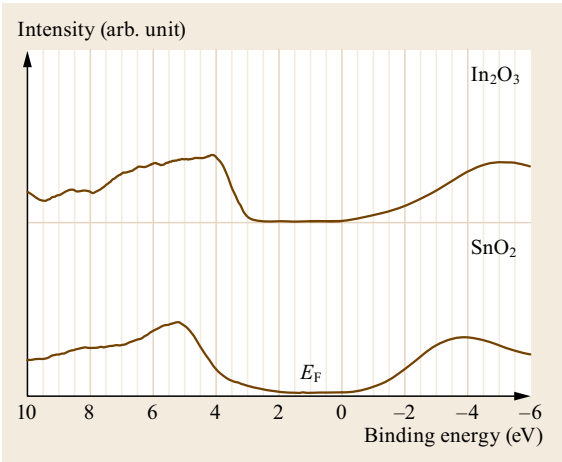


Fig. 58.16 Normal and inverse photoemission spectra of SnO_2 and In_2O_3

the amorphous TCO thin films can be formed at low temperatures and durable during bending, they can be deposited on not only in inorganic glasses but also in organic plastic plates or films. Their advantages are utilized in growing applications such as flexible electronics and printable electronics.

Some TCOs such as TiO_2 and $KTaO_3$ have the conduction bands that are not composed of s^0 orbitals but d^0 orbitals [58.40–42]. These materials are considered to have conduction paths consisting of the d^0 orbitals. Because the d^0 orbitals are not isotropic in shape, the electrical conductivity would be largely influenced by the symmetry of the crystals.

58.2.3 Materials Design for p-Type TCOs

In p-type TCOs, in contrast to n-type TCOs, it is the filled valence bands rather than the the empty conduction bands that form the conduction paths for carriers, namely positive holes. However, the formation of the hole conduction paths are not as easy as that of the electron conduction paths. This is why p-type TCOs are rare [58.43].

The valence bands of oxides are generally composed of filled oxygen 2p orbitals. Accordingly, the oxygen 2p orbitals must be sufficiently overlapped with each other to form the conduction paths for holes. However, the oxygen 2p orbitals at the valence band maximum consist of nonbonding states, which indicate almost no interaction, namely no orbitals overlap between oxygen ions. This makes it very difficult to

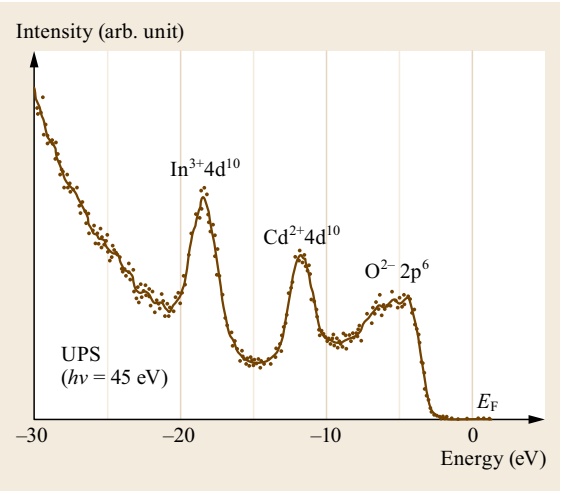


Fig. 58.17 Ultraviolet photoemission spectrum of $CdIn_2O_4$

form the hole conduction paths at the valence band, resulting in considerably small hole mobility even if the holes are doped into the valence band. Hole doping is also substantially difficult in oxides because oxygen has large electronegativity indicating that O^{2-} ions with closed shells are too stable to accept holes. Actually, the energy of oxygen 2p orbitals is very deep from the vacuum level, giving oxygen the large electron affinity. Therefore, it is considered that the formation of a hole conduction path by using only oxygen 2p orbitals is almost impossible.

To improve the nature of the low hole mobility and difficulty in hole-doping in oxides, chemical modulation of the valence band was proposed by hybridizing cations' orbitals with oxygen 2p orbitals [58.43, 44]. The hybridization of orbitals usually occurs between orbitals with similar energies. Therefore, the energy of the candidate cations' orbitals must be close to that of oxygen 2p orbitals. Figure 58.17 shows the photoemission spectrum of a $CdIn_2O_4$ n-type TCO. As similar to In_2O_3 and SnO_2 in Fig. 58.16, the oxygen $2p^6$ band consists of the valence band. In addition to the oxygen 2p valence band, Cd^{2+} $4d^{10}$ and In^{3+} $4d^{10}$ bands are located at higher binding energies with almost the same energy intervals. Because the Cd and In are located at adjacent positions on periodic table as in Fig. 58.12, the periodicity predicts that the energy of Ag^{+} $4d^{10}$ or Cu^{+} $3d^{10}$ orbitals are very close to the oxygen 2p energy and they are easily hybridized with the oxygen 2p orbitals. From the viewpoint of chemical stability, Cu $3d^{10}$ orbitals are considered to be the most favorable for the chemical modulation of the valence band.

Hybridization between Cu $3d^{10}$ and oxygen 2p nonbonding orbitals newly forms Cu–O bonding and antibonding states, and the Cu–O antibonding states become the top of the valence band as shown in Fig. 58.18 [58.45]. Therefore, the valence band is mainly composed of a Cu–O bonding band (A), an M–O bonding band (B), a Cu nonbonding band (C) and a Cu–O antibonding band (D). The chemical modulation by Cu $3d^{10}$ orbitals varies the nonbonding oxygen 2p orbitals to the Cu–O antibonding band at the top of the valence band and elevates the energy of the valence band maximum simultaneously. Accordingly, the shallow Cu–O antibonding band is expected to increase the hole mobility and make hole doping easier.

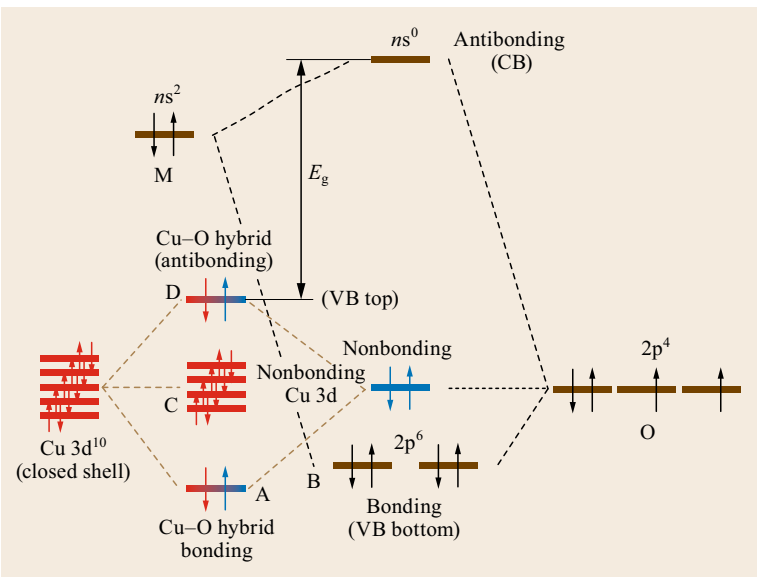


Fig. 58.18 Chemical modulation of the valence band by Cu $3d^{10}$ orbitals in oxides shown in a schematic molecular orbital diagram

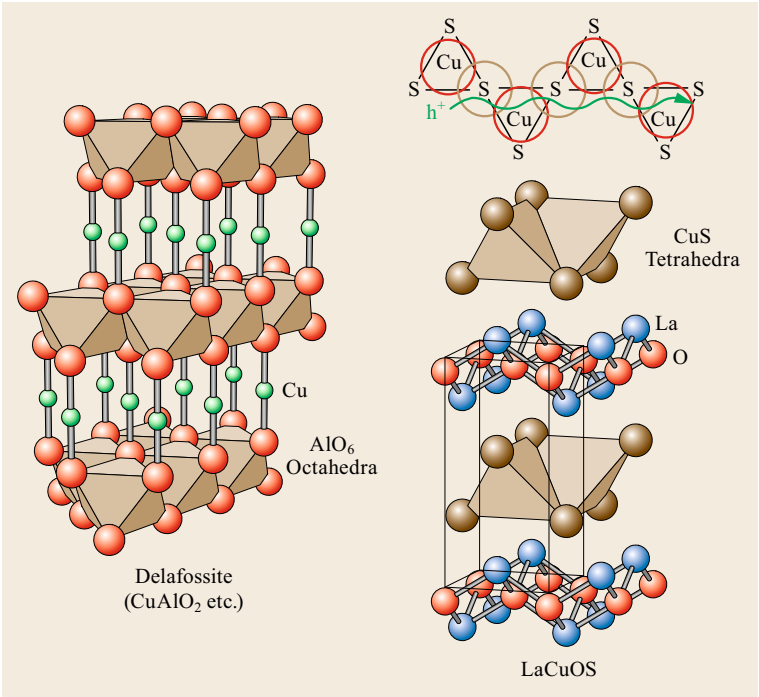


Fig. 58.19 Crystal structures of typical p-type TCOs

The chemical modulation works more effectively in oxysulfides, in which Cu 3d orbitals hybridize not oxygen 2p orbitals but sulfur 3p orbitals forming Cu–S bonding (A′) and antibonding (D′) band [58.46]. Because sulfur 3p orbitals are shallower in energy and more covalent than oxygen 2p orbitals, larger hybridization is expected at the top of the valence band. The larger modulation leads to higher hole mobility by forming a steady hole conduction path at the valence band maximum.

The chemical modulation usually narrows the bandgap by the formation of the shallow antibonding band at the top of the valence band. Because the degree of the modulation is usually dependent on the dimension of crystal structures, lower dimensional crystal structure gives a moderate hybridization, namely a moderately shallow antibonding band. Therefore, layered structures and chain structures are preferable to maintain wide bandgap and transparency.

Typical p-Type TCOs and Their Electronic Structure

From the materials design for p-type TCOs, the Cu^+ containing oxides with layered structures or chain structures can be leading candidates. $CuAlO_2$ with delafossite structure was proposed as a prototype of p-type TCOs [58.43, 47]. $CuAlO_2$ is a layered structure composed of Cu^+ and AlO_6 octahedra layers as shown in Fig. 58.19. Although p-type Cu_2O, in which an O^{2-} ion bonds to four Cu^+ ions, shows deep dark red color, $CuAlO_2$ is almost transparent in the visible region because an O^{2-} ion in $CuAlO_2$ bonds to only one Cu^+ and three Al^{3+} ions resulting in a moderate Cu–O hybridization. Another example of a transparent p-type material is LaCuOS, which is also a layered structure composed of CuS and LaO layers [58.48]. The hybridization between Cu 3d and S 3p orbitals occurs in the CuS layer separated by the LaO layer. Because the LaO layers are inserted between CuS layers reducing dimensionality, the layered structure brings about a moderate Cu–S hybridization to maintain transparency.

The chemical modulation at the top of the valence band in $CuAlO_2$ and LaCuOS is seen in the energy diagram of Fig. 58.20 [58.47, 49, 50]. In contrast to the energy diagram of the valence band for n-type TCOs, especially In_2O_3, dispersed bands are observed in the energy range from 0 to −2 eV above several flat bands densely gathered at around −3 eV. The dispersed bands and flat bands are derived from the Cu–O/S antibonding bands and Cu nonbonding bands respectively. The dispersed bands are considered to form a hole conduction path and increase hole mobility at the top of the valence band. These features in the valence band can be experimentally observed in the photoemission spectra of Cu_2O, $CuGaO_2$ delafossite and LaCuOS as shown in Fig. 58.21. The Fermi energy E_F is located

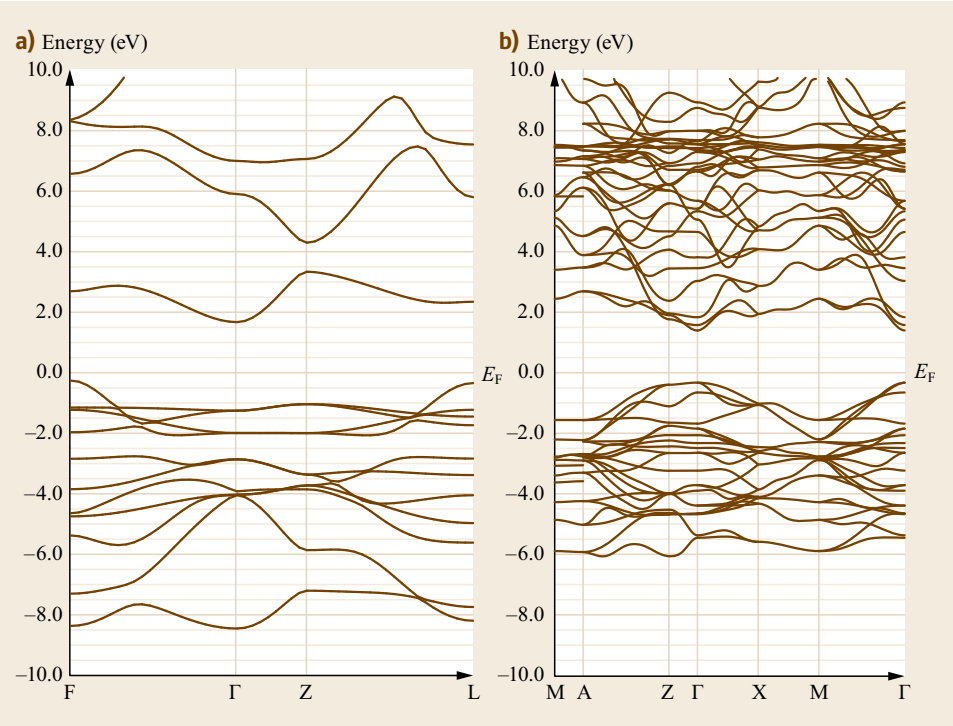


Fig. 58.20a,b Energy band diagrams of $CuAlO_2$ delafossite **(a)** and LaCuOS **(b)**

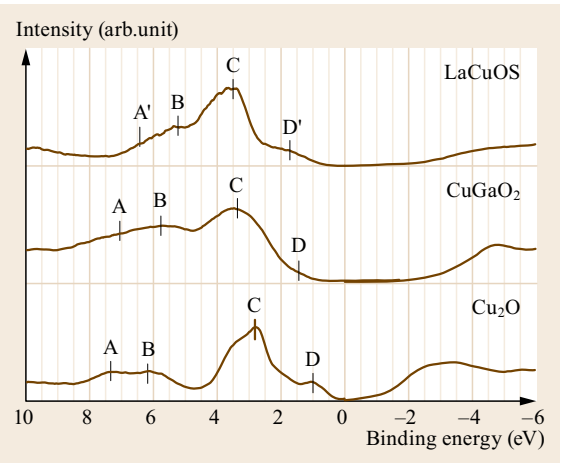


Fig. 58.21 Normal and inverse photoemission spectra of Cu_2O, $CuGaO_2$ and LaCuOS

near the top of the valence band indicating these materials show p-type conductions. The narrow bandgap of Cu_2O is also observed in contrast to wide bandgaps of $CuGaO_2$ and LaCuOS. In the valence band, the main four bands A–D, explained in molecular orbital diagram of Fig. 58.18, can be found in all materials. The Cu–O hybridized band D is located just below E_F and the presence of the band D demonstrates that the chemical modulation by Cu $3d^{10}$ orbitals is effective in the materials design of p-type TCOs. The typical electrical properties of some p-type materials are listed in Table 58.2.

Table 58.2 Electrical properties of Cu^+-based p-type TCOs

Materials	σ [a] ($S\,cm^{-1}$)	μ [b] ($cm^2\,V^{-1}\,s^{-1}$)	E_g (eV)	Reference
$CuAlO_2$	0.34	0.13	3.5[d], 1.8[c]	[58.47]
$CuGaO_2$	0.063	0.23	3.6[d]	[58.51]
$SrCu_2O_2$	0.048	0.46	3.3[c]	[58.52]
LaCuOS	5.9	0.5	3.1[c]	[58.53, 54]
LaCuOSe	910	8.0	2.8[c]	[58.54, 55]
BaCuSeF	43	–	3.0[c]	[58.56]

[a] Conductivity, [b] Mobility, [c] Direct transition, [d] Indirect transition

Other p-Type TCOs

The concept of the chemical modulation by Cu $3d^{10}$ orbitals is also applicable to other Cu-containing materials such as $SrCu_2O_2$, LaCuOSe and BaCuFSe as listed in Table 58.2 [58.47, 51–56]. The crystal structure of $SrCu_2O_2$ has low dimensional zigzag chains of Cu–O bonds, which enables a moderate Cu–O hybridization and transparency. LaCuOSe and BaCuFSe have the similar crystal structures to LaCuOS, in which

CuSe and BaF layers replace CuS and LaO layers respectively.

From other concepts of materials design for p-type TCOs, SnO [58.57, 58] and $ZnRh_2O_4$ [58.59] have been developed. Sn^{2+} in SnO has a lone pair of $5s^2$ electrons, which mainly occupy the bottom of the valence band. However, partial densities of Sn 5s orbitals also compose the top of the valence band along with Sn 5p orbitals and O 2p orbitals. This feature is similar to the modulation by Cu 3d orbitals and can be interpreted that the top of the valence band is composed of their hybridized states. The bandgap of SnO is as small as 0.2–0.7 eV for indirect transition, but as large as 2.7–3.1 eV for direct transition. Although the conductivity and transparency is not as high as Cu^+-based materials, SnO shows transparent p-type conductivity of approximately $10^{-1}\,S\,cm^{-1}$.

$ZnRh_2O_4$ with spinel structure includes Rh^{3+} ions with open shells of $4d^6$, which is not preferable for transparency. Rh^{3+} ions occupy the octahedral sites in the spinel structure and are influenced by the crystal field. Because the crystal field in $ZnRh_2O_4$ is sufficiently strong, the Rh^{3+} 4d orbitals are no longer usual open shells but two largely separated bands of occupied $4d^6$ and empty $4d^0$ orbitals, namely t_{2g}^6 band and e_g^0 band, with the bandgap as large as 2.1 eV. The p-type conductivity by holes, which probably originate from cationic vacancies, is reported to be approximately $1\,S\,cm^{-1}$.

There are several controversial reports on novel p-type TCOs such as p-type SnO_2. The experimental determination of p- or n-type conduction is sometimes difficult when carrier concentration is extremely high/low or both electrons and holes are present in conductors. Theoretical verification such as evaluation of acceptor formation energy becomes important, as well as experimental demonstration, to identify novel TCOs [58.60].

58.3 New Approach to Explore Candidate Materials: Materials Genome Approach

The materials design of TCOs explained above is empirically deduced from simple theories and some experimental data on electronic structures. However, recent remarkable progress in computational ability has enabled the calculation of essential parameters for TCOs such as bandgaps and carrier effective masses in almost all oxides (except strongly electron correlated systems) accumulated in databases. Because the candidates for TCOs are extracted from materials databases, this process is a kind of materials genome approach, which is similar to the biomedical approach in the human genome project.

For instance, by targeting new p-type TCOs through high-throughput computations and screening, groups at Université Catholique de Louvain (IMCN) and the Massachusetts Institute of Technology predicted several candidates of novel p-type TCOs such as $K_2Sn_2O_3$ as shown in Fig. 58.22 [58.61]. The electronic structures of more than 3000 binary and ternary oxides in the inorganic crystal structure database (ICSD) were computed and oxides with exceptionally low effective mass less than 1.5 were selected. The bandgaps of the selected oxides were estimated more accurately by a calculation based on many-body perturbation theory. In addition to the calculation of bandgaps and hole effective masses, the p-type dopability was also evaluated by calculating the defect formation energies. From the viewpoint of the electronic structure and p-type dopability, several candidates were selected and two design principles for p-type TCOs were drawn; one is to hybridize cations' $(n-1)d^{10}ns^2$ $(n = 5, 6)$ orbitals such as Sn^{2+} $5s^2$ with oxygen $2p^6$ orbitals; and the other is to include anions' p orbitals such as S^{2-} and P^{3-} $3p^6$ orbitals, which are more delocalized than oxygen $2p^6$ orbitals. This approach looks promising and having applicability as methodology. However, one has to keep in mind that practical TCO materials have to meet various fundamental properties such as chemical stability and etching ability.

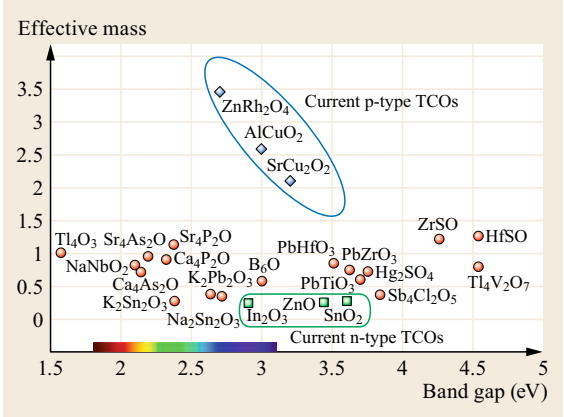


Fig. 58.22 Candidates for p-type TCOs selected from materials database through high-throughput computations

References

58.1 D.S. Ginley, H. Hosono, D.C. Paine: *Handbook of Transparent Conductors* (Springer, Berlin 2011)
58.2 D.S. Ginley, C. Bright: Transparent conducting oxides, MRS Bulletin **25**, 15 (2000)
58.3 H. Hosono: Thin Solid Films **515**, 6000 (2007)
58.4 K. Ellmer: Nat. Photonics **6**, 809 (2012)
58.5 C.G. Granqvist: Sol. Energy Mater. Sol. Cells **91**, 1529 (2007)
58.6 P.D.C. King, T.D. Veal: J. Phys.: Condens. Matter **23**, 334214 (2011)
58.7 S. Calnan, A.N. Tiwari: Thin Solid Films **518**, 1839 (2010)
58.8 T. Minami, H. Sato, H. Nanto, S. Takata: Jpn. J. Appl. Phys. **24**, L781 (1985)
58.9 I. Hamberg, C.G. Granqvist: J. Appl. Phys. **60**, R123 (1986)
58.10 M. Mizuhashi: Thin Solid Films **70**, 91 (1980)
58.11 H. Ohta, M. Orita, M. Hirano, H. Tanji, H. Kawazoe, H. Hosono: Appl. Phys. Lett. **76**, 2740 (2000)
58.12 T. Ishida, O. Tabata, J. Park, S.H. Shin, H. Magara, S. Tamura, S. Mochizuki, T. Mihara: Thin Solid Films **281/282**, 228 (1996)
58.13 X. Feng, J. Ma, F. Yang, F. Ji, F. Zong, C. Luan, H. Ma: Mater. Lett. **62**, 1779 (2008)
58.14 K. Budzynska, E. Leja, S. Skrzypek: Sol. Energy Mater. **12**, 57 (1985)
58.15 N. Ito, Y. Sato, P.K. Song, A. Kaijio, K. Inoue, Y. Shigesato: Thin Solid Films **496**, 99 (2006)
58.16 M. Orita, H. Ohta, M. Hirano, S. Narushima, H. Hosono: Philos. Mag. B **81**, 501 (2001)
58.17 C. Kittel: *Introduction to Solid State Physics*, 8th edn. (Wiley, New York 2005)
58.18 S.M. Sze: *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York 1981)
58.19 H.K. Müller: Phys. Status Solidi (b) **27**, 723 (1968)
58.20 W.D. Kingery, H.K. Bowen, D.R. Uhlmann: *Introduction to Ceramics*, 2nd edn. (Wiley, New York 1976)
58.21 W.R. Hunter, S.A. Malo: J. Phys. Chem. Solids **30**, 2739 (1969)
58.22 G. Hass, W.R. Hunter, R. Tousey: J. Opt. Soc. Am. **47**, 1070 (1957)
58.23 P.A. Cox: *The Electronic Structure and Chemistry of Solids* (Oxford Univ. Press, Oxford 1987)
58.24 H. Kawazoe, N. Ueda, H. Un'no, T. Omata, H. Hosono, H. Tanoue: J. Appl. Phys. **76**, 7935 (1994)
58.25 N. Ueda, H. Hosono, R. Waseda, H. Kawazoe: Appl. Phys. Lett. **71**, 933 (1997)
58.26 H. Kawazoe, K. Ueda: J. Am. Ceram. Soc. **82**, 3330 (1999)
58.27 L. Pauling: *The Nature of the Chemical Bond*, 3rd edn. (Cornell Univ. Press, New York 1960)
58.28 R.D. Shannon, J.L. Gillson, R.J. Bouchard: J. Phys. Chem. Solids **38**, 877 (1977)
58.29 N. Ueda, T. Omata, N. Hikuma, K. Ueda, H. Mizoguchi, T. Hashimoto, H. Kawazoe: Appl. Phys. Lett. **61**, 1954 (1992)
58.30 A.J. Freeman, K.R. Poeppelmeier, T.O. Mason, R.P.H. Chang, T.J. Marks: MRS Bulletin **25**, 45 (2000)
58.31 J. Robertson: Inst. Phys. Conf. Ser. **43**, 1183 (1979)
58.32 K. Ellmer, A. Klein, B. Rech (Eds.): *Transparent Conductive Zinc Oxide* (Springer, Berlin 2008)
58.33 T. Minami: Thin Solid Films **516**, 5822 (2008)
58.34 N. Ueda, H. Hosono, R. Waseda, H. Kawazoe: Appl. Phys. Lett. **70**, 3561 (1997)
58.35 A.J. Nozik: Phys. Rev. B **6**, 453 (1972)
58.36 H. Mizoguchi, T. Kamiya, S. Matsuishi, H. Hosono: Nat. Commun. **2**, 470 (2011)
58.37 T. Sasabayashi, N. Ito, E. Nishimura, M. Kon, P.K. Song, K. Utsumi, A. Kaijo, Y. Shigesato: Thin Solid Films **445**, 219 (2003)
58.38 H. Hosono, N. Kikuchi, N. Ueda, H. Kawazoe, K. Shimidzu: Appl. Phys. Lett. **67**, 2663 (1995)
58.39 H. Hosono: J. Non-Cryst. Solids **352**, 851 (2006)
58.40 D.C. Cronemeyer: Phys. Rev. **87**, 876 (1952)
58.41 Y. Furubayashi, T. Hitosugi, Y. Yamamoto, K. Inaba, G. Kinoda, Y. Hirose, T. Shimada, T. Hasegawa: Appl. Phys. Lett. **86**, 252101 (2005)
58.42 S.H. Wemple: Phys. Rev. **137**, A1575 (1965)
58.43 H. Kawazoe, H. Yanagi, K. Ueda, H. Hosono: MRS Bulletin **25**, 28 (2000)
58.44 H. Kawazoe, M. Yasukawa, H. Hyodo, M. Kurita, H. Yanagi, H. Hosono: Nature **389**, 939 (1997)
58.45 H. Yanagi, H. Kawazoe, A. Kudo, M. Yasukawa, H. Hosono: J. Electroceram. **4**, 407 (2000)
58.46 K. Ueda, H. Hiramatsu, M. Hirano, T. Kamiya, H. Hosono: Thin Solid Films **496**, 8 (2006)
58.47 H. Yanagi, S. Inoue, K. Ueda, H. Kawazoe, H. Hosono: J. Appl. Phys. **88**, 4159 (2000)
58.48 K. Ueda, S. Inoue, S. Hirose, H. Kawazoe, H. Hosono: Appl. Phys. Lett. **77**, 2701 (2000)
58.49 B.J. Ingram, T.O. Mason, R. Asahi, K.T. Park, A.J. Freeman: Phys. Rev. B **64**, 155114 (2001)
58.50 K. Ueda, H. Hosono, N. Hamada: J. Phys.: Condens. Matter **16**, 5179 (2004)
58.51 K. Ueda, T. Hase, H. Yanagi, H. Kawazoe, H. Hosono, H. Ohta, M. Orita, M. Hirano: J. Appl. Phys. **89**, 1790 (2001)
58.52 A. Kudo, H. Yanagi, H. Hosono, H. Kawazoe: Appl. Phys. Lett. **73**, 220 (1998)
58.53 H. Hiramatsu, K. Ueda, H. Ohta, M. Orita, M. Hirano, H. Hosono: Appl. Phys. Lett. **81**, 598 (2002)
58.54 H. Hiramatsu, K. Ueda, H. Ohta, M. Orita, M. Hirano, T. Kamiya, H. Hosono: Appl. Phys. Lett. **82**, 1048 (2003)
58.55 H. Hiramatsu, K. Ueda, H. Ohta, M. Hirano, M. Kikuchi, H. Yanagi, T. Kamiya, H. Hosono: Appl. Phys. Lett. **91**, 012104 (2007)
58.56 H. Yanagi, J. Tate, S. Park, C.H. Park, D.A. Keszler: Appl. Phys. Lett. **82**, 2814 (2003)
58.57 Y. Ogo, H. Hiramatsu, K. Nomura, H. Yanagi, T. Kamiya, M. Hirano, H. Hosono: Appl. Phys. Lett. **93**, 032113 (2008)
58.58 N.F. Quackenbush, J.P. Allen, D.O. Scanlon, S. Sallis, J.A. Hwelett, A.S. Nandur, B. Chen, K.E. Smith, C. Weiland, D.A. Fischer, J.C. Woicik, B.E. White, G.W. Watson, L.F.J. Piper: Chem. Mater. **25**, 3114 (2013)

58.59 H. Mizoguchi, M. Hirano, S. Fujitsu, T. Takeuchi, K. Ueda, H. Hosono: Appl. Phys. Lett. **80**, 1207 (2002)
58.60 D.O. Scanlon, G.W. Watson: J. Mater. Chem. **22**, 25236 (2012)
58.61 G. Hautier, A. Miglio, G. Ceder, G.-M. Rignanese, X. Gonze: Nat. Commun. **4**, 2292 (2013)

59. Inorganic Perovskite Oxides

Tatsumi Ishihara

Crystal structure and important functions of inorganic perovskite oxides are introduced. Perovskite oxides comprise large families among the structures of oxide compounds, and several perovskite-related structures are currently recognized. Typical structures (ABO_3) consist of large-sized 12-coordinated cations at the A-site and small-sized 6-coordinated cations at the B-site. Several complex halides and sulfides and many complex oxides have a perovskite structure. From a variety of compositions and structures, a variety of functions are observed in perovskite oxides. In particular high electronic conductivity, which is at a similar level as metal, and surface activity to oxygen dissociation, are highly attractive in this oxide. Perovskite oxide is now widely used for solid oxide fuel cells.

59.1 **Typical Properties of Perovskite Oxides** 1409
59.1.1 Dielectric Properties 1409
59.1.2 Electrical Conductivity and Superconductivity 1409
59.1.3 Catalytic Activity 1410
59.2 **Photocatalytic Activity** 1411
59.3 **Application for Solid Oxide Fuel Cells (SOFCs)** 1413
59.3.1 Cathode 1413
59.3.2 Anode 1414
59.3.3 Electrolyte 1415
59.3.4 Interconnector 1417
59.4 **Oxygen Separating Membrane** 1418
59.5 **Summary** 1419
References 1420

Oxide groups consisting of two or more different cations are called complex or mixed oxides, and many types of crystal structures are known. In some special cases, oxides consisting of single cations in different oxidation states are also classified as mixed oxides. For example, in Eu_3O_4, the mixed oxide consists of Eu(III) and Eu(II) in 6- or 8-coordination respectively. However, the most typical structure of a mixed oxide consists simply of two or more different cations with different oxidation states, ionic radii, and coordination numbers. This diversity, which comes from the complexity of these structures, results in a larger number of different properties as compared to those of simple oxides. One of the most well known and important complex oxide structures is the spinel structure (AB_2O_4), which shows important magnetic properties. The structure of such oxides displays a most interesting complexity. Since the size of the A and B ions in this structure is close, oxides of this type are typical examples of the versatility of mixed oxides.

Another important structure of mixed oxide is perovskite and a variety of related structures are classified as this oxide. The typical chemical formula of the perovskite structure is ABO_3, where A and B denotes two different cations. The ilmenite structure has the same composition as the perovskite one, i. e., ABO_3; however, A and B in this structure are cations of approximately the same size, which occupy an octahedral site. Therefore, in spite of the fact that they share the same general chemical formula, structures classified as ilumenite- or ilmenite-related structures (e.g., $LiSbO_3$) are different from perovskite.

Perovskite oxides comprise large families among the structures of oxide compounds, and several perovskite-related structures are currently recognized. Typical structures consist of large-sized 12-coordinated cations at the A-site and small-sized 6-coordinated cations at the B-site. Several complex halides and sulfides and many complex oxides have a perovskite structure. In particular, $(Mg,Fe)SiO_3$ or $CaSiO_3$ is thought to be the predominant compound in the geosphere [59.1, 2]. Perovskite compounds with different combinations of charged cations in the A and B-sites, for example $1+5$, $2+4$ and $3+3$, have been discovered. Even more complex combinations are observed, such as $Pb(B'_{1/2}B''_{1/2})O_3$, where $B' = Sc$, Fe and $B'' = Nb$, Ta, or $La(B'_{1/2}B''_{1/2})O_3$, where $B' = Ni$, Mg, etc. and $B'' =$ Ru(IV) or Ir(IV). In addition, many ABO_3 compounds

© Springer International Publishing AG 2017
S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_59

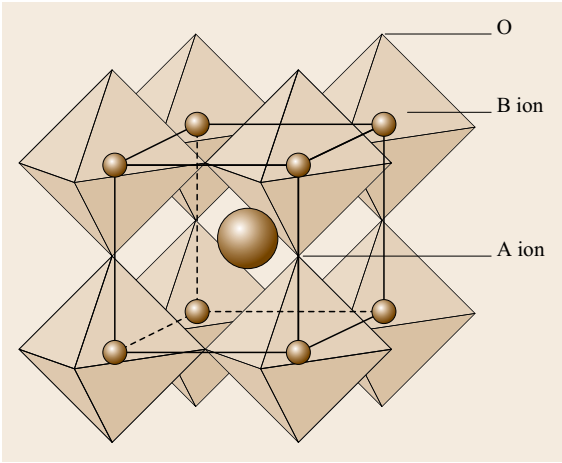


Fig. 59.1 Ideal cubic perovskite structure

crystallize in polymorphic structures, which show only a small distortion from the most symmetrical form of the perovskite structure.

The ideal structure of perovskite, which is illustrated in Fig. 59.1, is a cubic lattice. Although few compounds have this ideal cubic structure, many oxides have slightly distorted variants with lower symmetry (e.g., hexagonal or orthorhombic). Furthermore, even though some compounds have ideal cubic structure, many oxides display slightly distorted variants with lower symmetry. Several examples of perovskite oxides are listed in Table 59.1, where it is clear that a large number of perovskite oxides have a rombohedral lattice. Additionally, in many compounds a large extent of oxygen or cation deficiency has been observed. Due to the large lattice energy, many compounds are classified as perovskite oxides in spite of the large cation and/or oxygen deficiencies. There are various types of distortions in the perovskite structure that are strongly related to their properties, in particular their ferromagnetic or ferroelectricity.

In order to understand the deviations from the ideal cubic structure, these ABO_3 oxides are first regarded as purely ionic crystals. In the case of the ideal structure, the following relationship between the radii of the A, B, and O^{2-} ions holds true

$$r_A + r_O = \sqrt{2(r_B + r_O)}\,.$$

Therefore, the deviation from the ideal structure in perovskite oxides can be expressed through the following so-called *tolerance factor t*

$$t = \frac{(r_A + r_O)}{\sqrt{2(r_B + r_O)}}$$

Table 59.1 Typical peroskite compound

Compound	Lattice parameter (Å)		
	a	b	c
Cubic structure			
$KTaO_3$	3.989		
$NaTaO_3$	3.929		
$NaNbO_3$	3.949		
$BaMnO_3$	4.040		
$BaZrO_3$	4.193		
$SrTiO_3$	3.904		
$KMnF_3$	4.189		
$KFeF_3$	4.121		
Tetragonal structure			
$BiAlO_3$	7.61		7.94
$PbSnO_3$	7.86		8.13
$BaTiO_3$	3.994		4.038
$PdTiO_3$	3.899		4.153
$TlMnCl_3$	5.02		5.04
$LaAlO_3$ type			
$LaAlO_3$	5.357	$\alpha = 60°06'$	
$LaNiO_3$	5.461	$\alpha = 60°05'$	
$BiFeO_3$	5.632	$\alpha = 60°06'$	
$KNbO_3$	4.016	$\alpha = 60°06'$	
$GdFeO_3$ type			
$GdFeO_3$	5.346	5.616	7.668
$YFeO_3$	5.283	5.592	7.603
$NdGaO_3$	5.426	5.502	7.706
$CaTiO_3$	5.381	5.443	7.645
$NaMgF_3$	5.363	5.503	7.676

In perovskite-type compounds, the value of t lies between approximately 0.80 and 1.10. It is noted that the oxides with the lower t values crystallize in the ilmenite structure, which is a polymorph of the perovskite structure. It seems superfluous to say that for the ideal cubic structure, the value of t is close to 1 or at least higher than 0.89. Figure 59.2 shows the crystal groups for $A^{2+}B^{4+}O_3$ and $A^{3+}B^{3+}O_3$ combinations, which are related to deviation from the ideal structure [59.3]. As the value of t decreases, the structure of the unit lattice is shifted from cubic to triclinic due to the increased distortions. Figure 59.3 shows chemical elements that can be accommodated within the perovskite structure. It is evident that almost all elements except for noble gases can occupy either A or B lattice positions in the perovskite structure, including dopants. The stability and the crystal group is mainly determined by the ratio of the ionic radii of the A and B cations. Indeed, the structure is dependent not only on the size but also on the nature of the A and B atoms. For example, $AMnO_3$ compounds crystallize in the perovskite structure when A cation is La or Ce–Dy, whereas a new hexagonal structure with 5- and 7-coordination of Mn and A respectively, is formed when A = Ho–Lu or Y if A = La

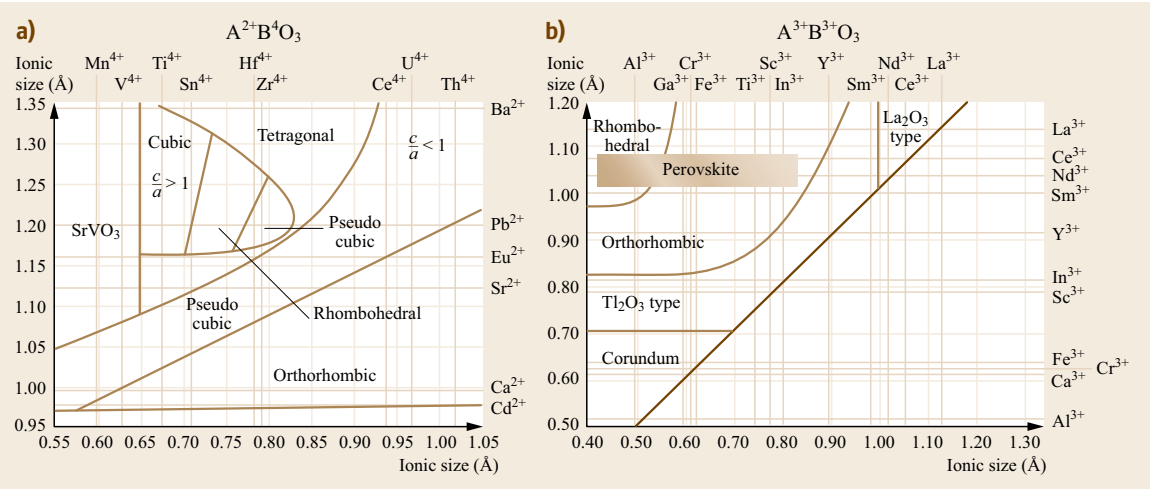


Fig. 59.2a,b The effect of ionic size of A- and B-site cations on the observed distortions of the perovskite structure; **(a)** $A^{2+}B^{4+}O_3$ case **(b)** $A^{3+}B^{3+}O_3$ case

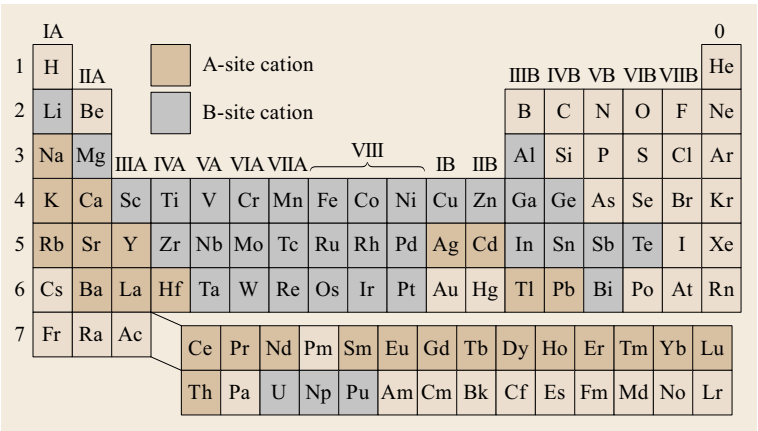


Fig. 59.3 Chemical elements that can occupy cation sites in the perovskite structure

or Ce–Dy [59.4]. Here, the attention should be paid to the nature of the B atom, where the nature of the bond is highly covalent, and therefore the coordination number is lower than 6. The typical example of this type structure is $BaGeO_3$. In spite of the t value close to $t = 1$, i. e., ideal ionic size combination, $BaGeO_3$ crystallizes not in the perovskite structure but in the silicate-related one. This is attributable to the fact that the preferred coordination number of Ge is 4. On the other hand, due to the progress in high pressure technology, the synthesis of new Ge-based perovskite oxides has been reported [59.5]. Since the coordination number of Ge increases with the pressure, the perovskite structures with higher coordination numbers are preferred, and a typical example of this is $CaGeO_3$. Another group of interesting perovskite compounds is the oxynitrides, i. e., $LaWO_{3-x}N_x$, $LaTaO_2N$, etc. Therefore, the value of t, which is determined by the ionic size, is an important index for the stability of perovskite structures; however, the contribution of the chemical nature, such as coordinating number of the constituent elements, needs to be taken into account.

The formation of superstructures in the perovskites is discussed next. If a B-site cation is progressively replaced by a dopant, a large difference in ionic radii tends to lead to the formation of the superstructures rather than random arrangements of the two kinds of ions. The typical case of this is Ba_2CaWO_6, which is regarded as $Ba_2(CaW)O_6$. Similarly, in the compounds with the general formula $Ba_3MTa_2O_9$, there is random distribution of M and Ta ions in the octahedral positions when M is Fe, Co, Ni, Zn, or Ca, whereas the formation of the superstructure with a hexagonal lattice is observed in $Ba_3SrTa_2O_9$. Another interesting type of superstructure observed in the perovskite is the ordering of cation vacancies located on A-sites: for example

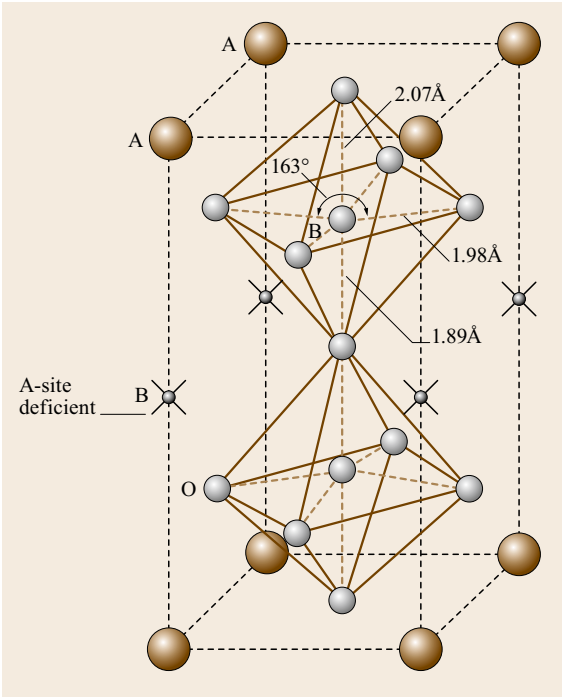


Fig. 59.4 Structure of $LaNb_3O_9$, an A-site-deficient perovskite oxide

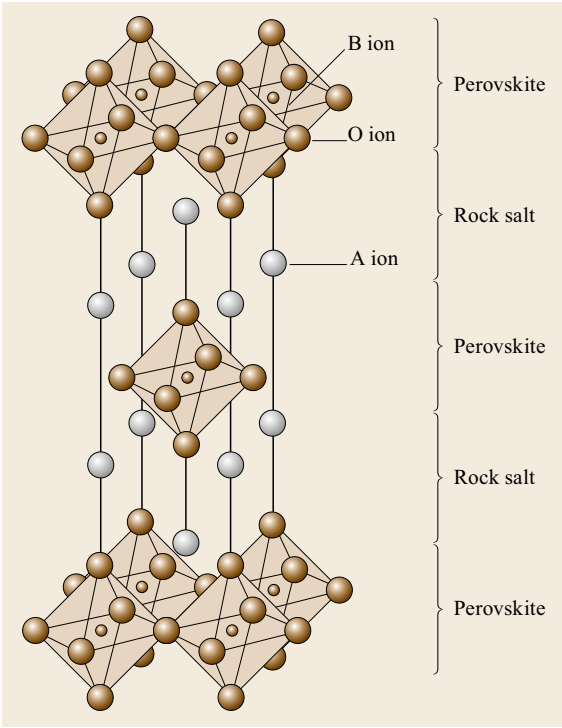


Fig. 59.5 K_2NiF_4 structure, a perovskite-related structure

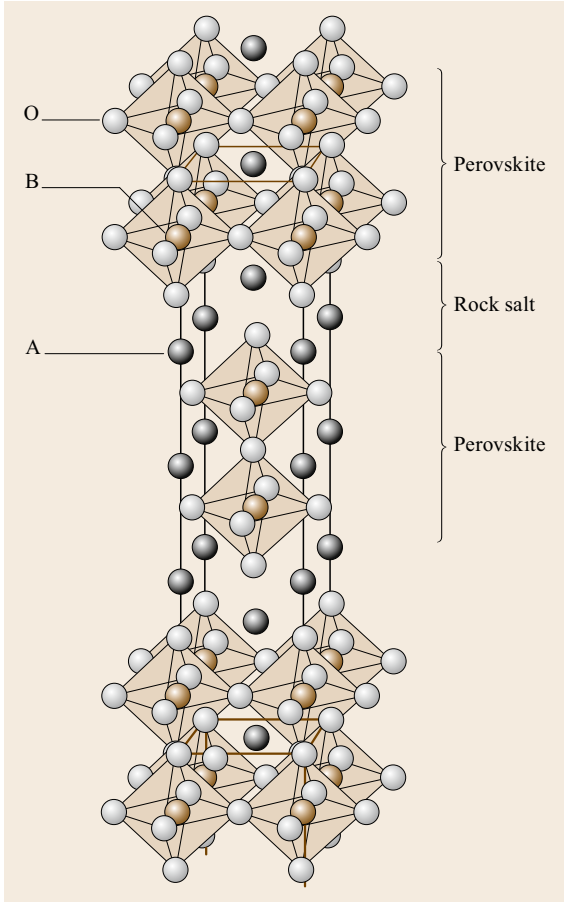


Fig. 59.6 Ruddelsden–Popper structure, another type of perovskite-related structure

MNb_3O_9 (M = La, Ce, Pr, Nb) and MTa_3O_9 (M = La, Ce, Pr, Nd, Sm, Gd, Dy, Ho, Y, Er). In these oxides, there is an octahedral framework of the ReO_3 type with incomplete occupancy of the 12-fold coordinated A-sites. Figure 59.4 shows the structure of $LaNb_3O_9$. The B-sites of the perovskite structure are occupied by Nb ion and two-thirds of the A-sites remain vacant.

Other typical polymorphs of the perovskite structure are Brownmillalite ($A_2B_2O_5$) and K_2NiF_4 structures. Brownmillalite ($A_2B_2O_5$) is an oxygen-deficient type of perovskite where oxygen vacancy is ordered. The unit cell contains BO_6 and BO_4 units in an ordered arrangement. Due to the oxygen deficiency, the coordination number of A-site cations decreases to 8. The lattice parameter of the Brownmillalite structure relates to cubic lattice parameter (a_p) of the ideal perovskite as $a = b = \sqrt{2}a_p$, $c = 4a_p$. Cu-based oxides or Ni-based oxides tend to adopt these oxygen-deficient structures because of the large amount of oxygen defects.

A combination of ordered B-sites and oxygen defects is seen in K_2NiF_4 structures, which are well-known as they show superconducting properties. The K_2NiF_4 structures consist of two units; a $KNiF_3$ perovskite unit and a KF rock salt unit (Fig. 59.5), which are connected in series along the *c* axis. Since the rock salt structure is embedded into the *c* axis direction, the K_2NiF_4 compound shows strong two-dimensional properties. Based on the intergrowth of the different number of $KNiF_3$ and KF units, there are many structures called Ruddelsden–Popper compounds with the general formula $(ABO_3)_nAO$ (Fig. 59.6), i. e., $Sr_3Ti_2O_7$ ($n = 2$), $Sr_4Ti_3O_{10}$ ($n = 3$). It is interesting to compare the isostructural Sr_2TiO_4 or Ca_2MnO_4 with $SrTiO_3$ or $CaMnO_3$, which crystallize in the perovskite structures. Two different A cations forming the perovskite and the rock salt units are also possible, and $LaO \cdot nSrFeO_3$ is the typical example of this arrangement. Another interesting variant of these K_2NiF_4 structures is when two different anions occupy the two building blocks exclusively, i. e., $SrFeO_3 \cdot SrF$ or $KNbO_3 \cdot KF$. In any case, it is evident that perovskite oxides comprise a large family of oxides. As a result, a variety of crystal structures and properties is expected in these compounds. For further detailed discussion on the perovskite-related oxides, the reader is referred to references [59.5–8].

59.1 Typical Properties of Perovskite Oxides

Due to the variety of structures and chemical compositions, perovskite oxides exhibit a large variety of properties. The well-known properties of the perovskite oxides are ferroelectricity in $BaTiO_3$-based oxides and superconductivity in $Ba_2YCu_3O_7$ etc. In addition to these well-known properties, several perovskite oxides exhibit good electrical conductivity, which is close to those of metals, ionic conductivity as well as mixed ionic and electronic conductivity. Based on these variations in the electrical conducting property, perovskite oxides are chosen as the components for solid oxide fuel cells (SOFCs). It is also well known that several perovskite oxides exhibit high catalytic activity with respect to various reactions, in particular oxidation reactions [59.9]. Table 59.2 provides examples of the typical properties of perovskite oxides. In this section, several typical properties of the perovskite oxides, namely ferroelectricity, magnetism, superconductivity, and catalytic activity, are briefly overviewed.

59.1.1 Dielectric Properties

Ferroelectricity, piezoelectricity, electrostriction, and pyroelectricity are special properties inherent to dielectric materials, and are important properties of electroceramics. The most well-known property of perovskite oxides is ferroelectric behavior, where $BaTiO_3$, $PdZrO_3$, and their doped compounds are representative examples. The study of the ferroelectricity in $BaTiO_3$ has long history, and many detailed reviews have been published. Furthermore, since the ferroelectric behavior of $BaTiO_3$ has a strong relationship with the crystal structure, detailed studies of the crystal structure have been reported for $BaTiO_3$. $BaTiO_3$ undergoes mainly three phase transformations from, i. e., monoclinic to tetragonal and cubic as the temperature increases. Above 303 K, $BaTiO_3$ crystallizes in the cubic perovskite structure, which does not show ferroelectric behavior. The high dielectric constant observed in $BaTiO_3$ can be explained by the basis of the anisotropy of the crystal structure.

59.1.2 Electrical Conductivity and Superconductivity

One of the most well-known properties of perovskite oxides is superconductivity. In 1984, superconductivity was first reported by *Bednorz* and *Müller* in La–Ba–Cu–O perovskite oxide [59.10]. After their report, much attention has been paid to new types of high-temperature oxide superconductors, mainly Cu-based oxides. As a result, several superconducting oxides with different A-site cations have been discovered. However, the presence of Cu on the B-site is found to be essential

Table 59.2 Typical properties of perovskite oxides

Typical property	Typical compound
Ferromagnetism	$BaTiO_3$, $PdTiO_3$
Piezoelectricity	$Pb(Zr, Ti)O_3$, $(Bi, Na)TiO_3$
Electrical conductivity	ReO_3, $SrFeO_3$, $LaCoO_3$, $LaNiO_3$, $LaCrO_3$
Superconductivity	$La_{0.9}Sr_{0.1}CuO_3$, $YBa_2Cu_3O_7$, $HgBa_2Ca_2Cu_2O_8$
Ion conductivity	$La(Ca)AlO_3$, $CaTiO_3$, $La(Sr)Ga(Mg)O_3$, $BaZrO_3$, $SrZrO_3$, $BaCeO_3$
Magnetism	$LaMnO_3$, $LaFeO_3$, La_2NiMnO_6
Catalytic properties	$LaCoO_3$, $LaMnO_3$, $BaCuO_3$
Electrode materials	$La_{0.6}Sr_{0.4}CoO_3$, $La_{0.8}Ca_{0.2}MnO_3$

for superconductivity to occur. High-temperature oxide superconductors of the $YBa_2Cu_3O_7$ system [59.11] and the $Bi_2Sr_2Ca_2Cu_3O_{10}$ system [59.12] were reported in 1987 and 1988 respectively, and currently the critical temperature of the superconducting transition (T_c) has been further increased to 130−155 K in the $HgBa_2Ca_2Cu_3O_{8+\delta}$ system [59.13]. Since all high-temperature superconducting oxides are cuprites (Cu-based oxides), superconductivity is clearly related to the Cu–O layers. The critical temperature for superconductivity, T_c, is related to the number of Cu–O layers in the crystal structure:

1. Cu–O layer $T_c \approx 30\,K$
2. Cu–O layers $T_c \approx 90\,K$
3. Cu–O layers $T_c \approx 110\,K$
4. Cu–O layers $T_c \approx 120\,K$.

It is expected that further increase in the number of Cu–O layers may result in the higher T_c values. However, due to the low chemical stability, synthesis of five or more Cu–O layered compounds has not been successful so far. $YBa_2Cu_3O_7$ is one of the most important superconductor systems with high T_c, and detailed studies of its crystal structure have been performed. Also, the content of oxygen nonstoichiometry is an important factor for high T_c. When the value of d is smaller than 0.5, $YBa_2Cu_3O_7$ crystallizes in an orthorhombic structure, which is superconductive, while for $d > 0.5$, $YBa_2Cu_3O_7$ has a tetragonal structure, which does not exhibit superconductivity.

In addition to superconductivity, there are many perovskite oxides showing high electronic conductivity, which is close to those of metals like Cu. The typical examples of such perovskite oxides are $LaCoO_3$, $LaFeO_3$, and $LaMnO_3$, which are now commonly used as cathodes in SOFCs. These perovskite oxides show superior hole conductivity, which is as high as $\sigma = 100\,S/cm$ originated from excess oxygen [59.14, 15]. Doping of aliovalent cations on the A-site is also highly effective in enhancing the electrical conductivity due to the increased number of mobile charge carriers generated by the charge compensation.

59.1.3 Catalytic Activity

Because of the variety of component elements and the high chemical stability, perovskite oxides have also been extensively studied as catalysts for various reactions. Two types of research trends clearly emerged from the above reasons. The objective of the first one is the development of the oxidation catalysts or oxygen-activated catalysts as an alternative to catalysts containing precious metals. The second trend regards perovskite as a model for active sites. The stability of the perovskite structure allows preparation of compounds with unusual valence states of elements or a high extent of oxygen deficiency. It is also noted that the high catalytic activity of perovskite oxides is based partially on the high surface activity to oxygen reduction or oxygen activation due to the large number of oxygen vacancies presented.

Among the various catalytic reactions studied, the ones applicable to environmental catalysis (e.g., automobile exhaust gas cleaning catalysts) attract particular attention. Initially, it was reported that perovskite oxide consisting of Cu, Co, Mn or Fe exhibited superior activity to NO direct decomposition at higher temperature [59.16–18]. The direct NO decomposition reaction ($2NO = N_2 + O_2$) is one of the *dream reactions* in the catalysis field. In this reaction, the ease in the removal of surface oxygen as a product of the reaction plays an important role, and due to the facility of oxygen deficiency present, perovskite oxides are active with respect to this reaction at high temperatures. It is pointed out that the doping is highly effective in enhancing NO decomposition activity. Under an oxygen enriched atmosphere (up to 5%), a relatively high NO decomposition activity was reported for $Ba(La)Mn(Mg)O_3$ perovskite [59.19].

Recently, another interesting application of perovskite oxides as an automobile catalyst has been reported, namely the so-called *intelligent catalyst* [59.20]. Up to now, three-way Pd–Rh–Pt catalysts have been widely used for the purpose of removal of NO, CO, and uncombusted hydrocarbons. In order to decrease the amount of precious metals, the catalyst consisting of fine particles with high surface-to-volume ratio is required. However, these fine particles are not stable under the operating conditions and easily sinter, resulting in the deactivation of the catalyst. In order to maintain a high dispersion state, the redox property of perovskite oxides has been proposed, i. e., under oxidation conditions, palladium is oxidized and exists as $LaFe_{0.57}Co_{0.38}Pd_{0.05}O_3$, and under reducing conditions, palladium is deposited as fine metallic particles with a radius of 1−3 nm. This cycling of the catalyst through oxidizing and reducing conditions results in the partial substitution of Pd into, and deposits from, the perovskite framework, thus maintaining a high dispersion state of Pd. This was found to be highly effective in improving the long-term stability of Pd during removal of NO_x, CO, and hydrocarbons from the exhaust gas. The high dispersion state of Pd can be recovered by exposing the catalyst to oxidation and reduction environments. As a result this catalyst is called an *intelligent* catalyst. This unique property also originates from the high stability of the perovskite crystal structure and

charge compensation is automatically done by redox couples of another cation in the lattice.

Another interesting application of perovskite oxide is as a photocatalyst for water splitting. Among the catalysts for water splitting into H_2 and O_2, it is reported that Ta- or Nb-based perovskite oxide shows high activity by using ultraviolet light. This will be introduced in detail in the next section.

59.2 Photocatalytic Activity

Photo-excited electrons and holes can be used for splitting water into H_2 and O_2 and this reaction is attracting much interest for converting solar energy to hydrogen. Various inorganic catalysts have been studied as photocatalysts for water splitting, in particular, Pt/TiO_2 is a well-known inorganic semiconductor for photocatalysis. Among the various catalysts reported, in this section, photocatalysts based on perovskite structure are briefly introduced. The Ta-based oxide is generally active in the photocatalytic water splitting reaction [59.22]. In particular, the Ta-based perovskite oxide, $ATaO_3$ (A = alkaline cation) shows high activity in water splitting [59.23]. The activity is strongly affected by the A cation and this is because crystal structure is related to the electronic configuration of the oxide. The bond angles of Ta–O–Ta are 143° ($LiTaO_3$), 163° ($NaTaO_3$), and 180° ($KTaO_3$). As the bond angle is close to 180°, migration of excited energy in the crystal occurs more easily and the band gap becomes smaller. Therefore, the order of the delocalization of excited energy is $LiTaO_3 < NaTaO_3 < KTaO_3$, while that of the band gap is reversed in the order as shown in Fig. 59.7. Table 59.3 shows photocatalytic activities for water splitting into H_2 and O_2 in pure water on alkali tantalite photocatalysts with and without NiO cocatalysts. $NaTaO_3$ photocatalysts showed the highest photocatalytic activity when NiO cocatalysts were loaded. In this case, excess sodium in the starting material was indispensable for showing the high activity [59.23]. The conduction band level of the $NaTaO_3$ photocatalyst was higher than that of NiO (-0.96 eV) as shown in Fig. 59.7 [59.23]. Moreover, the excited energy was delocalized in the $NaTaO_3$ crystal. Therefore, the photogenerated electrons in the conduction band of the $NaTaO_3$ photocatalyst were able to transfer to the conduction band of the NiO cocatalyst of an active site for H_2 evolution, resulting in the enhancement of the charge separation. Therefore, NiO loading was effective for the $NaTaO_3$ photocatalyst even without special pretreatment.

It is reported that the activity of photocatalytic water splitting is also much increased by additives. For

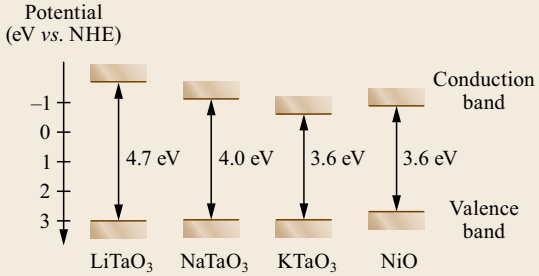


Fig. 59.7 Band structures of alkali tantalates $ATaO_3$ (A : Li, Na and K) with perovskite-type structures in comparison to a normal hydrogen electrode (NHE). (After [59.21])

Table 59.3 Photocatalytic activities for water splitting into H_2 and O_2 in pure water on alkali tantalite photocatalysts with and without NiO cocatalysts. (After [59.21])

Catalyst[a]	Ratio of alkali to Ta[b]	Band gap[c] (eV)	Surface area ($m^2\,g^{-1}$)	Activity ($\mu mol\,h^{-1}$)	
				H_2	O_2
$LiTaO_3$	1.05	4.7	0.3	430	220
NiO (0.10 wt%)/ $LiTaO_3$	1.05	4.7	–	98	52
$NaTaO_3$	1.00	4.0	0.5	11	4.4
NiO (0.05 wt%)/ $NaTaO_3$	1.00	4.0	–	480	240
$NaTaO_3$	1.05	4.0	0.4	160	86
NiO (0.05 wt%)/ $NaTaO_3$	1.05	4.0	–	2180	1100
$KTaO_3$	1.10	3.6	1.6	29	13
NiO (0.10 wt%)/ $KTaO_3$	1.10	3.6	–	7.4	2.9

[a] Catalyst: 1 g, pure water: 350 ml, 400 W high-pressure mercury lamp, inner irradiation cell made of quartz
[b] In starting materials
[c] Estimated from the onset of absorption

example, Table 59.4 shows band gaps, surface areas and photocatalytic activities for water splitting into H_2 and O_2 in pure water on various lanthanide-doped $NaTaO_3$ (denoted as $NaTaO_3$:Ln hereafter) with NiO cocatalysts [59.21]. $NaTaO_3$:Ln powders had larger surface areas than that of nondoped $NaTaO_3$. Band gaps of $NaTaO_3$:Ln are slightly higher than that of nondoped $NaTaO_3$. The activity of the NiO/$NaTaO_3$ photocatalyst is remarkably improved by doping of Ln, except for Eu and Yb. In particular, obviously, NiO/$NaTaO_3$:La is the most active: H_2 and O_2 evolved steadily and efficiently as shown in Fig. 59.8. The optimized NiO(0.2 wt%)/$NaTaO_3$:La(1.5%) photocatalyst evolves H_2 and O_2 with the rates of 14.6 and 7.2 m mol h^{-1} respectively. The apparent quantum yield is approximately 50% at 270 nm. Thus, it is demonstrated that the photocatalytic water splitting is able to proceed efficiently using a photocatalyst powder system.

On the other hand, effects of substitution of Ta-sites of $KTaO_3$ are also reported [59.24]. Not only dopants at A-sites, but also those at B-sites are strongly influenced by photocatalytic activity. Since Pyrex glass was used for the reactor, the catalytic activity is different from those in Table 59.4 because of negligible ultraviolet light, and it is reported that $KTaO_3$ is more active than that of $NaTaO_3$, which corresponds with the most narrow band gap among $ATaO_3$. The effects of substitution of various elements on mainly Ta-sites in $KTaO_3$ on photocatalytic activity to water splitting are shown in Table 59.5. Obviously, H_2 and O_2 formation rates are

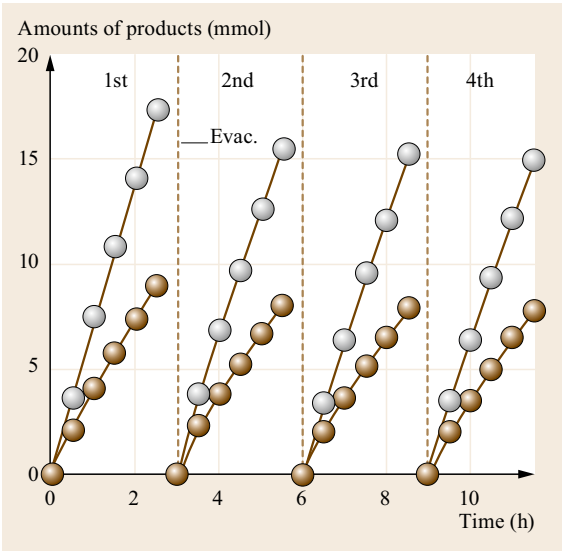


Fig. 59.8 Photocatalytic water splitting over NiO(0.05 wt%)/$NaTaO_3$:La(1 mol%). The cell was evacuated (evac.) after each experimental run. Catalyst: 1 g, pure water: 390 ml, 400 W high-pressure mercury lamp, inner irradiation cell made of quartz. Open marks: H_2, closed marks: O_2. (After [59.23])

Table 59.4 Band gap, surface areas and photocatalytic activities for water splitting into H_2 and O_2 in pure water on various lanthanidedoped $NaTaO_3$ with NiO cocatalysts. (After [59.21])

Ln-doped	Band gap (eV)	Surface area ($m^2\,g^{-1}$)	Activity (m mol h^{-1})[a]	
			H_2	O_2
None	4.01	0.44	2.18[b]	1.10[b]
La	4.07	2.5	5.90	2.90
Pr	4.09	3.1	5.29	2.58
Nd	4.07	3.0	5.19	2.51
Sm	4.08	2.6	5.29	2.63
Eu	4.08	2.5	0.254	0.122
Gd	4.08	1.9	4.29	2.11
Tb	4.07	1.4	4.30	2.19
Dy	4.07	1.7	4.46	2.23
Yb	4.05	1.3	1.72	0.820

[a] Catalyst: 1 g, pure water: 390 ml, inner irradiation cell made of quartz, 400 W high-pressure mercury lamp
[b] Initial activity

Table 59.5 Effect of substitution of various elements on mainly Ta-sites in $KTaO_3$ on photocatalytic activity to water splitting. (After [59.24])

Catalyst[a]	Formation rate (μ mol/h)	
	H_2	O_2
$LiTaO_3$	0.0	0.0
$NaTaO_3$	0.8	0.0
$Rb_4Ta_6O_{17}$	0.0	0.0
$KTaO_3$	4.5	0.0
$KT_{0.9}M_{0.1}O_3$		
$M = Zn^{2+}$	29.1	0.0
Y^{3+}	4.4	0.0
Al^{3+}	9.7	2.7
Ga^{3+}	67.7	22.3
In^{3+}	21.3	8.1
Ce^{4+}	trace	0.0
Ti^{4+}	50.6	12.6
Zr^{4+}	93.5	42.1
Hf^{4+}	98.5	39.8
Si^{4+}	17.2	4.1
Ge^{4+}	8.3	0.0
Nb^{5+}	3.7	0.0
Sb^{5+}	trace	0.0
W^{6+}	1.0	0.0
Pt/TiO_2 (0.3 wt%)	106.1	0.0

[a] 1 wt% NiO loaded

significantly increased by doping Ta-sites, in particular, the highest H_2 formation rate can be achieved by substituting Ta partially with Zr or HF. Since positive effects are tended to be obtained by doping tetravalent cations, it seems that a decrease in carrier density is effective for increasing photocatalytic activity for water splitting.

In addition to Ta-based perovskite, another important oxide based on the perovskite structure is $SrTiO_3$, which has narrower band gap than that of Ta-based perovskite oxide [59.25]. Since the valence band is not low enough for the formation of O_2, complete decomposition of water can be achieved by combining with oxygen-formation catalysts such as $BiVO_4$ with $Pt/SrTiO_3$ doped with Rh [59.22]. Among the various catalysts reported, it is obvious that perovskite oxide is an important family for photocatalysts from a unique semiconducting property viewpoint, and it is expected that research is expanded to the oxygen-deficient-type perovskite oxides such as K_2NiF_4-type structures.

59.3 Application for Solid Oxide Fuel Cells (SOFCs)

An important application area of inorganic perovskite oxide is solid oxide fuel cells and air electrodes of metal-air batteries because of high catalytic activity to oxygen reduction and superior mixed conductivity achieved simultaneously. In this section, the application of perovskite oxide for SOFCs is briefly mentioned. Further details are available in the another book [59.26].

Table 59.6 summarizes the important applications of perovskite oxides for solid oxide fuel cell technology. As shown in Table 59.6, $LaCoO_3$ or $LaMnO_3$ are promising candidates for SOFC cathodes and $LaGaO_3$-based oxides for the electrolyte. In addition, recently there have been several reports on the application of Cr-based perovskites as anodes. Therefore, the concept of SOFCs based entirely on perovskite components, *all-perovskite SOFCs*, is also proposed and some preliminary results have been reported [59.27]. In contrast to the SOFCs using oxide-ion-conducting electrolytes, the development of SOFCs using high-temperature proton-conducting electrolytes is slightly delayed, particularly when compared with development of polymer electrolyte-type fuel cells. However, the Toyota group has quite successfully demonstrated a high-power SOFC using a $BaCeO_3$-based electrolyte film on Pd foil [59.28]. Their data suggest that the proton-conducting perovskite oxides might also be an important component in real SOFCs in the near future.

59.3.1 Cathode

The principal requirement for a cathode of a SOFC is to electrochemically reduce oxygen molecules into oxide ions, and there are several requirements for oxide applied for cathodes of SOFCs i. e., catalytic activity, thermodynamic stability and compatibility in mechanical and chemical properties. Perovskite oxide is the most suitable material satisfying these requirements for cathodes. Figure 59.9 shows the reaction route considered for cathodes of SOFCs. Oxygen reduction proceeds on the electrode surface or at the electrode/electrolyte/gas-phase interface; the so-called triple phase boundary (TPB). The electrode material catalyzes the oxygen molecules to be dissociated into atoms, charged and incorporated into the electrolyte (Fig. 59.9). For the cathode material, the electrocatalytic activity is an important parameter to be considered. The surface reaction rate constant in oxygen isotope exchange is a good measure for the catalytic activity. *Kilner* et al. [59.29] compared various oxides in isotope diffusion coefficient and found a positive correlation between those parameters. A highly mixed electronic and ionic conductor may be a promising candidate in terms of the electrode performance. At the early stage of SOFC development, $La_{0.8}Ca_{0.2}MnO_3$ or $La_{0.6}Sr_{0.4}MnO_3$ were widely used, however, because of low oxide ion conductivity in Mn-based perovskite oxide, the reaction is limited to a three-phase boundary resulting in large cathodic overpotential. Recently, several oxides have been reported to show extremely high surface exchange rate for oxygen activation. *Baumann* et al. [59.30] compared several Co- and Fe-based perovskites in a controlled shape and found $Ba_{0.5}Sr_{0.5}Co_{0.8}Fe_{0.2}O_3$ shows 100 times smaller electrochemical resistance than that of $La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_3$ (LSCF) which is

Table 59.6 Important materials for perovskite oxide for solid oxide fuel cell applications

Component	Typical Materials
Cathode	$La(Sr)MnO_3$, $La(Sr)CoO_3$, $Sm_{0.5}Sr_{0.5}CoO_3$, $La(Sr)Fe(Co)O_3$
Electrolyte	$La(Sr)Ga(Mg)O_3(O^{2-})$, $BaCeO_3(H^+)$, $BaZrO_3(H^+)$, $SrZrO_3(H^+)$, $Ba_2In_2O_5(O^{2-})$
Anode	$La_{1-x}Sr_xCr_{1-y}M_yO_3$ (M = Mn, Fe, Co, Ni), $SrTiO_3$
Interconnector	$La(Ca)CrO_3$

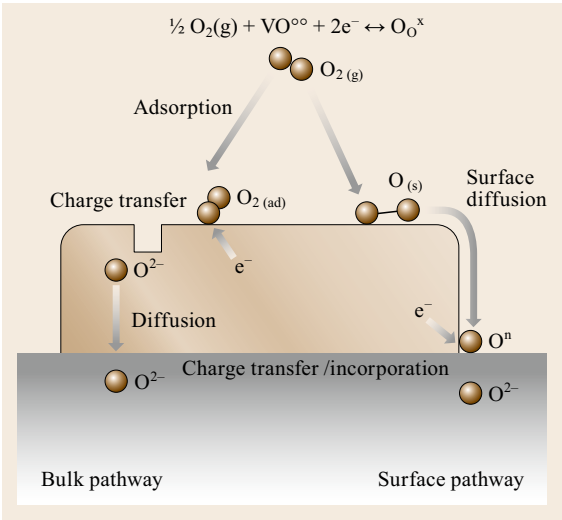


Fig. 59.9 Reaction route considered for a cathode of SOFC

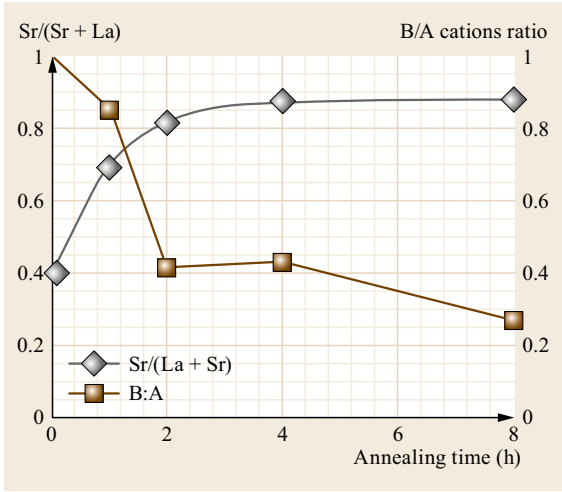


Fig. 59.10 Change in surface composition of $La_{0.6}Sr_{0.4}Fe_{0.8}Co_{0.2}O_3$ (LSCF) analyzed with low-energy ion scattering spectroscopy

often used for the intermediate-temperature SOFC cathode. One reason for such high cathodic activity is assigned to the high mixed conductivity and large reaction area available, i. e., the two phase boundary (route 2 in Fig. 59.9) is also contributed to the oxygen dissociation reaction. Another active composition to the cathode is $Sm_{0.5}Sr_{0.5}CoO_3$, which also has high oxide ion conductivity [59.31]. However, the most popular composition used for SOFC is $La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_3$ (LSCF) due to surface activity and stability. Recently, research has shifted to the more oxygen-deficient perovskites such as double (A_2BaO_6), Rudrusden–Poppered ($A_xB_yO_z$), or K_2NiF_4 structure. *Sase* et al. [59.32] reported that existence of a $(La, Sr)_2CoO_4$ phase on the $(La, Sr)CoO_3$ electrode enhances the oxygen exchange reaction rate. The double perovskite phase of $PrBaCo_2O_6$ is also reported as an active phase for oxygen dissociation [59.33] and so not only perovskite but also perovskite-related phases, in particular the oxygen-deficient perovskite phase, is now attracting as much attention as the active cathode catalyst for intermediate temperature operations.

On the other hand, one of another important issues for SOFC development is long-term stability and decrease in cathode performance is pointed out. There are several reasons decreasing cathodic performance is reported, i. e., chemical poisoning with Cr, S and B and sintering, and phase separation. Among them, surface segregation with Sr on the perovskite cathode has been suggested recently [59.34]. Figure 59.10 shows the change in surface composition of LSCF by low-energy ion scattering techniques, which are sensitive to the elements in the outermost surface layer. Obviously, the surface of LSCF is immediately enriched with Sr and it is considered that an Sr-enriched surface is highly reactive with Cr or S, resulting in the decrease in surface activity. Apparently, the surface composition of perovskite is slightly different from that of bulk, however, reactivity for oxygen dissociation is more strongly affected by B-site cations. Therefore, there is still some unclear points on parameters determining the activity for oxygen dissociation and this will be discussed more intensively in the future.

59.3.2 Anode

For the anode of SOFC, metal-oxide ion-conducting oxide composites named *cermet* have been widely used, in particular, Ni-Y_2O_3-stabilized ZrO_2 ($Y_{0.16}Zr_{0.84}O_2$, YSZ) cermet has been widely used. However, Ni is well known to be deactivated easily by aggregation and coarsening. In addition, re-oxidation of Ni is also occurs easily, resulting in the permanent failure of the cell. Therefore, recently, the application of oxide for the anode has also been proposed and among the proposed oxide anodes, perovskite oxides such as $SrTiO_3$ doped with La, Nd, etc. [59.35] or $La_{0.75}Sr_{0.25}Mn_{0.5}Cr_{0.5}O_3$ (LSCrM) [59.36] show interesting performance and are promising as oxide anodes.

In particular, improved performance has been obtained with complex perovskites based upon Cr and Mn at the B-sites forming compositions $(La,Sr)Cr_{1-x}M_xO_{3-\delta}$. Tao and Irvine have focused upon doped lanthanum chromite doped with Sr and Mn up to 20% dopant on the B-site, usually 5 or

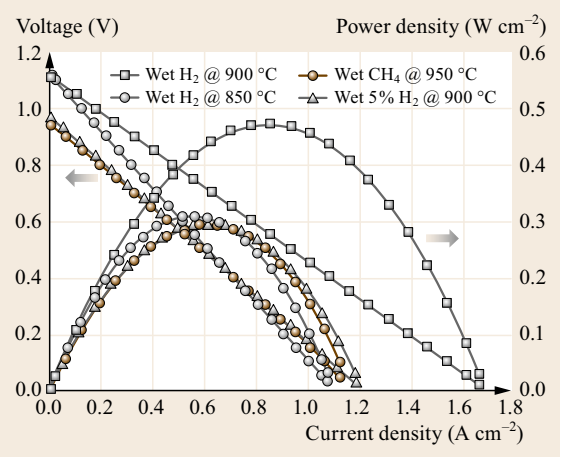


Fig. 59.11 Power generation property of the cell using $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_3$ for the anode

10%. $(La_{0.75}Sr_{0.25})Cr_{0.5}Mn_{0.5}O_3$ (LSCrM) exhibits comparable electrochemical performance to Ni/YSZ cermets [59.27, 36]. Figure 59.11 shows the I–V, I–P curves of the cell using LSCrM for the anode at 1173 and 1123 K. Apparently, the open circuit potential (OCV) is achieved to a theoretical level and reasonable power density of 0.4 W/cm^2 was achieved at 1173 K. The electrode polarization resistance approaches 0.2 Ω cm^2 at 1173 K in 97%H_2/3%H_2O. Good performance is achieved for methane oxidation without using excess steam. The anode is stable in both fuel and air conditions and shows stable electrode performance in methane. Thus both redox stability and operation in low steam hydrocarbons have been demonstrated. Catalytic studies of LSCM demonstrate that it is primarily a direct-oxidation catalyst for methane oxidation as opposed to a reforming catalyst, with the redox chemistry involving the Mn–O–Mn bonds. Although oxygen ion mobility is low in the oxidized state, the diffusion coefficient for oxide ions in reduced LSCrM is comparable to YSZ.

Another important double perovskite is $Sr_2MgMoO_{6-\delta}$, which has recently been shown to offer good performance, with power densities of 0.84 W/cm^2 in H_2 and 0.44 W/cm^2 in CH_4 at 1073 K, and good sulfur tolerance [59.37]. The molybdenum-containing double perovskite was initially prepared at 1473 K in flowing 5% H_2 and then deposited on top of a lanthanum ceria buffer layer before testing. Although there is some decomposition recognized, it is also reported that $La_{0.5}Sr_{0.5}MnO_3$ is active to not only the cathode but also the anode. The maximum power density of 1 W/cm^2 was reported at 1273 K when $La_{0.9}Sr_{0.1}Ga_{0.8}Mg_{0.2}O_3$ (LSGM, 0.5 mm thickness) is used for the electrolyte [59.38].

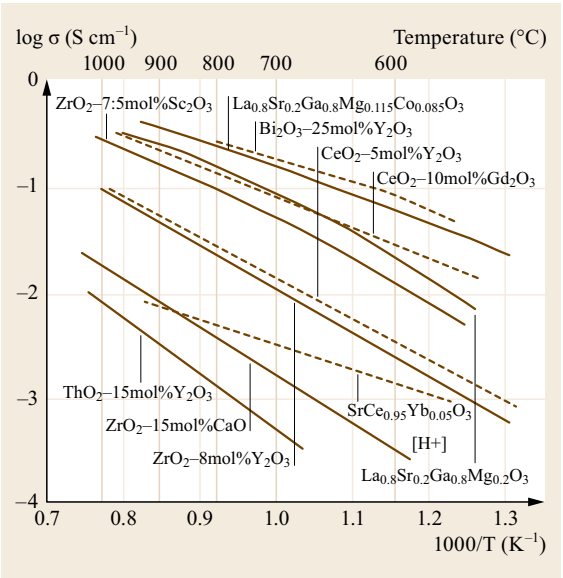


Fig. 59.12 Comparison of the oxide ion conductivity of LSGM with typical fluorite oxide

59.3.3 Electrolyte

An oxide ion conductor is widely used for electrolytes of SOFCs. In the case of SOFCs, solid proton conducting ceramics such as $BaZrO_3$ or $SrCeO_3$ doped with various rare earth cations to B-sites have been also studied [59.39], in particular for low-temperature SOFCs; however, considering the chemical stability and power density achieved, SOFCs using oxide ion conducting electrolytes are more promising. At present, almost all SOFCs under development for commercialization use YSZ for the electrolyte, with only a few exceptions. However, for increasing power density, there is still strong demand for replacing YSZ with an alternative electrolyte showing higher oxide ion conductivity. CeO_2 with fluorite structure and $LaGaO_3$ with perovskite structure are showing promise as electrolytes for this purpose. In particular, $LaGaO_3$ doped with Sr and Mg is the first pure oxide ion conductor with a perovskite structure [59.40]. Figure 59.12 shows a comparison of the oxide ion conductivity of LSGM with typical fluorite oxide. Apparently, oxide ion conductivity in LSGM is much higher than that of YSZ and almost comparable with Gd-doped CeO_2. In the case of CeO_2, partial electron conductivity appears in reducing atmospheres, however, LSGM exhibits the pure oxide ion conductivity across a wide p_{O_2} range, i. e., from pure oxygen to pure hydrogen.

The application of LSGM for electrolytes of SOFCs is also studied and the power generation property of the cell using LSGM is shown in Fig. 59.13. Since

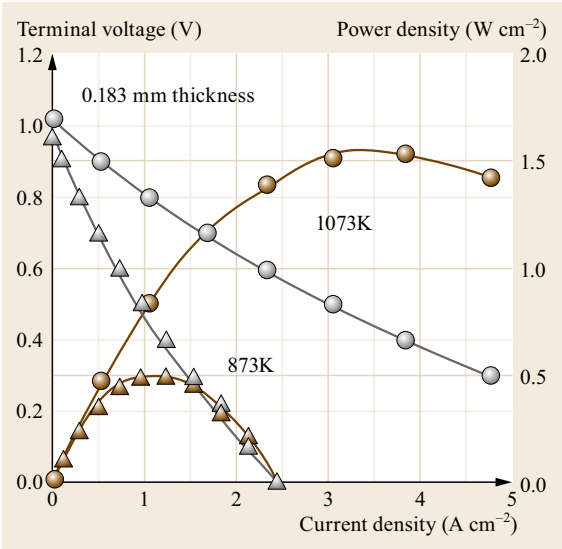


Fig. 59.13 Power generation property of the cell using LSGM with 0.182 mm thickness

transport number of the oxide ion in LSGM is close to unity from pure O_2 to pure H_2, the open circuit potential is 1.15 V, which is almost the same as that of theoretical open circuit potential. In addition, in spite of the thick electrolyte (0.5 mm), high maximum power density is achieved compared with that of the cell using YSZ. Therefore, it is expected that the operating temperature of SOFCs could be decreased by using LSGM for the electrolytes. Making thin films of LSGM is another important subject, however, it is reported that LSGM has some reactivity with Ni in the anode substrate and so the power density is not as high as expected from the film thickness and ionic conductivity. Therefore, for LSGM film cells, preventing reaction with Ni in the anode substrate is an important issue. *Huang* et al. reported that La-doped CeO_2 is effective for preventing Ni diffusion [59.41], however, electrical conductivity of La-doped CeO_2 is insufficient. Recently, *Hong* et al. reported that Ti- and La-codoped CeO_2 is promising as a buffer layer for preventing Ni diffusion with reasonable electrical conductivity and sintering properties [59.42]. In any case, LSGM perovskite oxide has a high potential as a fast oxide ion conductor for intermediate-temperature SOFCs and as an alternative to YSZ. Another interesting oxide ion conductor is $Ba_2In_2O_5$, which is a Brownmillerite structure and although oxide ion conductivity is slightly lower than that of YSZ, this new family of oxide ion conductors is also considered as a promising electrolyte for SOFCs [59.43].

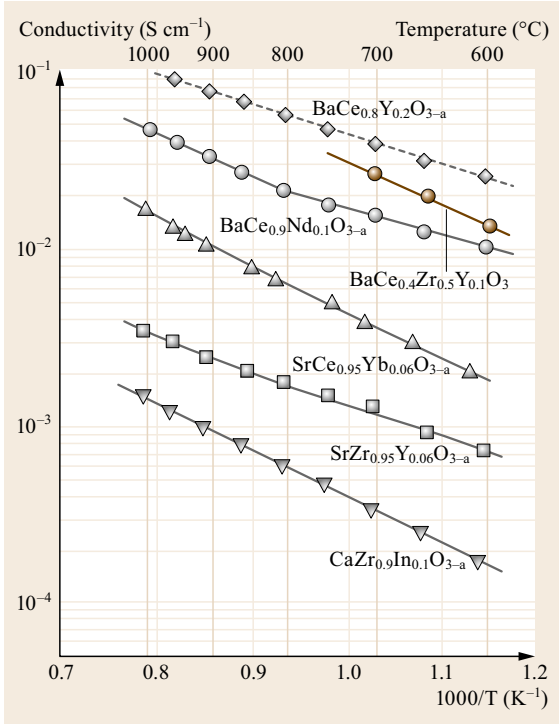


Fig. 59.14 Comparison of proton conductivity in perovskite oxide. (After [59.44].) Plots for $BaCe_{0.4}Zr_{0.5}Y_{0.1}O_3$ were added from [59.45]

Another important family for electrolytes of SOFC is oxide proton conductor and perovskite oxide of $ACeO_3$ or $BZrO_3$(A = Ba and Sr), which shows fast proton conductivity. Protons in perovskite oxide were formed by the following equation and the interstitial protons are mobile at the interstitial position of the perovskite lattice. Figure 59.14 shows the comparison of proton conductivity in perovskite oxide [59.44]. $BaCeO_3$ doped with Y shows reasonably high proton conductivity at low temperature. It is reported that extremely large power density is achieved by using $BaCeO_3$ doped with a Y thin film for the electrolyte of SOFC; however, these proton-conducting perovskites show high reactivity with CO_2 resulting in decomposition by formation of carbonate. Therefore, an increase in chemical stability is required for these oxide perovskite. Recently, there has been high interest on $ACe_{1-x}Zr_xO_3$-based oxides doped with Y from chemical stability; in particular it is reported that $BaCe_{0.4}Zr_{0.5}Y_{0.1}O_3$ shows reasonably high proton conductivity and reasonable chemical stability to CO_2 and hydrolysis with water [59.45].

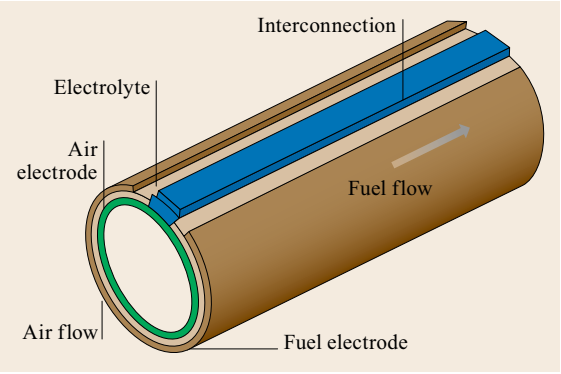


Fig. 59.15 Schematic view of the SOFC stack

59.3.4 Interconnector

The interconnector is an important component for stacking SOFCs and requires high stability in reducing and oxidizing atmospheres. The interconnector is used for connecting the SOFC single cell as shown in Fig. 59.15 and so high electrical conductivity and no oxide ion conductivity are also required for the high performance of the SOFC stack. Perovskite oxides of $LaCrO_3$ have been also widely used for interconnectors. Although $LaCrO_3$ is stable under reducing and oxidizing atmospheres, the electrical conductivity is still insufficient and also sintering is rather difficult. Therefore, doping Ca^{2+} for La-site in $LaCrO_3$ is generally performed for increasing conductivity as well as sintering properties.

Figure 59.16 shows electronic conductivity of doped $LaCrO_3$ as a function of temperature [59.46, 47]. The electronic conductivity increases with increasing temperature, suggesting the semiconductor temperature dependence. An increasing of the Ca concentration in $La_{1-x}Ca_xCrO_{3-\delta}$ enhanced the electronic conductivity due to the increase of Cr^{4+} concentration. There are some deviations of the electrical conductivity among the examined alkaline earth elements: Ca-doped $LaCrO_3$ shows larger electrical conductivity than Sr-doped $LaCrO_3$. This difference was reported to be due to the difference of lattice distortion and phase stability. The activation energy for conductivity was 0.12–0.14 eV and the mobility was $0.066–0.075\,cm^2/(Vs)$ at 1173–1323 K.

The electronic conductivity decreases with a reduction of oxygen partial pressure because of the decrease of Cr^{4+} concentration in a reducing atmosphere. The

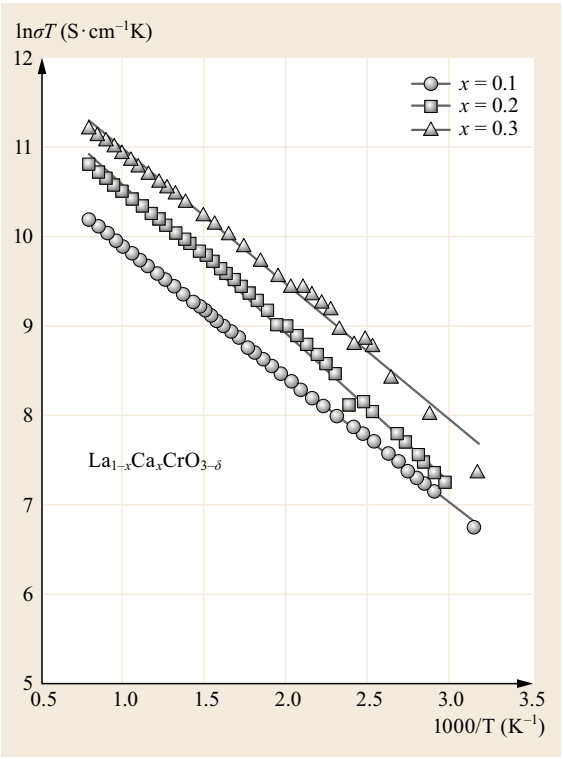


Fig. 59.16 Electrical conductivity of $La_{1-x}Ca_xCrO_{3-\delta}$ ($x = 0.1–0.3$) in air as a function of inverse temperature. (After [59.47])

electrical conductivity decreases with a reduction of oxygen partial pressures. The electrical conductivity is proportional to $p_{O_2}^{1/4}$, which is consistent with the defect chemistry of $La_{1-x}Ca_xCrO_{3-\delta}$ [59.47]. A doping of the B-site has been also considered by several authors [59.48]. A typical dopant cation is Mg^{2+}, replaced into Cr^{3+} sites. This substitution also increases the concentration of Cr^{4+}, and eventually increases the electrical conduction. Because of low sintering property, $SrTiO_3$ doped with La for Sr sites or Nd for Ti sites is also studied as an interconnector. Perovskite oxide is also important for interconnectors of SOFCs.

In summary, perovskite oxide is widely used for current SOFCs and becomes important compounds for SOFC components. Therefore, the all-perovskite concept is also proposed for SOFC and the device consisting of materials with the same structure is highly interesting.

59.4 Oxygen Separating Membrane

Perovskite oxide shows high electronic and oxide ionic conductivity and these conductors are called mixed conductors. Since charge compensation by ion transport can be automatically achieved with electronic conductivity in mixed conductors, so ions can be transported in mixed conductors without outside circuits. Therefore, an important application of such mixed conductors is as a separation membrane for oxygen from air. The oxygen permeation rate is shown as the following equation if the bulk diffusion is a rate-limiting step

$$J_{O_2} = \frac{RT\sigma_{el}\sigma_{ion}}{16F^2(\sigma_{el}+\sigma_{ion})t}\ln\frac{p_h}{p_l} \qquad (59.1)$$

Here, J_{O_2}; oxygen flux, σ_{el}; electron conductivity, σ_{ion}; oxide ion conductivity, F; Faraday constant, R; gas constant, t; membrane thickness, T; temperature and p_h, p_l means high and low oxygen partial pressure respectively.

From (59.1), the oxygen permeation rate is limited by bulk diffusivity of oxygen. Figure 59.17 shows the diffusivity of oxide ions in several perovskite oxides considered for oxygen permeation membranes [59.49]. Apparently, Fe- or Co-based perovskite oxides show fast oxide ion conductivity and so large oxygen permeation rate is expected on these perovskite oxides. In fact, a large oxygen permeation rate of 1.05 ml/(cm^2 min) at 1173 K is reported for $Ba_{0.5}Sr_{0.5}Fe_{0.8}Co_{0.2}O_3$ (BSCF) (2 mm thickness) [59.50]. According to (59.1), the oxygen permeation rate is determined by the diffusivity of oxide ions and the thickness of the membrane when the oxygen partial pressure differential across the membrane is the same. Therefore, the oxygen permeation rate should be increased with decreasing membrane thickness; however, it is reported that the oxygen permeation rate is dependent on membrane thickness when thickness is large, but with decreasing thickness, the oxygen permeation rate becomes independent of oxygen permeation rate because of the limitation by surface reactions of oxygen dissociation. To achieve the large oxygen permeation rate, not only high diffusivity of oxide ion in bulk but also high surface activity to oxygen dissociation is required.

Recently, perovskite-related oxides are also attracting much interest as mixed conductors, and applied use for oxygen permeation membranes. Among the perovskite-related oxides, there is strong interest for the K_2NiF_4-type oxide because of the large amount of oxygen deficiency. As explained in Figure 59.5, K_2NiF_4 oxides consist of a series of connected perovskite and rock salt blocks, and a large free volume exists in the rock salt block. Therefore, interstitial oxygen can be easily introduced into the rock salt block. For the K_2NiF_4 oxides, the oxide ion permeation property in La_2NiO_4-based oxide is now studied in detail and it is reported that doping Cu and Ga is effective for increasing oxygen permeation rate and $Pr_2Ni_{0.71}Cu_{0.24}Ga_{0.05}O_3$ (denoted as PNCG) is the optimized composition for oxygen permeation and permeation rate of approximately 3 cc/(min cm^2) was reported on PNCG with 0.5 mm thickness from air to He [59.51]. The oxygen transport route is also studied by neutron diffraction analysis and as shown in Fig. 59.18, oxide ion transports through interstitial positions in the rock salt block is clearly demonstrated. Therefore, in this PNCG, oxide ions are mainly transported through rock salt but not in the perovskite block. On the other hand, PNCG shows high hole conduction and holes are mainly transported in the perovskite block; and so in PNCG, it is interesting that two different routes for oxide ion and hole conduction are expected [59.52].

The most interesting use of mixed conductors is to combine the catalytic reaction and the so-called membrane reactor system; and several catalytic reactions such as NO_x decomposition and partial oxidation of alkanes are reported by using mixed conductors for removing the reactant from or into reaction systems. To permeate oxygen through mixed conductor membranes, a gradient in oxygen partial pressure is required as a driving force for oxygen transport; however, in membrane reactor system, differences in oxygen partial pres-

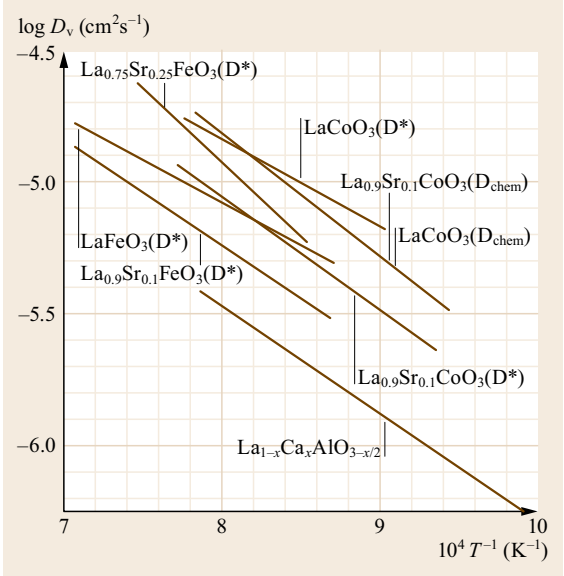


Fig. 59.17 Diffusivity of oxide ion in several perovskite oxides considered for an oxygen permeation membrane. (After [59.49])

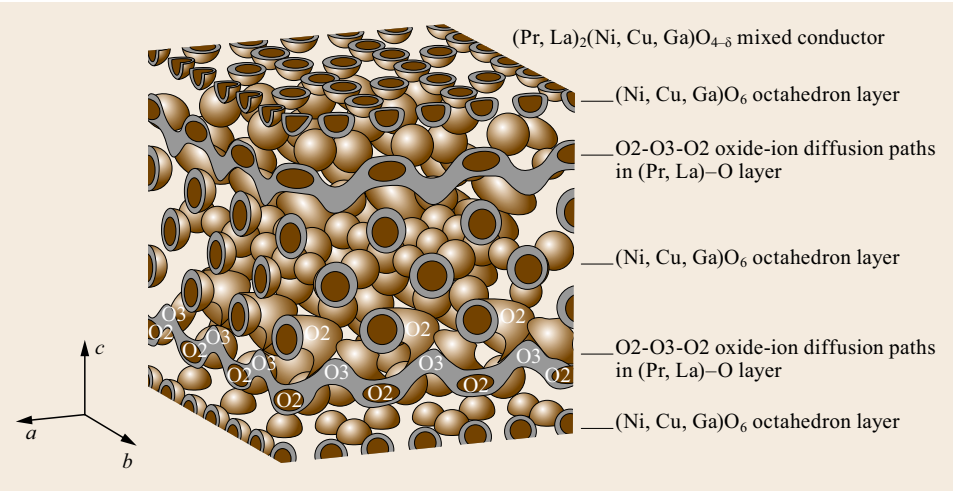


Fig. 59.18 Oxygen transport route in $Pr_2Ni_{0.81}Cu_{0.24}Ga_{0.05}O_4$ estimated by neutron diffraction analysis. (After [59.52])

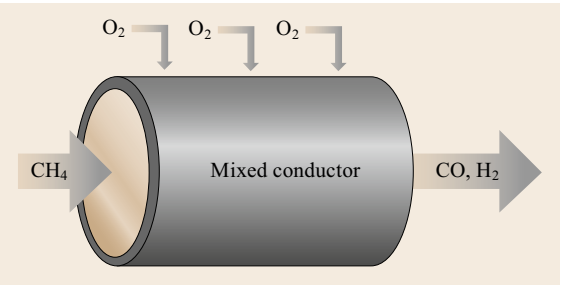


Fig. 59.19 Schematic image of the catalytic membrane reactor using mixed conductors

sure are automatically achieved. Therefore, the combination of mixed conductors with catalytic reactions is the most ideal usage. From this aspect, partial oxidation of CH_4 has been studied with perovskite mixed conductors for oxygen separation from air [59.53]. The schematic image of this catalytic reactor system was shown in Fig. 59.19. Since oxygen partial pressure in CH_4 is as low as approximately 10^{-20} atm, a large oxygen permeation rate is achieved under CH_4 partial oxidation conditions and it is reported that an oxygen permeation rate of $5\,\mathrm{cc/(min\,cm^2)}$ was achieved by using $SrFe_{0.8}Co_{0.2}O_3$ perovskite [59.54]. However, the oxygen permeation rate is decreased by phase changes in a reducing atmosphere of CH_4. Therefore, although Fe- or Co-based perovskites have been mainly studied, the most important issues for application of perovskite oxides to catalytic membrane reactors are stability in a wide oxygen partial pressure range and it is reported that Ni- or Fe-doped $LaGaO_3$ shows hole and oxide ion conductivity stably over wide p_{O_2} ranges and the oxygen permeation rate of $12\,\mathrm{cc/(min\,cm^2)}$ was exhibited on $La_{0.7}Sr_{0.3}Ga_{0.6}Fe_{0.4}O_3$ of 0.2 mm thick at 1273 K in CH_4 partial oxidation [59.55]. As a result, obviously, perovskites are important materials for oxygen separation membranes and the application of catalytic membrane reactors is an important area; albeit the chemical stability is required to be much improved.

59.5 Summary

In this chapter, crystal structures of perovskites and related oxides were explained. Perovskite oxide has a variety of composition and component elements. In addition, there are many isomorphs in crystal structure. Therefore, based on variety of crystal structures, there are many functions and rich application areas. In particular, in this chapter, the application of perovskite oxides for photocatalytic properties and solid oxide fuel cells were briefly overviewed. Since high electric conductivity and surface activity to oxygen dissociation are achieved simultaneously, perovskite oxides, mainly Co-, Fe- and Mn-based oxides are widely used for SOFCs and oxygen permeation membranes. On the other hand, for the application to photocatalysts, Ta- or Ti-based perovskites are highly active. Therefore, perovskite oxide is a highly important compound for these areas.

References

59.1 R.M. Hazen: Sci. Amer. **258**(6), 74 (1988)
59.2 T. Yagi, H.K. Mao, P.M. Bell: Phys. Chem. Minerals **3**, 97 (1978)
59.3 F. Kanamura: Perovskite related compound. In: *Kikan Kigaku Sasetsu, No.32*, ed. by Japanese Society of Chemistry (Japanese Society of Chemistry, Tokyo 1997) p. 9
59.4 S. Geller, J.B. Jeffries, P.J. Curlander: Acta Cryst. B **31**, 2770 (1975)
59.5 R.C. Liebermann, L.E.A. Jones, A.E. Ringwood: Phys. Earth Planet. Inter. **14**, 165 (1977)
59.6 A.F. Well: *Structural Inorganic Chemistry*, 5th edn. (Oxford Univ. Press, Oxford 1984)
59.7 A.F. Cotton, G. Wilkinson: *Advanced Inorganic Chemistry* (Wiley, New York 1988)
59.8 F.S. Galasso: *Perovskites and High Tc Superconductors* (Gordon Breach, New York 1990)
59.9 R.H. Mitchell, T. Bay: *Perovskites Modern and Ancient* (Almaz, Thunder Bay 2002)
59.10 H. Arai, T. Yamada, K. Eguchi, T. Seiyama: Appl. Catalysis **26**, 265 (1986)
59.11 J.B. Bednorz, K.A. Müller: Z. Phys. B **64**, 189 (1986)
59.12 P.H. Hor, R.L. Meng, Y.Q. Wang, L. Gao, Z.J. Huang, J. Bechtold, K. Forster, C.W. Chu: Phys. Rev. Lett. **58**, 1891 (1987)
59.13 H. Maeda, Y. Tanaka, M. Fukutomi, T. Asano: Jpn. J. Appl. Phys. **27**, L209 (1988)
59.14 L. Gao, Y.Y. Xue, F. Chen, Q. Xiong, R.L. Meng, D. Ramirez, C.W. Chu, J.H. Eggert, H.K. Mao: Phys. Rev. B **50**, 4260 (1994)
59.15 J. Mizusaki, M. Yoshihiro, S. Yamauchi, K. Fueki: J. Solid State Chem. **58**, 257 (1985)
59.16 J. Mizusaki, I. Yasuda, J. Shimoyama, S. Yamaguchi, K. Fueki: J. Electrochem. Soc. **140**, 467 (1993)
59.17 S. Shin, H. Arakawa, Y. Hatakeyama, K. Ogawa, K. Shimomura: Mater. Res. Bull. **14**, 633 (1979)
59.18 Y. Teraoka, T. Harada, S. Kagawa: J. Chem. Soc. Faraday Trans. **1998**, 94 (1887)
59.19 H. Yasuda, T. Nitadori, N. Mizuno, M. Misono: Bull. Chem. Soc. Jpn. **66**, 3492 (1993)
59.20 H. Iwakuni, Y. Shinmyou, H. Yano, H. Matsumoto, T. Ishihara: Appl. Catal. B **299**, 74 (2007)
59.21 H. Kato, A. Kudo: Catal. Today **78**, 561 (2003)
59.22 Y. Nishihata, J. Mizuki, T. Akao, H. Tanaka, M. Uenishi, M. Kimura, T. Okamoto, N. Hamada: Nature **418**, 164 (2002)
59.23 K. Maeda: J. Photochem. Photobiol. C Photochem. Rev. **12**, 237 (2011)
59.24 A. Kudo, H. Kato: Chem. Phys. Lett. **331**, 373 (2000)
59.25 T. Ishihara, H. Nishiguchi, K. Fukamachi, Y. Takita: J. Phys. Chem. B **103**, 1 (1999)
59.26 K. Domen, S. Naito, M. Soma, T. Onishi, K. Tamaru: J. Chem. Soc. Chem. Commun. **12**, 543 (1980)
59.27 T. Ishihara (Ed.): *Perovskite Oxide for Solid Oxide Fuel Cells* (Springer, New York 2009)
59.28 S. Tao, J.T.S. Irvine, J.A. Kilner: Adv. Mater. **17**, 1734 (2005)
59.29 N. Ito, M. Iijima, K. Kimura, S. Iguchi: J. Power Source **152**, 200 (2005)
59.30 J.A. Kilner, R.A. De Souza, I.C. Fullarton: Solid State Ion. **86–88**, 703 (1996)
59.31 F.S. Baumann, J. Fleig, G. Cristiani, B. Stuhlhofer, H.U. Habermeier, J. Maier: J. Electrochem. Soc. **154**, B931 (2007)
59.32 T. Ishihara, M. Honda, T. Shibayama, H. Nishiguchi, Y. Takita: J. Electrichem. Soc. **145**(9), 3177 (1998)
59.33 M. Sase, K. Yashiro, K. Sato, J. Mizusaki, T. Kawada, N. Sakai, K. Yamaji, T. Horita, H. Yokokawa: Solid State Ion. **178**, 1843 (2008)
59.34 M. Burriel, J. Pena-Martinez, R.J. Chater, S. Fearn, A.V. Berenov, S.J. Skinner, J.A. Kilner: Chem. Mat. **24**, 613 (2012)
59.35 J. Druce, H. Tellez, M. Burriel, M. Sharp, L. Fawcett, S. Cook, D. McPhail, T. Ishihara, H.H. Brongersma, J.A. Kilner: Energy Environ. Sci. **7**, 3593 (2014)
59.36 J.T.S. Irvine, P.R. Slater, P.A. Wright: Ionics **2**, 213 (1996)
59.37 S.W. Tao, J.T.S. Irvine: Nat. Mater. **2**, 320 (2003)
59.38 Y.H. Huang, R.I. Dass, Z.L. Xing, J.B. Goodenough: Science **312**, 254 (2006)
59.39 T. Ishihara, S. Fukui, M. Enoki, H. Matsumoto: J. Electrochem. Soc. **153**, A2085 (2006)
59.40 H. Iwahara, T. Esaka, H. Uchida, N. Maeda: Solid State Ion. **3–4**, 359 (1981)
59.41 T. Ishihara, H. Matsuda, Y. Takita: J. Am. Chem. Soc. **116**, 3801 (1994)
59.42 K. Huang, R. Tichy, J.B. Goodenough, C. Milliken: J. Am. Ceram. Soc. **81**, 2581 (1998)
59.43 J.E. Hong, T. Inagaki, S. Ida, T. Ishihara: J. Am. Ceram. Soc. **95**(11), 3588 (2012)
59.44 J.B. Goodenough, J.E. Ruiz-Diaz, Y.S. Zhen: Solid State Ion. **44**, 21 (2000)
59.45 H. Iwahara: Solid State Ion. **86–88**, 9 (1996)
59.46 I. Higuchi, T. Tsukamoto, N. Sata, S. Yamaguchi, S. Shin, I. Hattori: Solid State Ion. **176**(39/40), 2963 (2005)
59.47 W.J. Weber, C.W. Griffin, J.L. Bates: J. Am. Ceram. Soc. **70**, 265 (1987)
59.48 I. Yasuda, T. Hikita: J. Electrochem. Soc. **140**, 1699 (1993)
59.49 J.W. Fergus: Solid State Ion. **171**, 1 (2004)
59.50 J.A. Kilner: Solid State Ion. **129**, 13 (2000)
59.51 L.A. Tan, X.H. Gu, L. Yang, W.Q. Jin, L.X. Zhang, N.P. Xu: J. Membrane Sci. **212**, 157 (2003)
59.52 T. Ishihara, S. Miyoshi, T. Furuno, O. Sanguanruang, H. Matsumoto: Solid State Ion. **177**, 3087 (2006)
59.53 M. Yashima, H. Yamada, S. Nuansaeng, T. Ishihara: Chem. Mater. **24**, 4100 (2012)
59.54 B. Ma, U. Balachandran, J.H. Park, C.U. Segre: J. Electrochem. Soc. **143**, 1736 (1996)
59.55 T. Ishihara, Y. Tsuruta, T. Todaka, H. Nishiguchi, Y. Takita: Solid Sate Ion. **152/153**, 709 (2002)

About the Authors

Martin Abkowitz (deceased)

Chapter D.37

Sadao Adachi

Chapter D.30

Gunma University
Faculty of Science and Technology
Gunma, Japan
adachi@gunma-u.ac.jp

Sadao Adachi received his PhD from Osaka University and is a Professor Emeritus of the Faculty of Science and Technology at Gunma University. He has published and presented over 300 technical papers and 25 textbooks on semiconductor physics and technology. His current research interests include physical properties of semiconductors and new functional materials.

Guy J. Adriaenssens

Chapter A.7

University of Leuven
Lab. for Semiconductor Physics
Leuven, Belgium
guy.adri@fys.kuleuven.be

Guy Adriaenssens received his PhD from the University of Washington (1971) for an NMR study of ferroelectric phase transitions. After postdoctoral studies at the University of Saarland, Germany, he joined the University of Leuven, Belgium, in 1973. There his main research interests centered on the electronic transport properties and the electronic density of states of amorphous semiconductors, chalcogenide glasses and conjugated polymers, and on defect structures in CVD diamonds.

Wilfried von Ammon

Chapter A.5

Von Ammon Consulting
Roth, Germany
wilfried.ammon@aon.at

Wilfried von Ammon studied Physics at the Technical University of Munich and at the University of Regensburg; he received his PhD in Physics in 1981. He then joined Wacker Siltronic, working in research and development. His main focus was on silicon crystal growth, silicon-layer growth by chemical vapor deposition, neutron transmutation doping, thermal treatment of silicon, and behavior of defects in silicon. He is now retired.

Hajime Asahi

Chapter C.23

Osaka University
Inst. of Scientific and Industrial Research
Osaka, Japan
asahi@sanken.osaka-u.ac.jp

Hajime Asahi received his PhD degree from the University of Tokyo, Japan in 1976. He then joined NTT Laboratories, where he engaged in research on MBE of III-V semiconductors and their optical device applications. In 1987 he joined ISIR, Osaka University, then focussing on research on MBE of III-nitride, diluted magnetic and temperature-insensitive semiconductors, quantum nanostructures, and their device applications.

Peter Ashburn

Chapter C.22

University of Southampton
School of Electronics and Computer Science
Southampton, UK
pa@ecs.soton.ac.uk

Peter Ashburn received a PhD degree in Electrical and Electronic Engineering in 1974 from the University of Leeds and then joined the Philips Research Laboratories. In 1978 he joined the University of Southampton and is currently a Professor of Microelectronics. His research interests include SiGe heterojunction bipolar transistors (HBTs), ultimate CMOS and carbon nanotubes. He has published over 200 papers in the scientific literature and has authored two books on bipolar transistors.

Mark Auslender

Chapter C.21

Ben-Gurion University of the Negev Beer Sheva
Dept. of Electrical and Computer Engineering
Beer Sheba, Israel
marka@bgu.ac.il

Mark Auslender received a PhD in Solid-State Theory in 1977. He was a Certified Senior Researcher at the Institute of Metal Physics (Ural Branch, USSR Academy of Sciences). In 1991 he joined the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel, where he is a Grade A Researcher and Adjunct Full Professor in the department. His present interests focus on micro- and nanophotonics, optical sensors, graphene, and manganites.

Darren M. Bagnall

University of New South Wales
School of Photovoltaic and Renewable Energy Engineering
Sydney, Australia
d.bagnall@unsw.edu.au

Chapter C.22

Darren Bagnall is a senior lecturer at the School of Electronics and Computer Science at Southampton University. His research has included pioneering work on molecular-beam epitaxy (MBE) and chemical vapor deposition (CVD) for the development of new nano and quantum devices based on ZnO, silicon and silicon germanium. He has published over 50 refereed papers.

Ian M. Baker

Leonardo MW Ltd
Southampton, UK
ian.m.baker@leonardocompany.com

Chapter D.34

Ian Baker received his PhD in Silicon Sensors from Southampton University in 1973. His career in solid state imaging has spanned visible imaging CCDs to advanced infrared detectors. He has presented over 90 papers and has 32 patents. He was awarded an MBE for services to industry in 2008. His current interests are avalanche photodiode arrays for long range surveillance and astronomy.

Sergei Baranovskii

Philipps University
Dept. of Physics
Marburg, Germany
baranovs@staff.uni-marburg.de

Chapter A.9

Sergei Baranovskii received his PhD (1981) at Ioffe Institute, St. Petersburg, Russia, where he worked as a Senior Researcher till 1990. Since 1990 he has been working at Philipps University Marburg, Germany. His research interests are devoted to optical properties and charge transport in disordered semiconductors. He has edited two books, co-authored seven book chapters, and published 230 peer-reviewed articles.

Mark Baxendale

Queen Mary University of London
School of Physics and Astronomy
London, UK
m.baxendale@qmul.ac.uk

Chapter E.47

Mark Baxendale has been a Reader in Nanotechnology at Queen Mary University of London since 2002. His general research interest is in self-organization and emergent phenomena in the physical and life sciences with a focus on the science and applications nanoscale carbon macromolecules (fullerenes, carbon nanotubes, and graphene) in the physical and life sciences.

Mohammed L. Benkhedir

Tebessa University
Lab. of Applied and Theoretical Physics
Tebessa, Algeria
benkhedir@gmail.com

Chapter A.7

Mohammed L. Benkhedir received an MS in Physics from the University of Annaba, Algeria, and holds a lectureship at the Centre Universitaire de Tebessa, Algeria. He gained a PhD in Physics from the University of Leuven, where he studied the electronic properties and density of states in amorphous selenium by means of photoconductivity techniques. He is currently Professor at the University of Tebessa, Algeria.

Monica Brinza

ASML Netherlands B.V.
Veldhoven, The Netherlands
monica.brinza@asml.com

Chapter A.7

Monica Brinza received a PhD from the University of Leuven in 2004 with a thesis on electronic transport properties of hydrogenated amorphous silicon. She continued her research work at Utrecht University, with a post-doc focused on low temperature amorphous and microcrystalline solar cells. In 2010 she joined ASML as a design engineer in the Development and Engineering Department of EUV lithographic systems.

Paul D. Brown

University of Nottingham
Dept. of Mechanical, Materials and Manufacturing Engineering
Nottingham, UK
paul.brown@nottingham.ac.uk

Chapter B.17

Paul D. Brown received a PhD in Applied Physics from the University of Durham in 1989. He is presently Professor of Materials Characterisation at the University of Nottingham. His current research interests are concerned with the interrelationship between the processing, structure, and property of structural, functional, and biomedical materials, assessed using a broad range of characterization techniques, with emphasis on novel variants of electron microscopy.

Peter Capper Chapters B.12, B.14, B.15 For biographical profile, please see the section "About the Editors".

R. Lawrence Comstock (deceased) Chapter E.49

Ray DeCorby

University of Alberta
Electrical and Computer Engineering
Edmonton, Canada
rdecorby@ualberta.ca

Chapter D.41

Ray DeCorby is a Professor of Electrical and Computer Engineering at the University of Alberta, Edmonton, Canada. His research is in the area of integrated optics, with a current focus on air-core waveguides and microcavity devices for sensing, lab-on-a-chip, and quantum information applications.

M. Jamal Deen

McMaster University
Dept. of Electrical and Computer Engineering
Hamilton, Canada
jamal@mcmaster.ca

Chapter B.20

M. Jamal Deen is a Distinguished University Professor and Senior Canada Research Chair in Information Technology at McMaster University, Canada. He has published over 500 peer-reviewed articles, two textbooks, six patents, and has 14 best paper/poster/presentation awards. His current research interests are nanoelectronics, optoelectronics, nanotechnology, and their emerging applications to health and environmental sciences. Dr Deen is currently serving as President of the Academy of Science, The Royal Society of Canada.

Leonard Dissado

University of Leicester
Dept. of Engineering
Leicester, UK
lad4@leicester.ac.uk

Chapter A.10

Professor Emeritus Leonard Dissado works on the theory of dielectric response and electrical breakdown phenomena in a wide range of materials from glasses and high voltage insulators to bio-tissues. He was awarded a DSc from the University of London (1990), a Docteur (hc) by Paul Sabatier University, Toulouse, France (2007), and an Honorary Professorship from Xi'an Jiaotong University, China (2009).

David Dunmur

University of Southampton
Chemistry
Southampton, UK
d.dunmur@tiscali.co.uk

Chapter D.36

David Dunmur received his DPhil degree from the University of Oxford in 1965. His research has been concerned with the dielectric, optical, electro-optical and elastic properties of liquid crystals. He was Founding Editor of Liquid Crystals Today and the 1999 recipient of the G.W. Gray Medal of the British Liquid Crystal Society. Professor Dunmur has now retired from the University of Southampton.

Lester F. Eastman (deceased) Chapter D.32

Andy Edgar

Victoria University
School of Chemical and Physical Sciences
Wellington, New Zealand
andy.edgar@vuw.ac.nz

Chapter D.38

Andy Edgar is an Associate Professor and Senior Research Fellow at the School of Chemical and Physical Sciences of Victoria University, Wellington, New Zealand. His research interests are in materials science, particularly glasses, ceramics, and glass ceramics for opto-electronic applications such as radiation imaging and detection. He is also interested in the development of optical camera-based systems for X-ray imaging.

Vassili Fedotov

University of Southampton
Optoelectronics Research Centre
Southampton, UK
vaf@orc.soton.ac.uk

Chapter E.56

Vassili Fedotov received a PhD degree in Laser Physics in 2003 from the University of Southampton and is presently a Principal Research Fellow at the university's Optoelectronics Research Centre. His research interests lie in the fields of metamaterials, plasmonics, and toroidal electrodynamics, and he has authored and co-authored over 70 peer-reviewed journal publications on these topics.

Romualdo A. Ferreyra

Kyoto Institute of Technology
Innovation Center
Kyoto, Japan
raferrey@kit.ac.jp

Chapter D.31

Romualdo Alejandro Ferreyra received his PhD from Virginia Commonwealth University, Richmond, USA, in 2014. Since 2015, he has held a post-doctoral researcher position at Kyoto Institute of Technology, Japan. His main areas of research interest are III-nitride materials' growth, deposition of insulators and their related solid state physics, design, fabrication, and characterization of electronic devices, and their mathematical and computational calculations.

Paul Fons

National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan
paul-fons@aist.go.jp

Chapter E.46

Paul Fons joined the National Institute of Advanced Industrial Science and Technology in 1993. At AIST he has worked on both molecular beam epitaxy growth and characterization of materials using synchrotron radiation techniques such as diffraction, EXAFS, and HAXPES. He also employs density-functional theory calculations. He is currently a Chief Senior Researcher in the Systematic Materials Design Group.

Brian E. Foutz

Charlottesville, USA
foutz@cadence.com

Chapter D.32

Brian E Foutz completed his PhD at Cornell University studying electron transport and semiconductor device physics. He is a member of the IEEE and recently served on the Standards Committee for IEEE 1687. Currently, he works for Cadence Design Systems developing new techniques and logic for automated testing of semiconductor systems on chips.

Mark Fox

University of Sheffield
Dept. of Physics and Astronomy
Sheffield, UK
mark.fox@sheffield.ac.uk

Chapter D.40

Mark Fox is Professor of Optical Physics at the University of Sheffield. He obtained his DPhil degree from the University of Oxford in 1987 and worked at AT&T Bell Laboratories and at Oxford before joining the University of Sheffield in 1998. His research interests include low-dimensional semiconductor structures, quantum optics, and ultrafast laser spectroscopy. He has authored more than 150 papers and two texts in the Oxford Masters Series in Physics.

Darrel Frear

NXP Semiconductor
Core Technology in Package Innovation
Chandler, USA
darrel.frear@nxp.com

Chapter E.53

Darrel Frear earned MS and PhD degrees in Materials Science from the University of California, Berkeley. He is currently the Director of Core Technology in Package Innovation at NXP. Previously, Darrel was with Sandia National Labs, Motorola and Freescale Semiconductor. Darrel works in advanced packaging research and development including materials, modeling, manufacturing, and reliability.

Jacek Gieraltowski

European Institute for Marine Studies
Oceanic Domains Laboratory
Plouzané, France
jacek.gieraltowski@univ-brest.fr

Chapter A.4

Jacek Gieraltowski holds a DSc from Warsaw University (Poland). He has been involved in research on magnetic materials (ferrites, thin films and nanowires) for more than 35 years. Presently, he is Professor Emeritus of Physics at Institut Universitaire Européen de la Mer (IUEM), working on fast switching magnetization problems for high density magnetic recording, giant magnetoimpedance sensors, and the application of magnetism to biological and medical applications.

Robert D. Gould (deceased)

Chapter C.28

Yaser M. Haddara

McMaster University
Dept. of Electrical and Computer Engineering
Hamilton, Canada
yaser@mcmaster.ca

Chapter C.22

Yaser Haddara is Associate Professor of Electrical Engineering at McMaster University. He received his PhD in Electrical Engineering from Stanford University in 1997. His research interest is in semiconductor fabrication process modeling. He has published on modeling diffusion in GaAs, Si, and SiGe, as well as modeling of defect formation, nanowire growth, and SiGe oxidation. His current research is on cobalt germanidation, polymer sensors, and SiC process technology.

Walid A. Hadi

University of Windsor
Dept. of Electrical and Computer Engineering
Windsor, Canada
abdulhw@uwindsor.ca

Chapter D.32

Walid Hadi received his PhD in Electrical Engineering from the University of Windsor. He is currently an Assistant Professor at Saint Clair College and an Adjunct Professor at Lawrence Technological University. Dr Hadi is a post-doctoral fellow at the University of British Columbia. His research is on III-V and II-VI semiconductors and their application in high electron mobility transistors.

Shlomo Hava

Ben-Gurion University of the Negev Beer Sheva
Dept. of Electrical and Computer Engineering
Beer Sheba, Israel
hava@bgu.ac.il

Chapter C.21

Shlomo Hava received a PhD in Electrical Engineering from the University of Delaware in 1982. Since then he has worked in the Department of Electrical and Computer Engineering at Ben-Gurion University of the Negev, Israel. Currently he is a Faculty Professor and Head of the Microelectronics Laboratory in this department. His present research interests include micro- and nanometer-scaled diffraction gratings in optical elements and light emitters, and the effects of gamma radiation and vacuum on optoelectronic devices.

Masahiro Hiramoto

Institute for Molecular Science
Dept. of Materials Molecular Science
Okazaki, Japan
hiramoto@ims.ac.jp

Chapter E.54

Masahiro Hiramoto received a PhD in Chemistry from Osaka University in 1986. He started research on organic semiconductors and organic solar cells in 1988 at the Graduate School of Engineering, Osaka University. He joined the Institute for Molecular Science in 2008 as Professor. He has published over 130 papers and is the inventor of the blended junction and tandem junction for organic solar cells.

Hideo Hosono

Tokyo Institute of Technology
Materials Research Center of Element Strategy
Yokohama, Japan
hosono@msl.titech.ac.jp

Chapter E.58

Hideo Hosono obtained his PhD in Applied Chemistry from Tokyo Metropolitan University in1982 and became a Professor at Tokyo Institute of Technology in 1999. He is a pioneer of transparent oxide semiconductors represented by IGZO ($InGaZnO_x$)-thin film transistors used to drive LCDs and OLEDs and discovered the iron pnictide superconductor. His current concern is design and exploration of novel semiconductors, superconductors, and catalysts.

Colin Humphreys

University of Cambridge
Dept. Materials Science and Metallurgy
Cambridge, UK
colin.humphreys@msm.cam.ac.uk

Chapter 1

Colin Humphreys is Professor of Materials Science and Director of Research at the University of Cambridge. He is the founder and the Director of the Cambridge Centre for Gallium Nitride, which researches GaN-based LEDs and electronic devices. He is a Fellow of the Royal Society and the Royal Academy of Engineering.

Mehrdad Irannejad

University of Waterloo
Dept. of Mechanical and Mechatronics Engineering
Waterloo, Canada
mehrdad.irannejad@uwaterloo.ca

Chapter D.41

Mehrdad Irannejad received his MSc degrees from the University of Shiraz, Iran (2000) and the University of Leeds, UK (2008). He received his PhD degree in Photonic Materials Science Engineering from the University of Leeds in 2012. Currently, he is a Research Associate in Mechanical and Mechatronics Engineering at the University of Waterloo. His research interests include photonic glasses and polymers, and their integration with 2D materials in nanosensor platforms.

Stuart Irvine

Swansea University
College of Engineering
Swansea, UK
s.j.c.irvine@swansea.ac.uk

Chapters B.14, E.43

Stuart Irvine is Professor of Solar Energy Materials in the College of Engineering, Swansea University, and Director of the Centre for Solar Energy Research. He was awarded his BSc in Physics (1974, Loughborough University), his PhD in Materials Science (1978), and his DSc in Physics (1994, University of Birmingham). His research on metalorganic chemical vapor deposition (MOCVD) of II-VI semiconductors has been applied to infrared detectors and PV solar energy.

Tatsumi Ishihara

Kyushu University
Dept. of Applied Chemistry
Fukuoka, Japan
ishihara@cstf.kyushu-u.ac.jp

Chapter E.59

Tatsumi Ishihara is Professor at the International Institute for Carbon-Neutral Energy Research (I2CNER) at Kyushu University (Japan). He received his PhD from Kyushu University in 1991 and has been working there since 2003. His research interests lie in the fields of catalysts, solid state electrochemistry, solid state ionics, the development of energy-related materials like fuel cells, rechargeable batteries and photocatalysts, environment catalysts like deNOx, and low temperature PM oxidation catalysts for cars.

Radu Ispasoiu

ON-Semiconductor
Image Sensor Group
San José, USA
radu.ispasoiu@onsemi.com

Chapter D.40

Radu Ispasoiu has held engineering management and technical leadership positions in the hi-tech optoelectronics industry in Silicon Valley, California, since 2001. He is currently Senior Manager of Pixel Design and Image Sensor Characterization at ON Semiconductor, Image Sensor Group (formerly Aptina Imaging). He has a PhD in Physics (1996), with a dissertation on the optical spectroscopy of III-V semiconductor quantum well structures.

Minoru Isshiki

Tohoku University
Inst. of Multidisciplinary Research for Advanced Materials
Sendai, Japan
isshiki@agate.plala.or.jp

Chapters B.16, D.33

Minoru Isshiki is an Emeritus Professor, Tohoku University. He retired in 2012 as a Professor at the Institute for Advanced Materials Processing, Tohoku University. He received his PhD in Materials Science from Tohoku University in 1976. His research interests focus mainly on the purification of metals, single crystal growth of bulk, and thin film II-VI compound semiconductors and impurity effects on their properties.

Tim Joyce

University of Liverpool
School of Engineering
Liverpool, UK
t.joyce@liverpool.ac.uk

Chapter B.14

Tim Joyce has 35 years' experience in semiconductor growth and materials characterization, mainly at Liverpool University. He migrated from MOVPE to MOMBE of III-V materials and then to III-nitride MBE and to ALD of oxide films. More recently he has concentrated on AFM and electron microscopy. He is an active committee member of BACG.

M. Zahangir Kabir

Concordia University
Dept. of Electrical and Computer Engineering
Montreal, Canada
zahangir.kabir@concordia.ca

Chapter E.45

Zahangir Kabir is an Associate Professor in the Department of Electrical and Computer Engineering at Concordia University, Canada. His research interests include theory, modeling, and characterization of electronic materials and devices with applications to photoconductors for optical and x-ray detectors, HFETs, and thin-film solar cells. He has authored or co-authored more than 40 refereed journal papers in the above area.

Safa Kasap Chapters A.2, A.3, B.19, C.28, E.45 For biographical profile, please see the section "About the Editors".

Gudrun Kissinger

Innovations for High Performance
Microelectronics (IHP)
Frankfurt/Oder, Germany
gkissinger@ihp-microelectronics.com

Chapter A.5

Gudrun Kissinger obtained a PhD in Crystallography from the University of Leipzig and worked as a Research Scientist and Project Leader in IHP in Frankfurt (Oder), Germany, from 1984–2001. From 2001–2004 she co-founded and joined the start-up company Communicant Semiconductor Technologies. In 2004, she returned to IHP where she is now working as Project Leader in industrial cooperation with Siltronic AG.

Alexander V. Kolobov

National Institute of Advanced Industrial
Science and Technology
Tsukuba, Japan
a.kolobov@aist.go.jp

Chapter E.46

Alexander V. Kolobov has been involved in chalcogenide research since the late 1970s when he joined A.F. Ioffe Physico-Technical Institute in Leningrad (Russia). Since 1994 he has been working at the National Institute of Advanced Industrial Science and Technology in Tsukuba (Japan), where he is currently a Prime Senior Researcher.

Cyril Koughia

University of Saskatchewan
Dept. of Electrical and Computer
Engineering
Saskatoon, Canada
cyril.koughia@usask.ca

Chapters A.2, A.3

Cyril Koughia received his PhD from the A.F. Ioffe Physico-Technical Institute, St. Petersburg, Russia and is a Research Associate at the University of Saskatchewan, Canada. His current research interests include electronic and optical properties of amorphous semiconductors and rare-earth doped glasses used in photonics and electronics. He has written many invited papers and has published extensively in international journals in this field.

Roger Lewis

University of Wollongong
School of Physics
Wollongong, Australia
roger@uow.edu.au

Chapter E.55

Roger Lewis is a Professor of Physics and Associate Dean of Research in the Faculty of Engineering and Information Sciences at the University of Wollongong, Australia. Lewis holds a BSc (Honours) from the University of Sydney and a PhD from Griffith University, Australia. He is a fellow of the Australian Institute of Physics and of the Royal Microscopical Society.

Geoffrey Luckhurst

University of Southampton
Chemistry
Southampton, UK
gl@soton.ac.uk

Chapter D.36

Geoffrey Luckhurst was awarded his PhD in 1965 from the University of Cambridge, having obtained his BSc at the University of Hull. His research into liquid crystals started in Cambridge and continues at Southampton where he is Emeritus Professor of Chemical Physics. His work is concerned with many aspects of this interdisciplinary field including the design of novel materials, their molecular orientational order, macroscopic order and dynamics.

Jiří Málek

University of Pardubice
Dept. of Physical Chemistry
Pardubice, Czech Republic
jiri.malek@upce.cz

Chapter B.19

Jiří Málek received his PhD degree from the Institute of Chemical Technology, Pardubice (1986) and his DSc degree from the Faculty of Chemical Engineering in Prague (2000). He has been a Full Professor of Physical Chemistry at the University of Pardubice since 2002. His interests include kinetics of nucleation-growth processes, structural relaxation, and visco-elastic behavior of non-crystalline materials and highly supercooled, glass-forming liquids.

Akihisa Matsuda

Osaka University
Dept. of Systems Innovation
Osaka, Japan
matsuda@ee.es.osaka-u.ac.jp

Chapter C.25

Akihisa Matsuda has an MS degree from Waseda University, a PhD from Tokyo Institute of Technology, and was Chief Senior Researcher at ETL, MITI Japan. He was also the Director of the Research Initiative for Thin-Film Silicon Solar Cells at AIST, METI Japan. He was also a Professor at the Tokyo University of Science, Chiba, Japan.

Naomi Matsuura

University of Toronto
Dept. of Materials Science & Engineering
Toronto, Canada
naomi.matsuura@utoronto.ca

Chapter D.39

Naomi Matsuura is an Associate Professor in the Faculty of Applied Science and Engineering at the University of Toronto. She received her PhD from the University of Toronto in 2003 for the development of three-dimensional nanoscale architectures. Her research focuses on the design of new materials for enhanced interaction with medical imaging radiation for image-guided and targeted treatment of disease.

Jan Mistrik

University of Pardubice
Center of Materials and Nanotechnology
Pardubice, Czech Republic
jan.mistrik@upce.cz

Chapter A.3

Jan Mistrik is currently Senior Scientist at the Center of Materials and Nanotechnology at the University of Pardubice, Czech Republic. He obtained his MSc (1998) and PhD (2003) from Charles University in Prague (quantum optics and optoelectronics). In 2002, he received a doctoral degree in materials science from the University of Versailles, France. His research interest is optical characterization of nanostructures, in particular spectroscopic ellipsometry and magneto-optics with about 50 refereed journal papers in the field.

Winfried Mönch

University Duisburg-Essen
Faculty of Physics
Duisburg, Germany
winfried.moench@uni-due.de

Chapter A.8

Winfried Mönch received a Dr. rer. nat. degree from the Universität Göttingen in 1961. He spent three years at the AEG-Forschungsinstitut before joining the RWTH Aachen. There he qualified as university lecturer in 1968 and became Associate Professor two years later. In 1974 he was appointed Professor at the Universität Duisburg and retired in 1999. He was Walter Schottky Visiting Professor at Stanford University in 1981 and received the E.W. Müller Award 1984 of the University of Wisconsin–Milwaukee. He has authored two monographs on semiconductor surfaces and interfaces.

Donald T. Morelli

Michigan State University
Chemical Engineering and Materials Science
East Lansing, USA
dmorelli@egr.msu.edu

Chapter E.57

Donald Morelli is Professor of Materials Science and Adjunct Professor of Physics at Michigan State University. His research interests span a variety of topics, including: semimetals, conducting polymers, high temperature superconductors, wide and narrow band gap semiconductors, high thermal conductivity crystals, thermoelectric materials, and magnetism. Professor Morelli holds BS and PhD degrees in Physics from the University of Michigan.

Kazuo Morigaki

University of Tokyo
Tokyo, Japan
k.morigaki@yacht.ocn.ne.jp

Chapter C.24

Kazuo Morigaki received his PhD in Physics from Osaka University in 1959. Since then, he has worked for Osaka University, Sony Corporation Research Laboratory, Centre d'Etudes Nucléares de Saclay, University of Tokyo (Institute for Solid State Physics), Yamaguchi University, and Hiroshima Institute of Technology. He is now Professor Emeritus at the University of Tokyo. His current area of interest are light-induced phenomena and the electronic states of defects in amorphous and microcrystalline silicon.

Hadis Morkoç

Virginia Commonwealth University
Electrical and Computer Engineering
Richmond, USA
hmorkoc@vcu.edu

Chapter D.31

Hadis Morkoç received his PhD degree from Cornell University, Ithaca, NY, and a Honoris Causa from the University of Montpellier II. He is the Founders Professor of Engineering at Virginia Commonwealth University, Richmond, VA. He is among the most cited scientists in the fields of physics, engineering, and materials science. His research interests are semiconductors, oxides, electronic and optoelectronic devices, and sensors.

Jayanta Mukherjee

Compound Photonics UK Ltd
Durham, UK
jayanta.mukherjee@compoundphotonics.com

Chapter D.35

Jayanta Mukherjee is a Laser Design Engineer at Compound Photonics, United Kingdom, where he works on design, development and manufacturing of visible and infra-red high power semiconductor lasers and solid state lasers. He holds a BSc (Hons) in Physics, a Master's in Physics and Laser Engineering, and a PhD in Semiconductor Laser Physics.

Takashi Nagase

Osaka Prefecture University
Dept. of Physics and Electronics
Sakai, Japan
nagase@pe.osakafu-u.ac.jp

Chapter E.51

Takashi Nagase received his PhD degree in Electronic Engineering from Osaka Prefecture University in 2000. After a postdoctoral fellowship at the National Institute of Information and Communications Technology, he joined Osaka Prefecture University in 2006 where he is currently an Associate Professor in the Department of Physics and Electronics. His current research interests are organic electronics, molecular devices, and nanofabrication.

Hiroyoshi Naito

Osaka Prefecture University
Dept. of Physics and Electronics
Sakai, Japan
naito@pe.osakafu-u.ac.jp

Chapter E.51

Hiroyoshi Naito received his PhD degree in Electronic Engineering from Osaka Prefecture University in 1984 where he is now a Professor in the Department of Physics and Electronics. He has been engaged in the photoelectric characterization of disordered semiconductors such as amorphous chalcogenide glasses and of liquid crystalline materials. His current interests are the optical and electronic properties of organic semiconductors and their application to optoelectronic devices.

Arokia Nathan

University of Cambridge
Dept. of Electrical Engineering
Cambridge, UK
an299@eng.cam.ac.uk

Chapter E.44

Arokia Nathan holds the Chair of Photonic Systems and Displays in the Department of Engineering, Cambridge University. He has extensive experience in device physics and modeling, and materials processing. His present research interests lie in the integration of devices, circuits, and systems using a broad range of inorganic and organic thin film material systems on rigid and mechanically flexible substrates.

Stephen K. O'Leary

University of British Columbia
School of Engineering
Kelowna, Canada
stephen.oleary@ubc.ca

Chapter D.32

Stephen O'Leary received degrees from the University of Toronto. Currently, he is with The University of British Columbia. His current research interests are the material properties of electronic materials. Professor O'Leary is a registered Professional Engineer. He is currently a member of the American Physical Society, the Materials Research Society, and a Senior Member of the IEEE.

Chisato Ogihara

Yamaguchi University
Graduate School of Science and Engineering
Yamaguchi, Japan
ogihara@yamaguchi-u.ac.jp

Chapter C.24

Chisato Ogihara received his PhD degree in Physics from University of Tokyo in 1988. He joined the University of Strathclyde in 1989, Gifu University in 1991 and Yamaguchi University in 1993. His current research focuses on photoluminescence and light-induced creation of defects in hydrogenated amorphous silicon.

Fabien Pascal

University of Montpellier
Institute of Electronic and Systems
Montpellier, France
fabien.pascal@ies.univ-montp2.fr

Chapter B.20

Fabien Pascal is involved in a research group working in electronic devices (MOS, MODFET, MESFET, TBH Si–Ge, TBH III–V) by means of their proper background I–V/C–V characteristics and noise spectral analysis. He is also working on the technological qualification of contacts, semiconductor materials and carbon nanotubes based devices by the use of noise spectroscopy.

Michael C. Petty

Durham University
School of Engineering and Computing Sciences
Durham, UK
m.c.petty@durham.ac.uk

Chapter E.51

Michael Petty's higher education was at Sussex University, UK (BSc and DSc) and Imperial College, London (PhD). He is a Fellow of the Institute of Physics and of the Institution of Engineering and Technology. Mike's research activities focus on the properties of organized thin films of organic materials. He has a particular interest in the application of these layers to electronic and opto-electronic devices.

Henry H. Radamson

Chinese Academy of Science
Inst. of Microelectronics
Beijing, China
KTH Royal Institute of Technology
School of Information and Communication Technology
Stockholm, Sweden
rad@ime.ac.cn

Chapter E.48

Henry H. Radamson received the PhD degree in semiconductor materials from Linköping University in Sweden in 1996. He joined the Royal Institute of Technology in Stockholm in 1997 as a Senior Scientist. Since 2016, he is professor in Chinese Academy of Science in Beijing, China. His research field is nanoelectronics, nanophotonics, and nanosensing. He has published more than 200 scientific articles, including seven book chapters and one book about the integration of photonics and electronics.

Asim K. Ray

Brunel University London
Inst. of Materials and Manufacturing
Middlesex, UK
asim.ray@brunel.ac.uk

Chapters C.28, E.52

Asim K. Ray, BSc, MSc, PhD, DSc, holds the Chair of Electronic Materials and Devices at Brunel University London. Dr Ray is a Chartered Engineer and a Chartered Physicist. His nanoelectronics research activity is focused on electron transport in low dimensional nanostructures, organic/inorganic hybrid structures, and discotic liquid organic crystals and 2D graphene. He has published over 250 papers, including invited reviews and book chapters in his field of research.

Stephen Reynolds

University of Dundee
School of Science and Engineering
Dundee, UK
s.z.reynolds@dundee.ac.uk

Chapter A.7

Steve Reynolds received a PhD in Applied Physics (1983), and was BP Postdoctoral Fellow at the University of Edinburgh before moving to Dundee in 1987. He has had several periods in industry, and in 2005/2006 was Guest Scientist in the Institute of Photovoltaics, FZ Jülich, Germany. Steve is currently Programme Leader in Physics and Director of Solar Cities Scotland. His research interests include thin-film semiconductors, computer modeling, and renewable energy systems.

Oleg Rubel

McMaster University
Dept. of Materials Science and Engineering
Hamilton, Canada
rubelo@mcmaster.ca

Chapter A.9

Oleg Rubel received his PhD in Materials Science from the Zaporozhye National Technical University, Ukraine, in 2001. He has worked at the Central Technological Laboratory and the Semiconductor Theory Group, Philipps-University Marburg, Germany and at the Thunder Bay Regional Research Institute, Canada. His current research includes the development of first-principle approaches for predictive modeling of functional electronic materials.

Harry E. Ruda

University of Toronto
Dept. of Materials Science and Engineering, and Edward Rogers Dept. of Electrical and Computer Engineering
Toronto, Canada
harry.ruda@utoronto.ca

Chapters A.2, A.3, D.39

Harry Ruda received his PhD from MIT in 1982. He holds the Stanley Meek Chair in Nanotechnology at the University of Toronto and is Director of the Centre for Nanotechnology. His research interests focus on the fabrication and modeling of quantum functional nanostructures with applications in nanoelectronics and nanophotonics. He is a Fellow of the Royal Society of Canada and the Institute of Physics, and editor of several international scientific journals.

Andreas Sattler

Siltronic AG
München, Germany
andreas.sattler@siltronic.com

Chapter A.5

Andreas Sattler obtained his Doctor rerum naturalium in Physics from the University of Göttingen with a thesis on gettering of cobalt in silicon. His special fields of interest were solid state and semiconductor physics. He then joined the simulation group of Siltronic in 2002, where he is working on simulation of crystal growth with special emphasis on defect formation.

Peyman Servati

University of British Columbia
Electrical and Computer Engineering
Vancouver, Canada
peymans@ece.ubc.ca

Chapter E.44

Professor Peyman Servati is Director of Flexible Electronics and Energy Lab (FEEL) at the Department of Electrical and Computer Engineering, University of British Columbia. He has extensive experience in novel materials and devices for rigid and flexible displays and electronics and electronic textiles. His research interests lie in integrated sensors, electronics, and energy devices in mechanically flexible and textile substrates.

Derek Shaw

Hull University
Dept. of Physics
Hull, UK
derekshaw47@gmail.com

Chapter A.6

Derek Shaw has a BSc in Physics from Manchester University, UK, and a PhD from Imperial College. He worked for nine years in R&D in the electrical industry, primarily on thermionic emitters for magnetron valves and photoconductive CdS. He spent the remainder of his career at the University of Hull, where he initiated research into self and impurity diffusion in CdS, CdTe, HgCdTe, and GaSb.

Fumio Shimura

Shizuoka Institute of Science and Technology
Dept. of Materials and Life Science
Fukuroi, Japan
shimura@ms.sist.ac.jp

Chapter B.13

Fumio Shimura has been engaged in both fundamental and practical science and engineering related to semiconductor crystal technology and semiconductor device processing, and is the author of a book. Fumio Shimura received his PhD degree in Applied Physics from Nagoya University, Japanin 1982. He is currently a Professor at the Sizuoka Institute of Science and Technology and Adjunct Professor, North Carolina State University.

Yusuke Shinmura

Institute for Molecular Science
Dept. of Materials Molecular Science
Okazaki, Japan
shinmura@ims.ac.jp

Chapter E.54

Yusuke Shinmura received a PhD in Science from the Graduate University for Advanced Studies (SOKENDAI) in Hayama, Japan, in 2015. He joined the Institute for Molecular Science in Okazaki, Japan, in 2010. His current research interests are doping effects to organic semiconductors and organic photovoltaic cells.

Michael S. Shur

Rensselaer Polytechnic Institute
Dept. of Physics, Applied Physics, and Astronomy
New York, USA
shurm@rpi.edu

Chapter D.32

Dr Michael Shur is Patricia W. and C. Sheldon Roberts Professor at RPI. His interests are in semiconductor physics. He is Foreign Member of the Lithuanian Academy of Sciences and Fellow of IEEE, APS, IET, ECS, MRS, OSA, SPIE, EMA, and AAAS, received IEEE and other awards, and holds Honorary Doctorates from St. Petersburg Technical University and the University of Vilnius.

Poppy Siddiqua

University of British Columbia
School of Engineering
Kelowna, Canada
siddiquapoppy@gmail.com

Chapter D.32

Poppy Siddiqua is currently a PhD student at the University of British Columbia. Her research interests are focused on the electron transport within wide-energy gap compound semiconductor materials and their device implications. She worked as a Lecturer at Daffodil International University, Bangladesh prior to joining the University of British Columbia as a PhD student.

Jai Singh

Charles Darwin University
School of Engineerng and Information Technology
Darwin, Australia
jai.singh@cdu.edu.au

Chapter A.3

Jai Singh is Professor of Physics at the Charles Darwin University, Darwin, Australia and Fellow of the Australian Institute of Physics. His research interests are in condensed-matter theory, covering areas of excitonic processes in crystalline and amorphous semiconductors, nanostructures, and design of solar cells, OLEDs, and scintillators. He has written two books, edited three books and five conference proceedings, and published 200 research papers.

Tim Smeeton

SHARP Laboratories of Europe
Oxford, UK
tim.smeeton@sharp.co.uk

Chapter 1

Tim Smeeton is a Chief Technologist at SHARP Laboratories of Europe, leading development of novel light-emitting devices and photonic systems, including lasers, LEDs, deep ultraviolet (UVC) technologies, and photonics-based sensors. He received his PhD from the University of Cambridge in 2005 for research into the nanostructures of InGaN quantum wells using transmission electron microscopy and X-ray scattering techniques.

Hitoshi Suzuki

Hiroshima University
Graduate School of Advanced Sciences of Matter
Higashihiroshima, Japan
hitoshi-suzuki@hiroshima-u.ac.jp

Chapter E.51

Hitoshi Suzuki received his PhD degree in Applied Physics from Tokyo Institute of Technology and then joined the Communications Research Laboratory (National Institute of Information and Communications Technology). In 2007 he joined Hiroshima University. His current research focuses on the analysis of self-assembled structures of organic molecules using scanning probe microscopy and fabrication of functional nanostructures for biomolecules.

Roman Svoboda

University of Pardubice
Dept. of Physical Chemistry
Pardubice, Czech Republic
roman.svoboda@upce.cz

Chapter B.19

Roman Svoboda was born in 1980 in Hradec Krylova, Czech Republic. He received his PhD degree in Physical Chemistry from the University of Pardubice (2008), where he currently works as a senior scientist. His interests include advanced kinetic analysis of complex physical-chemical processes, structural relaxation phenomena in non-crystalline materials, and novel methods of thermal analysis.

Stephen J. Sweeney

University of Surrey
Dept. of Physics and Advanced Technology Institute
Guildford, UK
s.sweeney@surrey.ac.uk

Chapter D.35

Stephen John Sweeney is Professor of Physics and Head of the Department of Physics at the University of Surrey, UK. He holds a BSc (Hons) in Applied Physics and a PhD in Semiconductor Laser Physics. His research interests are in low-dimensional and novel semiconductors for use in photonic devices with applications including communications, energy generation and efficiency, sensing, and healthcare.

David Sykes

Loughborough Surface Analysis Ltd
Loughborough, UK
d.e.sykes@lsaltd.co.uk

Chapter B.18

David Sykes is a Director of Loughborough Surface Analysis Ltd. He was Chairman of the UK Surface Analysis Forum (UKSAF) 2000–2015, Chairman of the British Vacuum Council (BVC) 2006–2008, Recording Secretary of the International Union for Vacuum Science Technique and Application (IUVSTA) 2004–2016, and Chairman of the ISO/TC201 Subcommittee on Data Management and Treatment in Surface Chemical Analysis 1993–2017.

Keiji Tanaka

Hokkaido University
Dept. Applied Physics
Sapporo, Japan
keiji@eng.hokudai.ac.jp

Chapter D.42

After graduating from Hokkaido University in 1970, Keiji Tanaka worked at Canon Co. Ltd. for two years. He then returned to Hokkaido University and was promoted to Professor in 1991. In 2011 he became an Emeritus Professor. He has worked on physical properties and photoinduced phenomena of amorphous materials, specifically chalcogenide glasses, and received the first Ovshinsky Award in 2001.

Charbel Tannous

University of Western Brittany
Lab. of Magnetism of Brittany
Brest, France
charbel.tannous@univ-brest.fr

Chapters A.4, E.49

Charbel Tannous holds a DSc from Joseph Fourier University (Grenoble, France) and a PhD from the University of Sherbrooke (Quebec). He has worked on *1/f* noise in fractal circuits and microelectronic device simulation, and also on wireless communications and non-linear signal processing. He is now a Full Professor of Physics working on fast switching magnetization problems for high density magnetic recording and magnetic sensors.

Ali Teke

Balikesir University
Dept. of Physics
Balikesir, Turkey
ateke@balikesir.edu.tr

Chapter D.31

Ali Teke is currently an Associate Professor in the Physics Department at the Balikesir University, Turkey. He received his BSc (1992) and PhD (1997) degrees from Physics Departments of the Middle East Technical University, Turkey and Essex University, UK, respectively. His current research activities include the investigation of the electronic and optical properties of wide-band-gap semiconductors such as GaNand ZnO and exploring new devices for electronic and optoelectronic applications. He has published more than 30 journal and conference papers, and chapters in books.

Junji Tominaga

National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan
j-tominaga@aist.go.jp

Chapter E.46

Junji Tominaga is a Prime Senior Researcher at the National Institute of Advanced Industrial Science and Technology (AIST), Japan. He received his PhD from Cranfield Institute of Technology, UK (1991). He was the Director of the Center for Applied Near-Field Optics Research until 2009. His current research is on non-volatile solid-state phase change memory using superlattice structures.

Harry Tuller

Massachusetts Institute of Technology
Dept. of Materials Science and Engineering
Cambridge, USA
tuller@mit.edu

Chapter A.11

Harry Tuller received his EngScD from Columbia University (1973) and honorary doctorates from Universities of Marseille and Oulu. He is Editor-in-Chief, Journal of Electroceramics and President, International Society of Solid State Ionics, Fellow American Ceramic and Electrochemical Societies, von Humboldt awardee, and co-founder of Boston MicroSystems. His research focuses on defects and properties of metal oxides with applications to sensors, energy conversion, memory, and MEMS devices.

Kazushige Ueda

Kyushu Institute of Technology
Dept. of Materials Science
Kitakyushu, Japan
kueda@che.kyutech.ac.jp

Chapter E.58

Kazushige Ueda received a Master's degree in Inorganic Materials in 1993 and a Doctoral degree in Materials Science in 1998 from Tokyo Institute of Technology (Tokyo Tech.). He studied transparent conducting materials at the Materials and Structures Laboratory at Tokyo Tech. After moving to Kyushu Institute of Technology in 2003, he started research on luminescent inorganic materials.

Robert M. Wallace

University of Texas at Dallas
Dept. of Materials Science and Engineering
Richardson, USA
rmwallace@utdallas.edu

Chapter C.27

Robert M. Wallace earned his PhD in Physics from the University of Pittsburgh in 1988. In 1990, Wallace joined Texas Instruments, working on advanced device concepts and associated integration issues. In 2003, he joined the University of Texas at Dallas, where he is Professor of Materials Science and Engineering. His research interests include surfaces and interfaces of electronic materials. He has over 300 publications and 65 US and international patents.

Jifeng Wang

Rasa Industries Ltd.
Advanced Microelectronic Materials
Development Center
Tokyo, Japan
wypwm6866@yahoo.co.jp

Chapters B.16, D.33

Jifeng Wang is a Senior Manager at Rasa Industries, Ltd., Japan. He retired from Tohoku University in 2008 as an Associate Professor. He received his PhD in Solid State Physics from the Chinese Academy of Sciences in 1991. His research interests focus on the microelectronic materials and he is the author of over 90 publications.

David S. Weiss

Molecular Glasses, Inc.
Rochester, USA
dweiss2@rochester.rr.com

Chapter D.37

David S. Weiss is the Research Director of Molecular Glasses, Inc. and a Senior Scientist at the University of Rochester. He retired from Eastman Kodak in 2009 as a Scientist Fellow. He received his PhD in Chemistry from Columbia University in 1969. His research interests focus on the photochemistry, photophysics, and photoelectrical properties of organic materials. He holds 26 US patents and has authored over 100 publications and co-authored one book.

Rainer Wesche

Swiss Federal Institute of Technology
Lausanne
Swiss Plasma Center
Villigen, Switzerland
rainer.wesche@psi.ch

Chapter E.50

Rainer Wesche studied physics at the University of Constance, Germany. After completing his diploma in 1984, he was assistant from 1985 to 1989 (PhD in 1988). From 1989 to 1993 he was a research scientist at the Paul Scherrer Institute in Switzerland, where he led an experimental study of high-current applications of high-T_c superconductors funded by the Swiss National Science Foundation. Since 1994 he has been a research scientist at the Swiss Federal Institute of Technology Lausanne. His present research is in the field of applied superconductivity.

Roger Whatmore

Imperial College London
Dept. of Materials
London, UK
roger.whatmore@gmail.com

Chapter C.26

Roger Whatmore worked with ferroelectrics for more than 40 years. A winner of the Nelson Gold Medal and the Griffith Medal, he is SRA at Imperial College London and Emeritus Professor at University College Cork. He is a Fellow of the Royal Academy of Engineering and the Institutes of Physics and Materials, Minerals and Mining, and a Member of the Royal Irish Academy.

Neil White

University of Southampton
School of Electronics and Computer Science
Southampton, UK
nmw@ecs.soton.ac.uk

Chapter C.29

Neil White is Professor of Intelligent Sensor Systems in the Department of Electronics and Computer Science at the University of Southampton, UK. He is a Fellow of the Institution of Engineering and Technology (IET) and the Institute of Physics (IOP), as well as a Senior Member of the IEEE. Professor White has published over 200 papers in the field of sensor technology and novel sensing materials.

Congyong Zhu

Infineon Technologies
Chandler, USA
zhuc@mymail.vcu.edu

Chapter D.31

Congyong Zhu received his BSc from Wuhan University and his MPhil from The University of Hong Kong. He received his PhD in Electrical and Computer Engineering from Virginia Commonwealth University (VCU). His research experience in VCU expands from epitaxy growth, electrical characterization, simulation, phase noise, and reliability study of GaN based HEMTs to phase shifters based on $BaSrTiO_3$. In 2013, he joined International Rectifier Inc. (now Infineon Technologies Americas Corp.).

Detailed Contents

List of Abbreviations XXIX

1 **Perspectives on Electronic and Photonic Materials**
Tim Smeeton, Colin Humphreys 1
1.1 Tremendous Integration 2
1.2 The Silicon Age 3
1.2.1 The Transistor and Early Semiconductor Materials Development 3
1.2.2 The Integrated Circuit 5
1.3 The Compound Semiconductors 7
1.3.1 High-Speed Electronics 8
1.3.2 Light Emitting Devices 9
1.3.3 The III-Nitrides 11
References 14

Part A Fundamental Properties

2 **Electrical Conduction in Metals and Semiconductors**
Safa Kasap, Cyril Koughia, Harry E. Ruda 19
2.1 Fundamentals: Drift Velocity, Mobility and Conductivity 20
2.2 Matthiessen's Rule 22
2.3 Resistivity of Metals 23
2.3.1 General Characteristics 23
2.3.2 Fermi Electrons 25
2.4 Solid Solutions and Nordheim's Rule 26
2.5 Carrier Scattering in Semiconductors 28
2.6 The Boltzmann Transport Equation 29
2.7 Resistivity of Thin Polycrystalline Films 30
2.8 Inhomogeneous Media: Effective Media Approximation 32
2.9 The Hall Effect 35
2.10 High Electric Field Transport 37
2.11 Impact Ionization 38
2.12 Two-Dimensional Electron Gas 40
2.13 One-Dimensional Conductance 42
2.14 The Quantum Hall Effect 43
References 44

3 **Optical Properties of Electronic Materials: Fundamentals and Characterization**
Jan Mistrik, Safa Kasap, Harry E. Ruda, Cyril Koughia, Jai Singh 47
3.1 Optical Constants 47
3.1.1 Refractive Index and Extinction Coefficient 48
3.1.2 Kramers–Kronig Relations 49

3.2 Refractive Index 50
3.2.1 Cauchy Dispersion Equation 50
3.2.2 Sellmeier Dispersion Equation 50
3.2.3 Gladstone–Dale Formula 51
3.2.4 Wemple–DiDominico Dispersion Relation 52
3.2.5 Group Index (N_g) 53
3.3 Optical Absorption 53
3.3.1 Lattice or Reststrahlen Absorption and Infrared Reflection 53
3.3.2 Free-Carrier Absorption (FCA) 54
3.3.3 Band-to-Band or Fundamental Absorption 56
3.3.4 Exciton Absorption 62
3.3.5 Impurity Absorption 65
3.3.6 Effects of External Fields 68
3.4 Optical Characterization 69
3.4.1 Bulk Samples 71
3.4.2 Thin Film Optics 75
3.5 Optical Materials 79
3.5.1 Abbe Number or Constringence 79
3.5.2 Optical Materials 79
3.5.3 Optical Glasses 80
References 82

4 **Magnetic Properties: From Traditional to Spintronic**
Charbel Tannous, Jacek Gieraltowski 85
4.1 Traditional Magnetism 87
4.1.1 Fundamental Magnetic Quantities 88
4.1.2 The Hysteresis Loop 89
4.1.3 Intrinsic Magnetic Properties of a Material 93
4.1.4 Traditional Types of Magnetism and Classes of Magnetic Materials 96
4.2 Nonconventional Magnetism and Progress Toward Spintronics and Quantum Devices 99
4.2.1 Types of Exchange and Coupling in Magnetic Materials 99
4.2.2 Thin Magnetic Films 100
4.2.3 Electronic Properties: Localized, Free, Itinerant Magnetism and Spin-Polarized Band Structure 101
4.3 Spintronics and Quantum Information Devices 104
4.3.1 Magnetic Signal Processing Devices 106
4.3.2 Magnetic Quantum Dot Arrays 107
4.3.3 Magnetic Vortex Properties and Applications 107
References 109

5 **Defects in Monocrystalline Silicon**
Wilfried von Ammon, Andreas Sattler, Gudrun Kissinger 111
5.1 Technological Impact of Intrinsic Point Defects Aggregates 112
5.2 Thermophysical Properties of Intrinsic Point Defects 113

5.3 Aggregates of Intrinsic Point Defects 115
5.3.1 Experimental Observations 115
5.3.2 Theoretical Model: Incorporation of Intrinsic Point Defects 118
5.3.3 Theoretical Model: Aggregation of Intrinsic Point Defects 120
5.3.4 Effect of Impurities on Intrinsic Point Defect Aggregation 123
5.4 Formation of OSF Ring 127
References 129

6 **Diffusion in Semiconductors**
Derek Shaw 133
6.1 Basic Concepts 134
6.2 Diffusion Mechanisms 134
6.2.1 Vacancy and Interstitial Diffusion Mechanisms 135
6.2.2 The Interstitial–Substitutional Mechanism: Dissociative and Kick-Out Mechanisms 135
6.2.3 The Percolation Mechanism 135
6.3 Diffusion Regimes 136
6.3.1 Chemical Equilibrium: Self- and Isoconcentration Diffusion 136
6.3.2 Chemical Diffusion (or Diffusion Under Nonequilibrium Conditions) 136
6.3.3 Recombination-Enhanced Diffusion 138
6.3.4 Surface Processing Effects 138
6.3.5 Short-Circuit Paths 138
6.4 Internal Electric Fields 138
6.5 Measurement of Diffusion Coefficients 139
6.5.1 Anneal Conditions 139
6.5.2 Diffusion Sources 139
6.5.3 Profiling Techniques 139
6.5.4 Calculating the Diffusion Coefficient 140
6.6 Hydrogen in Semiconductors 140
6.7 Diffusion in Group IV Semiconductors 141
6.7.1 Germanium 141
6.7.2 Silicon 141
6.7.3 $Si_{1-x}Ge_x$ Alloys 141
6.7.4 Silicon Carbide 142
6.8 Diffusion in III–V Compounds 143
6.8.1 Self-Diffusion 143
6.8.2 Dopant Diffusion 143
6.8.3 Compositional Interdiffusion 144
6.9 Diffusion in II–VI Compounds 144
6.9.1 Self-Diffusion 145
6.9.2 Chemical Self-Diffusion 145
6.9.3 Dopant Diffusion 145
6.9.4 Compositional Interdiffusion 146
6.10 Nano Volume Diffusion 146
6.11 Diffusion in Molten Semiconductors 146
6.12 The Meyer–Neldel Rule 146

6.13 Conclusions 147
6.14 General Reading and References 147
References 147

7 **Photoconductivity in Materials Research**
Stephen Reynolds, Monica Brinza, Mohammed L. Benkhedir, Guy J. Adriaenssens 151
7.1 Steady-State Photoconductivity (SSPC) 153
7.1.1 Definitions and Overview 153
7.1.2 Example Applications in Materials Research 154
7.2 Constant Photocurrent Method (CPM) and Related Techniques 157
7.2.1 CPM 157
7.2.2 Dual Beam Photoconductivity (DBP) 158
7.2.3 Fourier-Transform Photocurrent Spectroscopy (FTPS) 159
7.3 Steady-State Photocarrier Grating Method (SSPG) 160
7.4 Modulated Photocurrent Spectroscopy (MPC) 161
7.4.1 MPC Background and Experiment 161
7.4.2 MPC Density of States Analysis 162
7.4.3 MPC Applications 163
7.5 Switch-on and Switch-off Transients 164
7.5.1 Switch-on Transient 164
7.5.2 Switch-off Transient 165
7.6 Transient Photocurrent Spectroscopy (TPC) 166
7.6.1 TPC Principles 166
7.6.2 TPC Experiment 166
7.6.3 TPC Density-of-States Analysis 167
7.6.4 TPC Applications 168
7.7 Time-of-Flight (TOF) and Related Techniques 168
7.7.1 TOF Mobility and DOS Measurements 168
7.7.2 Interrupted Field TOF (IFTOF) 170
7.8 Other Photoconductivity-Related Techniques 171
7.8.1 Surface Photovoltage (SPV) 171
7.8.2 Spin-Dependent Recombination 171
7.8.3 Time-Resolved Microwave Conductivity (TRMC) 171
References 172

8 **Electronic Properties of Semiconductor Interfaces**
Winfried Mönch 175
8.1 Experimental Database 177
8.1.1 Barrier Heights of Laterally Homogeneous Schottky Contacts 177
8.1.2 Band Offsets of Semiconductor Heterostructures 180
8.2 IFIGS-and-Electronegativity Theory 181
8.3 Comparison of Experiment and Theory 183
8.3.1 Barrier Heights of Schottky Contacts 183
8.3.2 Band Offsets of Semiconductor Heterostructures 184
8.4 Modifications of Schottky Contacts 187
8.4.1 Nonalloyed Ohmic Contacts or MUTIS Schottky Contacts 187
8.4.2 Atomic Interlayers 189

8.5 Graphene Schottky Contacts 190
8.6 Final Remarks 190
References 191

9 **Charge Transport in Disordered Materials**
Sergei Baranovskii, Oleg Rubel 193
9.1 General Remarks on Charge Transport in Disordered Materials 195
9.2 Charge Transport in Disordered Materials via Extended States 198
9.3 Hopping Charge Transport in Disordered Materials via Localized States 200
9.3.1 Nearest-Neighbor Hopping 201
9.3.2 Variable-Range Hopping 203
9.3.3 Description of Charge-Carrier Energy Relaxation and Hopping Conduction in Inorganic Noncrystalline Materials 205
9.3.4 Description of Charge-Carrier Energy Relaxation and Hopping Conduction in Organic Noncrystalline Materials 211
9.4 Concluding Remarks 215
References 216

10 **Dielectric Response**
Leonard Dissado 219
10.1 Definition of Dielectric Response 220
10.1.1 Relationship to Capacitance 220
10.1.2 Frequency-Dependent Susceptibility 220
10.1.3 Relationship to Refractive Index 221
10.2 Frequency-Dependent Linear Responses 222
10.2.1 Resonance Response 222
10.2.2 Relaxation Response 224
10.3 Information Contained in the Relaxation Response 228
10.3.1 The Dielectric Increment for a Linear Response χ_0 228
10.3.2 The Characteristic Relaxation Time (Frequency) 231
10.3.3 The Relaxation Peak Shape 237
10.4 Charge Transport 240
10.5 Data Presentation 243
10.6 A Few Final Comments 243
References 244

11 **Ionic Conduction and Applications**
Harry Tuller 247
11.1 Conduction in Ionic Solids 248
11.2 Fast Ion Conduction 251
11.2.1 Structurally Disordered Crystalline Solids 251
11.2.2 Amorphous Solids 254
11.2.3 Heavily Doped Defective Solids 254
11.2.4 Interfacial Ionic Conduction and Nanostructural Effects 255
11.3 Mixed Ionic–Electronic Conduction 256
11.3.1 Defect Equilibria 256
11.3.2 Electrolytic Domain Boundaries 257

11.4 Applications 258
11.4.1 Sensors 258
11.4.2 Solid Oxide Fuel Cells (SOFC) 260
11.4.3 Membranes 261
11.4.4 Batteries 261
11.4.5 Electrochromic Windows 261
11.5 Future Trends 262
References 263

Part B Growth and Characterization

12 **Bulk Crystal Growth: Methods and Materials**
Peter Capper 269
12.1 Background 269
12.2 History 270
12.3 Techniques 271
12.3.1 Verneuil 271
12.3.2 Czochralski 271
12.3.3 Kyropoulos 273
12.3.4 Stepanov 273
12.3.5 Edge-Defined Film Growth 273
12.3.6 Bridgman 273
12.3.7 Vertical Gradient Freeze 275
12.3.8 Float Zone 275
12.3.9 Traveling Heater Method 275
12.3.10 Low-Temperature Solution Growth 276
12.3.11 High-Temperature Solution Growth (Flux) 276
12.3.12 Hydrothermal 277
12.3.13 Growth from the Vapor 278
12.3.14 Multicrystalline Si Growth for Solar Cells 278
12.4 Materials Grown 278
12.4.1 Group IV 278
12.4.2 Groups III–V 280
12.4.3 Groups II–VI 283
12.4.4 Oxides/Halides/Phosphates/Borates 286
12.5 Conclusions 289
References 289

13 **Single-Crystal Silicon: Growth and Properties**
Fumio Shimura 293
13.1 Overview 294
13.2 Starting Materials 295
13.2.1 Metallurgical-Grade Silicon 295
13.2.2 Polycrystalline Silicon 295
13.3 Single-Crystal Growth 296
13.3.1 Floating-Zone Method 297
13.3.2 Czochralski Method 299
13.3.3 Impurities in Czochralski Silicon 301
13.4 New Crystal Growth Methods 304
13.4.1 Czochralski Growth with an Applied Magnetic Field (MCZ) 304

13.4.2 Continuous Czochralski Method (CCZ) 305
13.4.3 Neckingless Growth Method 306
References 306

14 **Epitaxial Crystal Growth: Methods and Materials**
Peter Capper, Stuart Irvine, Tim Joyce 309
14.1 Liquid-Phase Epitaxy (LPE) 309
14.1.1 Introduction and Background 309
14.1.2 History and Status 310
14.1.3 Characteristics 310
14.1.4 Apparatus and Techniques 311
14.1.5 Group IVs 313
14.1.6 Group III–Vs 314
14.1.7 Group II–VIs 316
14.1.8 Garnets 318
14.1.9 Oxides/Fluorides 318
14.1.10 Atomically Flat Surfaces 318
14.1.11 Conclusions 319
14.2 Metal Organic Chemical Vapor Deposition 319
14.2.1 Introduction and Background 319
14.2.2 Basic Reaction Kinetics 319
14.2.3 Precursors 321
14.2.4 Reactor Cells 322
14.2.5 III–V MOCVD 324
14.2.6 II–VI MOCVD 327
14.2.7 Conclusions 329
14.3 Molecular Beam Epitaxy (MBE) 329
14.3.1 Introduction and Background 329
14.3.2 Reaction Mechanisms 330
14.3.3 MBE Growth Systems 331
14.3.4 Gas Sources in MBE 333
14.3.5 Growth of III–V Materials by MBE 334
14.3.6 Conclusions 337
References 337

15 **Narrow Bandgap II-VI Semiconductors: Growth**
Peter Capper 343
15.1 Bulk Growth Techniques 344
15.1.1 Phase Equilibria 344
15.1.2 Crystal Growth 345
15.1.3 Material Characterization 347
15.2 Liquid Phase Epitaxy (LPE) 349
15.2.1 Hg-rich Growth 349
15.2.2 Te-rich Growth 350
15.2.3 Material Characteristics 351
15.2.4 Device Status 352
15.3 Metal-Organic Vapor Phase Epitaxy (MOVPE) 353
15.3.1 Substrate Type and Orientation 356
15.3.2 Doping 357
15.3.3 In situ Monitoring 357

15.4 Molecular Beam Epitaxy (MBE) 358
15.4.1 Double-Layer Heterojunction Structures 359
15.4.2 Multilayer Heterojunction Structures 359
15.4.3 MCT and CdZnTe Growth on Silicon 360
15.5 Alternatives to MCT 361
References 362

16 **Wide-Bandgap II-VI Semiconductors: Growth and Properties**
Minoru Isshiki, Jifeng Wang 365
16.1 Crystal Properties 367
16.1.1 Basic Properties 367
16.1.2 Phase Diagram 368
16.2 Epitaxial Growth 371
16.2.1 The LPE Technique 371
16.2.2 Vapor-Phase Epitaxy Techniques 372
16.3 Bulk Crystal Growth 375
16.3.1 The CVT and PVT Techniques 375
16.3.2 Hydrothermal Growth 377
16.3.3 Bridgman and Gradient Freezing (GF) Method 379
16.3.4 The Traveling Heater Method (THM) 380
16.3.5 Other Methods 380
16.4 Conclusions 381
References 381

17 **Structural Characterization**
Paul D. Brown 385
17.1 Radiation-Material Interactions 386
17.2 Particle-Material Interactions 387
17.3 X-ray Diffraction 390
17.4 Optics, Imaging and Electron Diffraction 393
17.4.1 Electron Diffraction and Image Contrast Analysis 397
17.4.2 Microdiffraction and Polarity 399
17.4.3 Reflection High-Energy Electron Diffraction 401
17.5 Characterizing Functional Activity 403
17.6 Sample Preparation 404
17.7 Case Studies – Complementary Characterization of Electronic and Optoelectronic Materials 406
17.7.1 Identifying Defect Sources Within Homoepitaxial GaN 407
17.7.2 Cathodoluminescence/Correlated TEM Investigation of Epitaxial GaN 408
17.7.3 Scanning Transmission Electron Beam Induced Conductivity of $Si/Si_{1-x}Ge_x/Si(001)$ 409
17.8 Concluding Remarks 411
References 412

18 **Surface Chemical Analysis**
David Sykes 413
18.1 Electron Spectroscopy 414
18.1.1 Auger Electron Spectroscopy 414
18.1.2 X-Ray Photoelectron Spectroscopy (XPS) 415
18.2 Glow-Discharge Spectroscopies (GDOES and GDMS) 416

18.3 Secondary Ion Mass Spectrometry (SIMS) 417
18.4 Conclusion 423

19 **Thermal Properties and Thermal Analysis: Fundamentals, Experimental Techniques and Applications**
Safa Kasap, Jiří Málek, Roman Svoboda 425
19.1 Heat Capacity 426
19.1.1 Fundamental Debye Heat Capacity of Crystals 426
19.1.2 Specific Heat Capacity of Selected Groups of Materials 428
19.2 Thermal Conductivity 431
19.2.1 Definition and Typical Values 431
19.2.2 Thermal Conductivity of Crystalline Insulators 432
19.2.3 Thermal Conductivity of Noncrystalline Insulators 433
19.2.4 Thermal Conductivity of Metals 435
19.3 Thermal Expansion 436
19.3.1 Grüneisen's Law and Anharmonicity 436
19.3.2 Thermal Expansion Coefficient α_L 437
19.4 Enthalpic Thermal Properties 438
19.4.1 Enthalpy, Heat Capacity and Phase Transformations 438
19.4.2 Conventional Differential Scanning Calorimetry (DSC) 440
19.5 Temperature-Modulated DSC (TMDSC) 446
19.5.1 TMDSC Principles 446
19.5.2 TMDSC Applications 447
19.5.3 Tzero Technology 448
References 449

20 **Electrical Characterization of Semiconductor Materials and Devices**
M. Jamal Deen, Fabien Pascal 453
20.1 Resistivity 454
20.1.1 Bulk Resistivity 454
20.1.2 Contact Resistivity 459
20.2 Hall Effect 462
20.2.1 Physical Principles 462
20.2.2 Hall Scattering Factor 463
20.3 Capacitance–Voltage Measurements 464
20.3.1 Average Doping Density by Maximum–Minimum High-Frequency Capacitance Method 464
20.3.2 Doping Profile by High-Frequency and High–Low Frequency Capacitance Methods 466
20.3.3 Density of Interface States 467
20.4 Current–Voltage Measurements 470
20.4.1 *I–V* Measurements on a Simple Diode 470
20.4.2 *I–V* Measurements on a Simple MOSFET 470
20.4.3 Floating Gate Measurements 471
20.5 Charge Pumping 472
20.6 Low-Frequency Noise 473
20.6.1 Introduction 473
20.6.2 Noise from the Interfacial Oxide Layer 474

20.6.3 Impedance Considerations During Noise Measurement ... 475
20.7 Deep-Level Transient Spectroscopy ... 477
References ... 479

Part C Materials for Electronics

21 **Single-Crystal Silicon: Electrical and Optical Properties**
Mark Auslender, Shlomo Hava ... 485
21.1 Silicon Basics ... 485
21.1.1 Structure and Energy Bands ... 485
21.1.2 Impurity Levels and Charge-Carrier Population ... 488
21.1.3 Carrier Concentration, Electrical and Optical Properties ... 490
21.1.4 Theory of Electrical and Optical Properties ... 491
21.2 Electrical Properties ... 495
21.2.1 Ohm's Law Regime ... 495
21.2.2 High-Electric-Field Effects ... 507
21.2.3 Review Material ... 514
21.3 Optical Properties ... 515
21.3.1 Diversity of Silicon as an Optical Material ... 515
21.3.2 Measurements of Optical Constants ... 515
21.3.3 Modeling of Optical Constants ... 516
21.3.4 Electric-Field and Temperature Effects on Optical Constants ... 520
References ... 520

22 **Silicon-Germanium: Properties, Growth and Applications**
Yaser M. Haddara, Peter Ashburn, Darren M. Bagnall ... 523
22.1 Physical Properties of Silicon-Germanium ... 524
22.1.1 Critical Thickness ... 524
22.1.2 Band Structure ... 526
22.1.3 Dielectric Constant ... 527
22.1.4 Density of States ... 527
22.1.5 Majority-Carrier Mobility in Strained $Si_{1-x}Ge_x$ ... 527
22.1.6 Majority-Carrier Mobility in Tensile-Strained Si on Relaxed $Si_{1-x}Ge_x$ ... 528
22.1.7 Minority-Carrier Mobility in Strained $Si_{1-x}Ge_x$ ... 529
22.1.8 Apparent Band-Gap Narrowing in $Si_{1-x}Ge_x$ HBTs ... 529
22.2 Optical Properties of SiGe ... 530
22.2.1 Dielectric Functions and Interband Transitions ... 530
22.2.2 Photoluminescence ... 530
22.2.3 SiGe Quantum Wells ... 532
22.3 Growth of Silicon-Germanium ... 534
22.3.1 In situ Hydrogen Bake ... 534
22.3.2 Hydrogen Passivation ... 534
22.3.3 Ultra-Clean Epitaxy Systems ... 534
22.3.4 $Si_{1-x}Ge_x$ Epitaxy ... 535
22.3.5 Selective $Si_{1-x}Ge_x$ Epitaxy ... 535
22.3.6 Heteroepitaxial Ge on Si ... 537

22.4 Polycrystalline Silicon-Germanium 537
22.4.1 Electrical Properties of Polycrystalline $Si_{1-x}Ge_x$ 538
References 539

23 **Temperature-Insensitive Band-Gap III-V Semiconductors: Tl-III-V and III-V-Bi**
Hajime Asahi 543
23.1 Tl-III-V Alloy Semiconductors 544
23.1.1 Expected Properties of Tl-Based III-V Alloys 544
23.1.2 Growth of Tl-III-V Alloys 546
23.1.3 Temperature Dependence of Physical Properties 547
23.1.4 Tl-III-V LD Application 549
23.2 III-V-Bi Alloy Semiconductors 550
23.2.1 Expected Properties of Bi-Based III-V Alloys 550
23.2.2 Growth of III-V-Bi Alloys 551
23.2.3 Optical Properties of III-V-Bi Alloys 552
23.2.4 III-V-Bi LD Application 553
23.3 Summary 554
References 554

24 **Amorphous Semiconductors: Structure, Optical, and Electrical Properties**
Kazuo Morigaki, Chisato Ogihara 557
24.1 Electronic States 557
24.2 Structural Properties 560
24.2.1 General Aspects 560
24.2.2 a-Si:H and Related Materials 560
24.2.3 Chalcogenide Glasses 561
24.3 Optical Properties 561
24.3.1 General Aspects 561
24.3.2 a-Si:H and Related Materials 563
24.3.3 Chalcogenide Glasses 564
24.4 Electrical Properties 565
24.4.1 General Aspects 565
24.4.2 a-Si:H and Related Materials 566
24.4.3 Chalcogenide Glasses 567
24.5 Light-Induced Phenomena 567
24.6 Nanosized Amorphous Structure 569
References 570

25 **Amorphous and Microcrystalline Silicon**
Akihisa Matsuda 573
25.1 Reactions in SiH_4 and SiH_4/H_2 Plasmas 574
25.2 Film Growth on a Surface 575
25.2.1 Growth of a-Si:H 575
25.2.2 Growth of μc-Si:H 576
25.3 Defect Density Determination for a-Si:H and μc-Si:H 580
25.3.1 Dangling Bond Defects 580
25.3.2 Dangling Bond Defect Density in μc-Si:H 582
25.4 Device Applications 582

25.5 Recent Progress in Material Issues Related to Thin-Film Silicon Solar Cells 583
25.5.1 Controlling Photoinduced Degradation in a-Si:H 583
25.5.2 High Growth Rates of Device-Grade μc-Si:H 584
25.6 Summary 586
References 586

26 **Ferroelectric Materials**
Roger Whatmore 589
26.1 Definitions and Background 589
26.1.1 Basic Ferroelectric Characteristics and Models 591
26.2 Ferroelectric Materials 593
26.2.1 Ferroelectric Oxides 593
26.2.2 Triglycine Sulphate (TGS) 598
26.2.3 Polymeric Ferroelectrics 599
26.3 Ferroelectric Materials Fabrication Technology 600
26.3.1 Single Crystals 600
26.3.2 Ceramics 601
26.3.3 Thick Films 604
26.3.4 Thin Films 605
26.4 Ferroelectric Applications 607
26.4.1 Dielectrics 607
26.4.2 Computer Memories 607
26.4.3 Piezoelectrics 608
26.4.4 Pyroelectrics 611
References 613

27 **Dielectric Materials for Microelectronics**
Robert M. Wallace 615
27.1 Overview 615
27.1.1 The Scaling of Integrated Circuits 615
27.1.2 Role of Dielectrics for ICs 619
27.2 Gate Dielectrics 620
27.2.1 Transistor Structure 620
27.2.2 Transistor Dielectric Requirements in View of Scaling 621
27.2.3 Silicon Dioxide 625
27.2.4 Silicon Oxynitride: SiO_xN_y 631
27.2.5 High-*k* Dielectrics 633
27.3 Isolation Dielectrics 635
27.4 Capacitor Dielectrics 636
27.4.1 Types of IC Memory 636
27.4.2 Capacitor Dielectric Requirements in View of Scaling 636
27.4.3 Dielectrics for Volatile Memory Capacitors 637
27.4.4 Dielectrics for Nonvolatile Memory 637
27.5 Interconnect Dielectrics 639
27.5.1 Tetraethoxysilane (TEOS) 640
27.5.2 Low-*k* Dielectrics 640
27.6 Summary 641
References 641

28 **Thin Films**
Robert D. Gould†, Safa Kasap, Asim K. Ray 645
28.1 Deposition Methods 647
28.1.1 Physical Deposition Methods 647
28.1.2 Chemical Deposition Methods 662
28.1.3 Sol–Gel Method 667
28.2 Structure 669
28.2.1 Crystallography 669
28.2.2 Film Structure 670
28.2.3 Morphology 676
28.2.4 Raman Spectroscopy 678
28.3 Properties 680
28.3.1 Optical Properties 680
28.3.2 Optical Absorption 684
28.3.3 Electrical Properties 685
28.4 Concluding Remarks 699
References 702

29 **Thick Films**
Neil White 707
29.1 Thick Film Processing 708
29.1.1 Screen Printing 708
29.1.2 The Drying and Firing Process 709
29.2 Substrates 710
29.2.1 Alumina 710
29.2.2 Stainless Steel 710
29.2.3 Polymer Substrates 710
29.3 Thick Film Materials 711
29.3.1 Conductors 711
29.3.2 Resistors 712
29.3.3 Dielectrics 713
29.3.4 Polymer Thick Films 713
29.4 Components and Assembly 714
29.4.1 Passive Components 714
29.4.2 Active Components 714
29.4.3 Trimming 715
29.4.4 Wire Bonding 716
29.4.5 Soldering of Surface-Mounted Components 716
29.4.6 Packaging and Testing 717
29.5 Sensors 717
29.5.1 Mechanical 718
29.5.2 Thermal 719
29.5.3 Optical 719
29.5.4 Chemical 719
29.5.5 Magnetic 720
29.5.6 Actuators 720
References 720

Part D Materials for Optoelectronics and Photonics

30 **III–V Ternary and Quaternary Compounds**
Sadao Adachi 725
30.1 Introduction to III–V Ternary and Quaternary Compounds 726
30.2 Interpolation Scheme 726
30.3 Structural Parameters 727
30.3.1 Lattice Parameters and Lattice-Matching Conditions Between III–V Quaternaries and Binary Substrates 727
30.3.2 Molecular and Crystal Densities 727
30.4 Mechanical, Elastic and Lattice Vibronic Properties 729
30.4.1 Microhardness 729
30.4.2 Elastic Constants and Related Moduli 729
30.4.3 Long-Wavelength Phonons 729
30.5 Thermal Properties 731
30.5.1 Specific Heat and Debye Temperature 731
30.5.2 Thermal Expansion Coefficient 732
30.5.3 Thermal Conductivity 732
30.6 Energy Band Parameters 733
30.6.1 Bandgap Energy 733
30.6.2 Carrier Effective Mass 734
30.6.3 Deformation Potential 736
30.7 Optical Properties 737
30.7.1 The Reststrahlen Region 737
30.7.2 The Interband Transition Region 739
30.8 Carrier Transport Properties 739
References 741

31 **Group III Nitrides**
Romualdo A. Ferreyra, Congyong Zhu, Ali Teke, Hadis Morkoç 743
31.1 Crystal Structures of Nitrides 745
31.2 Lattice Parameters of Nitrides 746
31.3 Mechanical Properties of Nitrides 748
31.4 Thermal Properties of Nitrides 752
31.4.1 Thermal Expansion Coefficients 752
31.4.2 Thermal Conductivity 753
31.4.3 Specific Heat 756
31.5 Electrical Properties of Nitrides 757
31.5.1 Low Field Transport 758
31.5.2 High Field Transport 766
31.6 Optical Properties of Nitrides 769
31.6.1 Gallium Nitride 769
31.6.2 Aluminum Nitride 777
31.6.3 Indium Nitride 781
31.7 Properties of Nitride Alloys 782
31.8 Doped GaN 786
31.8.1 n-Type Doping 787
31.8.2 p-Type Doping 787

31.9 Defects in GaN 789
31.9.1 Points Defects 789
31.9.2 Extended Defects 792
31.10 GaN-Based Nanostructures 794
31.10.1 Quantum Wells 794
31.10.2 Quantum Dots 796
31.10.3 Vertical Cavities 796
31.10.4 Nitride Nanorods 799
31.11 Summary and Conclusions 801
References 802

32 **Electron Transport Within III-V Nitride Semiconductors**
Stephen K. O'Leary, Poppy Siddiqua, Walid A. Hadi, Brian E. Foutz, Michael S. Shur, Lester F. Eastman† 829
32.1 Electron Transport Within Semiconductors and the Monte Carlo Simulation Approach 830
32.1.1 The Boltzmann Transport Equation 831
32.1.2 Our Ensemble Semi-Classical Monte Carlo Simulation Approach 832
32.1.3 Parameter Selections for Bulk Wurtzite GaN, AlN, and InN 832
32.2 Steady-State and Transient Electron Transport Within Bulk Wurtzite GaN, AlN, and InN 834
32.2.1 Steady-State Electron Transport Within Bulk Wurtzite GaN 835
32.2.2 Steady-State Electron Transport: A Comparison of the III–V Nitride Semiconductors with GaAs 836
32.2.3 Influence of Temperature on the Electron Drift Velocities Within GaN and GaAs 836
32.2.4 Influence of Doping on the Electron Drift Velocities Within GaN and GaAs 839
32.2.5 Electron Transport in AlN 841
32.2.6 Electron Transport in InN 842
32.2.7 Transient Electron Transport 843
32.2.8 Electron Transport: Conclusions 845
32.3 Electron Transport Within III–V Nitride Semiconductors: A Review 845
32.3.1 Evolution of the Field 846
32.3.2 Developments in the 21st Century 848
32.3.3 Future Perspectives 849
32.4 Conclusions 850
References 850

33 **II-IV Semiconductors for Optoelectronics: CdS, CdSe, CdTe**
Minoru Isshiki, Jifeng Wang 853
33.1 Background 853
33.2 Solar Cells 853
33.2.1 Basic Description of Solar Cells 853
33.2.2 Design of Cd-Based Solar Cells 854
33.2.3 Development of CdS/CdTe Solar Cells 855
33.2.4 CdZnTe Solar Cells 858
33.2.5 The Future of Cd-Based Solar Cells 858

33.3 Radiation Detectors 858
33.3.1 Basic Description of Semiconductor Radiation Detectors .. 858
33.3.2 CdTe and CdZnTe Radiation Detectors 859
33.3.3 Performance of CdTe and CdZnTe Detectors 860
33.3.4 Applications of CdTe and CdZnTe Detectors 862
33.4 Conclusions 863
References 863

34 **II–VI Narrow Bandgap Semiconductors: Optoelectronics**
Ian M. Baker 867
34.1 Overview 867
34.1.1 Introduction to Solar and Infrared Imaging 868
34.1.2 Historical Perspective and Early Detectors 868
34.1.3 HgCdTe Infrared Detector Market in the Modern Era 869
34.2 Applications and Sensor Design 870
34.3 Photoconductive Detectors in HgCdTe 871
34.4 Sprite Detectors 872
34.5 Introduction to Photovoltaic Devices in HgCdTe 873
34.5.1 Ideal Photovoltaic Devices 874
34.5.2 Nonideal Behavior in HgCdTe Diodes 875
34.5.3 Influence of Silicon on Photodiode Array Performance ... 875
34.5.4 Theoretical Foundations for HgCdTe Array Technology 876
34.6 Manufacturing Technology for Photodiode Arrays 879
34.6.1 Liquid Phase Epitaxial (LPE)-Based Processes 879
34.6.2 Vapour Phase Epitaxial (VPE)-Based Processes 882
34.7 Advanced HgCdTe Technologies 886
34.7.1 Small Pixel Technology 886
34.7.2 Higher Operating Temperature (HOT) Device Structures 886
34.7.3 Two-Color Array Technology 888
34.7.4 HgCdTe Electron Avalanche Photodiodes (e-APDs) 889
34.7.5 Multifunctional HgCdTe Detectors 891
34.7.6 Retina-Level Processing 892
34.7.7 Conclusions and Future Trends for Photovoltaic HgCdTe Arrays 892
34.8 Emission Devices in II–VI Semiconductors 892
34.9 Potential for Reduced Dimensionality in HgTe–CdTe 893
References 893

35 **Optoelectronic Devices and Materials**
Stephen J. Sweeney, Jayanta Mukherjee 897
35.1 Introduction to Optoelectronic Devices 898
35.1.1 Historical Perspective 898
35.2 Light-Emitting Diodes and Semiconductor Lasers 900
35.2.1 Carrier–Photon Interactions in Semiconductors 900
35.2.2 Direct- and Indirect-Gap Semiconductors 900
35.2.3 Emission and Absorption Rates and the Einstein Relations 901
35.2.4 Population Inversion 902
35.2.5 Gain in Semiconductors 902
35.2.6 Density of States 902

35.2.7 Optical Feedback in a Fabry–Perot Laser Cavity 905
35.2.8 Wave Guiding 906
35.2.9 Carrier Confinement 907
35.2.10 Current Confinement 908
35.2.11 Laser Threshold and Efficiency 909
35.2.12 Carrier Recombination Processes 910
35.2.13 Temperature Sensitivity and T_0 912
35.3 Single- and Multimode Lasers 913
35.3.1 DFB and Wavelength Tuneable Lasers 913
35.3.2 Vertical Cavity Surface-Emitting Lasers (VCSELs) 915
35.3.3 High Power Lasers 917
35.3.4 Quantum Cascade Lasers 920
35.3.5 Emerging Semiconductor Materials for Modern Semiconductor Lasers 921
35.4 Optical Amplifiers 922
35.4.1 An Introduction to Optical Amplification 922
35.4.2 Semiconductor Optical Amplifiers (SOAs) 922
35.5 Modulators 923
35.5.1 Modulator Theory 923
35.5.2 Polarization-Insensitive Modulators 924
35.5.3 High-Speed High-Power QCSE Modulators 925
35.5.4 The Electro-Optic Effect 926
35.6 Photodetectors 927
35.6.1 Photodetector Requirements 927
35.6.2 Photodetection Theory 927
35.6.3 Detectors with Internal Gain 928
35.6.4 Avalanche Photodetectors 928
35.7 Conclusions 929
References 930

36 **Liquid Crystals**
Geoffrey Luckhurst, David Dunmur 933
36.1 Introduction to Liquid Crystals 933
36.1.1 Calamitic Liquid Crystals 935
36.1.2 Chiral Liquid Crystals 937
36.1.3 Discotic Liquid Crystals 938
36.2 The Basic Physics of Liquid Crystals 940
36.2.1 Orientational Order 940
36.2.2 Director Alignment 940
36.2.3 Elasticity 941
36.2.4 Flexoelectricity 944
36.2.5 Viscosity 945
36.3 Liquid-Crystal Devices 947
36.3.1 A Model Liquid-Crystal Display: Electrically Controlled Birefringence (ECB) Mode 948
36.3.2 High-Volume Commercial Displays: The Twisted Nematic (TN) and Super-Twisted Nematic (STN) Displays 951
36.3.3 Complex LC Displays and Other Cell Configurations 952

36.4 Materials for Displays 956
36.4.1 Chemical Structure and Liquid- Crystal Phase Behaviour . 956
36.4.2 The Formulation of Liquid-Crystal Display Mixtures 957
36.4.3 Relationships Between Physical Properties and Chemical Structures of Mesogens 958
References 964

37 **Organic Photoconductors**
David S. Weiss, Martin Abkowitz† 967
37.1 Chester Carlson and Xerography 968
37.2 Operational Considerations and Critical Materials Properties 970
37.2.1 Dark Conductivity 970
37.2.2 Photodischarge–Charge Transport 971
37.2.3 Photogeneration 977
37.3 OPC Characterization 978
37.3.1 Dark Decay 979
37.3.2 Photosensitivity 979
37.3.3 Electrical-Only Cycling 980
37.4 OPC Architecture and Composition 980
37.4.1 OPC Architecture 980
37.4.2 Coating Technologies 981
37.4.3 Substrate and Conductive Layer 982
37.4.4 Smoothing Layer and Charge-Blocking Layer 983
37.4.5 Charge-Generation Layer (CGL) 983
37.4.6 Charge-Transport Layer (CTL) 988
37.4.7 Overcoat Layer 989
37.5 Photoreceptor Fabrication 990
37.6 Summary 990
References 992

38 **Luminescent Materials**
Andy Edgar 997
38.1 Luminescent Centres 999
38.1.1 Rare-Earth Ions 999
38.1.2 Transition-Metal Ions 1000
38.1.3 s^2 Ions 1001
38.1.4 Semiconductors 1001
38.2 Interaction with the Lattice 1001
38.3 Thermally Stimulated Luminescence 1003
38.4 Optically (Photo-)Stimulated Luminescence 1004
38.5 Experimental Techniques – Photoluminescence 1005
38.6 Applications 1006
38.6.1 Compact Fluorescent Lamps 1006
38.6.2 Phosphors for Solid State Lighting 1006
38.6.3 Long-Persistence Phosphors 1007
38.6.4 X-Ray Imaging 1008
38.6.5 Phosphors for Displays 1009
38.6.6 Scintillators 1010
38.7 Representative Phosphors 1011
References 1011

39 **Nano-Engineered Tunable Photonic Crystals**
Harry E. Ruda, Naomi Matsuura 1013
39.1 PC Overview 1014
39.1.1 Introduction to PCs 1014
39.1.2 Nanoengineering of PC Architectures 1015
39.1.3 Materials Selection for PCs 1016
39.2 Traditional Fabrication Methodologies for Static PCs 1017
39.2.1 2-D PC Structures 1017
39.2.2 3-D PC Structures 1022
39.3 Tunable PCs 1027
39.3.1 Tuning the PC Response by Changing the Refractive Index of the Constituent Materials 1027
39.3.2 Tuning PC Response by Altering the Physical Structure of the PC 1028
39.4 Summary and Conclusions 1030
References 1031

40 **Quantum Wells, Superlattices, and Band-Gap Engineering**
Mark Fox, Radu Ispasoiu 1037
40.1 Principles of Bandgap Engineering and Quantum Confinement 1038
40.1.1 Lattice Matching 1038
40.1.2 Quantum-Confined Structures 1039
40.2 Optoelectronic Properties of Quantum-Confined Structures 1040
40.2.1 Electronic States in Quantum Wells and Superlattices 1040
40.2.2 Interband Optical Transitions 1042
40.2.3 The Quantum-Confined Stark Effect 1044
40.2.4 Intersubband Transitions 1045
40.2.5 Vertical Transport 1046
40.2.6 Carrier Capture and Relaxation 1046
40.3 Emitters 1047
40.3.1 Interband Light-Emitting Diodes and Lasers 1047
40.3.2 Quantum Cascade Lasers 1049
40.4 Detectors 1050
40.4.1 Solar Cells 1050
40.4.2 Avalanche Photodiodes 1050
40.4.3 Infrared Detectors 1051
40.5 Modulators 1052
40.6 Quantum Dots and Nanowires 1053
40.7 Conclusions 1055
References 1055

41 **Glasses for Photonic Integration**
Ray DeCorby, Mehrdad Irannejad 1059
41.1 Main Attributes of Glasses as Photonic Materials 1061
41.1.1 The Glass Transition and Control of Viscosity 1062
41.1.2 Metastability 1064
41.1.3 Glass as Host Material 1068
41.2 Glasses for Integrated Optics 1068
41.2.1 Low-Index Glassy Films 1068
41.2.2 Medium-Index Glassy Films 1069
41.2.3 High-Index Glassy Films 1070

41.3 Laser Glasses for Integrated Light Sources 1071
41.3.1 Advantages of Glass-Based Light Sources 1071
41.3.2 Alternative Glass Hosts 1072
41.3.3 Progress Towards Integrated Light Sources in Glass 1075
41.3.4 Integration of Dissimilar Materials Using Laser Processing 1075
41.4 Summary 1077
References 1077

42 **Optical Nonlinearity in Photonic Glasses**
Keiji Tanaka 1081
42.1 Third-Order Nonlinearity in Homogeneous Glass 1083
42.1.1 Experimental 1083
42.1.2 Theoretical Treatment 1084
42.1.3 Stimulated Light Scattering, Supercontinuum Generation, and Other Applications 1087
42.2 Second-Order Nonlinearity in Poled Glass 1088
42.3 Particle-Embedded Systems 1089
42.4 Photoinduced Phenomena 1090
42.5 Summary 1091
References 1092

Part E Novel Materials and Selected Applications

43 **Solar Cells and Photovoltaics**
Stuart Irvine 1097
43.1 Figures of Merit for Solar Cells 1098
43.2 Crystalline Silicon 1100
43.3 Amorphous Silicon 1102
43.4 GaAs Solar Cells 1102
43.5 CdTe Thin-Film Solar Cells 1104
43.6 $CuInGaSe_2$ ($CIGS_2$) Thin-Film Solar Cells 1105
43.7 Excitonic PV 1107
43.8 Conclusions 1107
References 1108

44 **Disordered Semiconductors on Mechanically Flexible Substrates for Large-Area Electronics**
Peyman Servati, Arokia Nathan 1111
44.1 a-Si:H TFTs on Flexible Substrates 1112
44.2 Field-Effect Transport in Amorphous Films 1112
44.2.1 Localized and Extended States 1113
44.2.2 Density of States (DOS) 1114
44.2.3 Effective Carrier Mobility 1114
44.3 Electronic Transport Under Mechanical Stress 1118
44.3.1 Thin-Film Strain Gauges 1118
44.3.2 Strained Amorphous-Silicon Transistors 1120
References 1123

45 **Photoconductors for X-Ray Image Detectors**
M. Zahangir Kabir, Safa Kasap 1125
45.1 X-Ray Photoconductors 1127
45.1.1 Ideal Photoconductor Properties 1127
45.1.2 Potential Photoconductors 1128
45.1.3 Summary and the Future 1135
45.2 Dark Current Limitations 1136
45.2.1 Charge Carrier Depletion 1137
45.2.2 Steady-State Thermal Generation 1137
45.2.3 Carrier Injection 1138
45.3 Metrics of Detector Performance 1139
45.3.1 X-Ray Sensitivity 1140
45.3.2 Detective Quantum Efficiency 1141
45.3.3 Modulation Transfer Function (MTF) 1143
45.3.4 Image Lag and Ghosting 1144
45.4 Summary 1145
References 1145

46 **Phase-Change Memory Materials**
Alexander V. Kolobov, Junji Tominaga, Paul Fons 1149
46.1 Structure of Ge-Sb-Te Phase-Change Alloys 1150
46.1.1 Crystalline Phase 1150
46.1.2 Amorphous Phase 1152
46.2 Mechanism of the Phase-Change Process 1156
46.2.1 Structural Studies 1156
46.2.2 Photo-Assisted Amorphization of $Ge_2Sb_2Te_5$ 1156
46.3 Present Applications and Future Trends 1158
References 1160

47 **Carbon Nanotubes and Bucky Materials**
Mark Baxendale 1163
47.1 Carbon Nanotubes 1163
47.1.1 General 1163
47.1.2 Geometry 1165
47.1.3 Synthesis and Chemistry 1165
47.1.4 Electronic Structure and Transport 1166
47.1.5 Nanoelectronic Devices 1168
47.1.6 Networks 1168
47.1.7 Other Electronic Applications 1169
47.2 Bucky Materials 1170
References 1170

48 **Graphene**
Henry H. Radamson 1173
48.1 Graphene Synthesis 1174
48.2 Band Structure and Electronic Applications 1176
48.3 Characterization of Graphene Material 1178
48.3.1 AFM 1178
48.3.2 Transmission Electron Microscopy 1178
48.3.3 Raman Spectroscopy 1179
48.3.4 XRD Analysis 1180

48.4 Photonic Applications 1180
References 1182

49 **Magnetic Information-Storage Materials**
Charbel Tannous, R. Lawrence Comstock† 1185
49.1 Magnetic Recording Technology 1186
49.1.1 Magnetic Thin Films 1187
49.1.2 The Write Head 1189
49.1.3 Spin-Valve Read Head 1192
49.1.4 Longitudinal Recording Media (LMR) 1199
49.1.5 Perpendicular Magnetic Recording 1205
49.2 Magnetic Random-Access Memory 1215
49.2.1 Tunneling Magnetoresistive Heads 1218
49.3 Extraordinary Magnetoresistance (EMR) 1220
49.4 Summary 1220
References 1220

50 **High-Temperature Superconductors**
Rainer Wesche 1225
50.1 The Superconducting State 1228
50.1.1 Characteristic Properties of Superconductors 1228
50.1.2 Superconductor Electrodynamics 1229
50.1.3 Superconductivity: A Macroscopic Quantum Phenomenon 1230
50.1.4 Type II Superconductors 1232
50.2 Cuprate High-T_c Superconductors: An Overview 1234
50.2.1 Major Families of Cuprate Superconductors 1234
50.2.2 Generic Phase Diagram of Cuprate Superconductors 1235
50.2.3 Crystal Structures 1236
50.2.4 Critical Temperatures 1238
50.3 Physical Properties of Cuprate Superconductors 1239
50.3.1 Anisotropic Superconductors 1239
50.3.2 Irreversibility Line 1239
50.3.3 Limitations of the Transport Critical Current 1241
50.4 Superconducting Films 1244
50.5 The Special Case of MgB_2 1245
50.6 Iron-Based Superconductors 1248
50.7 Summary 1250
References 1252

51 **Molecular Electronics**
Michael C. Petty, Takashi Nagase, Hitoshi Suzuki, Hiroyoshi Naito 1257
51.1 Electrically Conductive Organic Compounds 1258
51.1.1 Orbitals and Chemical Bonding 1258
51.1.2 Band Theory 1259
51.1.3 Electrical Conductivity 1260
51.2 Materials 1261
51.3 Plastic Electronics 1263
51.3.1 Diodes and Transistors 1263
51.3.2 Organic Light-Emitting Structures 1264
51.3.3 Photovoltaic Devices 1265

Detailed Cont.

51.4 Molecular-Scale Electronics 1266
51.4.1 Moore's Laws 1266
51.4.2 Nanoscale Organic Films 1267
51.4.3 Patterning Technologies 1269
51.4.4 Molecular Device Architectures 1271
51.5 DNA Electronics 1274
51.6 Conclusions 1276
References 1276

52 **Organic Materials for Chemical Sensing**
Asim K. Ray 1281
52.1 Analyte Requirements 1282
52.2 Brief Review of Inorganic Materials 1283
52.2.1 Oxide Semiconductor Sensors 1283
52.2.2 Graphene Based Sensors 1285
52.3 Macrocylic Compounds for Sensing 1286
52.3.1 Preparation of Sensing Membranes 1286
52.3.2 Thin-Film Properties 1289
52.4 Sensing with Phthalocyanine and Porphyrin 1291
52.4.1 Amperometric Sensor 1291
52.4.2 Optical Sensors 1292
52.4.3 Detection of Volatile Organic Vapor Compounds 1295
52.4.4 Phthalocyanine for Biosensing 1296
52.5 Polymeric Materials 1297
52.5.1 Conducting Polymers 1298
52.5.2 Ion Sensing 1299
52.5.3 Examples of Other Polymeric Sensors 1300
52.6 Cavitand Molecules 1303
52.7 Concluding Remarks 1305
References 1305

53 **Packaging Materials**
Darrel Frear 1311
53.1 Package Applications 1312
53.2 The Materials Challenge of Electronic Packaging 1313
53.2.1 Materials Issues in High-Speed Digital Packaging 1313
53.2.2 RF Packaging Materials Issues 1314
53.3 Materials Coefficient of Thermal Expansion 1315
53.4 Wirebond Materials 1315
53.4.1 Wirebond for Digital Applications 1315
53.4.2 Wirebond for RF 1317
53.5 Solder Interconnects 1317
53.5.1 Flip Chip Interconnects 1319
53.5.2 Flip Chip for RF 1320
53.5.3 Pb-Free 1320
53.6 Substrates 1322
53.6.1 RF Substrate Materials 1322
53.7 Underfill and Encapsulants 1323
53.7.1 Underfill 1323
53.7.2 Encapsulation 1323

53.8 Electrically Conductive Adhesives (ECAs) 1324
53.8.1 Adhesive Polymers 1324
53.8.2 Metal Fillers 1325
53.8.3 Conduction Mechanisms 1325
53.8.4 Isotropic Versus Anisotropic Conduction 1326
53.8.5 Rework 1326
53.9 Thermal Issues 1326
53.9.1 Thermal Issues in Digital Packaging 1326
53.9.2 Thermal Issues in RF Packaging 1327
53.10 Summary 1327
References 1327

54 **Organic Solar Cells**
Masahiro Hiramoto, Yusuke Shinmura 1329
54.1 History 1330
54.2 Excitons 1331
54.3 Donor–Acceptor Sensitization 1332
54.4 Exciton Diffusion 1332
54.5 Blended Junctions 1333
54.6 Route Formation 1334
54.7 π–π Stacking 1335
54.8 HOMO–LUMO Gap 1335
54.9 Tandem Cells 1336
54.10 Conclusions 1337
References 1337

55 **Materials for Terahertz Engineering**
Roger Lewis 1339
55.1 Overview 1339
55.1.1 Bridging Electronics and Photonics 1340
55.1.2 Focus on Frequency 1340
55.1.3 Physical Phenomena at Terahertz Frequencies 1341
55.1.4 Terahertz Technology Timeline 1342
55.1.5 Entry into the Terahertz Literature 1342
55.2 Terahertz Sources 1343
55.2.1 Thermal 1343
55.2.2 Vacuum Electronic 1344
55.2.3 Solid-State Electronic 1344
55.2.4 Laser-Pumped 1345
55.2.5 Mechanical Excitation 1345
55.3 Terahertz Detectors 1345
55.3.1 Thermal 1345
55.3.2 Electro-Optic 1345
55.3.3 Electronic 1346
55.4 Terahertz Components 1346
55.4.1 Mirrors 1346
55.4.2 Lenses 1346
55.4.3 Polarizers 1346
55.4.4 Metamaterials 1346
55.5 Conclusion 1347
References 1347

56 **Metamaterials**
Vassili Fedotov 1351
56.1 Bulk Metamaterials 1351
56.1.1 Basic Principle of Metamaterial Operation 1352
56.1.2 The First Metamaterial 1353
56.1.3 Magnetic Metamaterials (Artificial Magnetism) 1353
56.1.4 Negative-Index Metamaterials (Left-Handed Media) 1355
56.1.5 Chiral Metamaterials 1358
56.1.6 Transformation Optics and Metamaterial Cloaks 1360
56.1.7 Hyperbolic Metamaterials 1362
56.2 Planar Metamaterials 1365
56.2.1 Frequency Selective Surfaces 1365
56.2.2 Magnetic Mirrors (High Impedance Surfaces) 1367
56.2.3 Perfect Metamaterial Absorbers 1368
56.2.4 Designer Metasurfaces and Flat Optics 1369
56.2.5 Polarization Manipulation with Planar Metamaterials 1370
References 1375

57 **Thermoelectric Materials**
Donald T. Morelli 1379
57.1 Overview of the Field 1380
57.2 Semiconductors as Thermoelectric Materials 1381
57.2.1 Electronic Properties of Charge Carriers in Semiconductors 1381
57.2.2 Semiconductors with Low Intrinsic Thermal Conductivity 1383
57.2.3 Semiconductor Solid Solutions 1384
57.3 New Concepts in Thermoelectric Materials Design 1385
57.3.1 Phonon-Glass Electron Crystal: Skutterudites 1385
57.3.2 Bulk Nanostructured Thermoelectrics: Pb-Chalcogenide Nanocomposites 1387
57.3.3 Crystals with Large Anharmonicity: Tetrahedrites 1388
57.4 Summary and Future Outlook 1389
References 1390

58 **Transparent Conductive Oxides**
Hideo Hosono, Kazushige Ueda 1391
58.1 Overview 1391
58.1.1 Electrical Conduction 1391
58.1.2 Transparency 1393
58.2 Materials Design for TCOs 1395
58.2.1 Electronic Structures of Oxides 1396
58.2.2 Materials Design for n-Type TCOs 1396
58.2.3 Materials Design for p-Type TCOs 1398
58.3 New Approach to Explore Candidate Materials: Materials Genome Approach 1402
References 1403

59 **Inorganic Perovskite Oxides**
Tatsumi Ishihara 1405
59.1 Typical Properties of Perovskite Oxides 1409
59.1.1 Dielectric Properties 1409

59.1.2 Electrical Conductivity and Superconductivity 1409
59.1.3 Catalytic Activity 1410
59.2 Photocatalytic Activity 1411
59.3 Application for Solid Oxide Fuel Cells (SOFCs) 1413
59.3.1 Cathode 1413
59.3.2 Anode 1414
59.3.3 Electrolyte 1415
59.3.4 Interconnector 1417
59.4 Oxygen Separating Membrane 1418
59.5 Summary 1419
References 1420

About the Authors 1421
Detailed Contents 1435
Glossary of Defining Terms 1461
Subject Index 1495

Glossary of Defining Terms

π Band
Equivalent of the valence band in an organic semiconductive material

π^* Band
Equivalent of the conduction band in an organic semiconductive material

π Bond
Chemical bond formed between the nuclei of atoms. There are two bonding regions above and below a nodal plane containing the nuclei. In ethylene, a π-bond is formed by the overlap of p orbitals on the carbon atoms

σ Bond
Bond formed along the internuclear axis, between the nuclei of the atoms. In ethylene, a σ bond is formed by the overlap of two carbon sp^2 hybrid orbitals

A

Abbe Number (ν_d)
A measure of the dispersion (wavelength dependence) of the refractive index n of transparent materials at visible wavelengths; $\nu_d = (n_d - 1)/(n_F - n_C)$, where $d = 587$ nm, $F = 486$ nm, and $C = 656$ nm. Typically ν_d takes values of 80–20, which decreases with increasing n_d from 1.5 to 1.9

Absorption Coefficient (α)
The absorption coefficient represents the logarithmic decrement of the incident light intensity per unit length in the material

Accelerated Crucible Rotation Technique (ACRT)
A means to stir liquids in sealed ampoules, leading to improvements in crystal properties

Acceptor
Impurity in a semiconductor or any other electroluminescent device capable of inducing hole conduction and accepting a valence-band electron to produce an acceptor energy level

Acceptor Concentration
Number of acceptors per cm^3. It refers to the total concentration N_a and the concentration of ionized acceptors N_a^-

Acceptor Level
A level with energy E_a that acquires an electron captured from the valence band to an orbit near the acceptor's atom

Accumulation
The condition in which, for an MIS structure, the applied voltage to the gate electrode V_G results in an increase in the density of majority carriers near the surface of the underlying semiconductor. The applied voltage is less than the flat band voltage, $V_G < V_{FB}$

Active Material
The main functional component of thick-film paste, usually in the form of a finely divided powder

Active Matrix Addressing
The technique used to write and change images on complex liquid-crystal displays. Each display consists of thousands of picture elements (pixels) arranged as a matrix of columns and rows. Each pixel has a transistor switch (thin-film transistor – TFT), which has been deposited onto one plate of the liquid-crystal display. A voltage is applied sequentially to the rows of pixels, and selected pixels within each row are switched on by data voltages on the columns. As the next row in the sequence is activated, the elements of the previous row are switched off from the data line by the TFT, but hold their charge as an electrical capacitor until the row is addressed next time in the sequence, after a full frame time. Thus the image is built up line by line, and individual pixels must hold their charge for the period of the full frame-address time

Air Mass (AM) Number
This is often abbreviated to AM followed by a number and characterizes the intensity and spectrum of light (modified by absorption in the atmosphere). The AM number is 0 for no air mass (i. e., for a satellite solar array), 1 for the sun directly overhead, and 1.5 for the sun at 45° to the horizon

All-Optical Switch
Optical inputs are switched by other optical signals. In many cases, the switching is produced by refractive-index changes. Comparing electronic devices such as transistors, in which an electrical input is controlled by an electrical current, we expect that the all-optical switch can work faster, being free from electrical noise. However, at present, in contrast to the μm scale of a transistor, an all-optical switch may have cm scales

Amorphous Metal oxides
Glassy alloys of a transition metal with oxygen, typical examples being TiO_2, Ta_2O_5, Nb_2O_5, and Y_2O_3. In bulk form, these materials are typically polycrystalline or crystalline ceramics. However, amorphous thin films can be deposited with relative ease and they have been widely used as high-index layers in optical filter design and as dielectric layers in the microelectronics industry

Amorphous Network
Network consisting of atoms distributed randomly with a short-range order, i. e., holding a regular coordination of atoms

Amorphous Semiconductors
Semiconducting materials having a random network without long-range order

Amorphous Silicon
Thin films of silicon with a structure that is noncrystalline but which have the advantage of a higher absorption coefficient compared to crystalline silicon

Amphiphilic
An organic molecule possessing both polar (hydrophilic) and non-polar (hydrophobic) chemical groups. The classic examples are fatty acids, which consist of a non-polar ($-CH_2-$) chain terminating in a polar $-$COOH acid group

Amphoteric Dopants
Dopants that may serve both as acceptors and donors

Amphotericity
This occurs when an impurity can act either as an acceptor or a donor. In GaAs, a Si atom can take up a site on either sublattice and can act as a shallow donor or shallow acceptor

Analytical Electron Microscopy
The application of transmission and/or scanning electron microscopy for the purpose of structural and/or chemical microanalysis

Anharmonicity
The deviation of lattice vibrations in solids from those of a simple harmonic oscillator, which lead to nonzero lattice thermal resistivity

Anion
The electronegative atomic component in a compound, e.g., As in GaAs, Se in ZnSe

Annealing
Low-temperature heat treatment, normally in the presence of Hg, with the aim of adjusting the electrical properties of materials. The process of heating and slowly cooling semiconductor materials to reverse lattice damage. As a metallurgical term, annealing describes the use of heat to reduce the internal energy of a crystal. This energy may be mechanical (strain) or may represent variations in concentration, which can be reduced by diffusion, etc.

Antiferromagnetically Coupled Media (AFC)
See *Synthetic Ferrimagnetic Media* (SFM)

Antisite Defect
Occurs in semiconductor compounds and is an anion (cation) replacing a cation (anion) on a regular cation (anion) lattice site, e.g., As_{Ga} (Ga_{As}). The anion As_{Ga} (cation Ga_{As}) anti-site is expected to be a donor (acceptor) in the III–Vs and similarly in the II–VIs

Apparent Band gap Narrowing
Definition of apparent band gap seen in bipolar transistors due to the combined effects of band gap narrowing due to heavy doping and the use of Fermi–Dirac statistics

Areal Density
Number of bits per unit area stored on a recording medium. In disk drives, it is the product of the bits/unit length times the number of tracks per unit length (usually inches; the areal density is frequently quoted in Gbit/in^2)

Arrhenius Relation
A general expression connecting the value of a parameter in a thermally activated process to the absolute temperature T. Thus, a diffusivity $D = D_0 \exp(-Q/(k_B T))$, where D_0, Q and k_B are the pre-exponential factor, the activation energy for diffusion, and Boltzmann's constant, respectively

Artificial Magnetism
A phenomenon characterized by the appearance of a strong magnetic response (most notably at optical frequencies) in a material composite engineered from inherently non-magnetic materials

Astigmatism
A defect in the performance of a lens arising from lens asymmetry about the optic axis, leading to loss of sharpness of an image at focus

A-Swirl
A network of dislocations loops. The origin is a stacking fault of Si interstitials. The size of the dislocation loops can be many μm

Asymmetric Transmission
A new fundamental phenomenon of electromagnetism, which manifests itself as a difference in the overall intensity of circularly polarized waves transmitted in the opposite directions. It resembles the non-reciprocity of the Faraday effect in magnetized media but requires no magnetic field for its observation, and is fully compliant with the Lorentz reciprocity principle. The effect is observed in planar chiral metamaterials and involves partial conversion of the incident wave into one of the opposite handedness, and it is the efficiency of this conversion that is asymmetric with respect to the direction of propagation

Attenuated Total Reflection (ATR)
Technique to measure infrared absorption spectra of thin films on a prism with enhanced sensitivity utilizing multiple reflections of an infrared beam in the prism

Auger Electron Spectroscopy (AES)
A method in which an electron spectrometer is used to measure the energy distribution of Auger electrons emitted from a surface

Auto-Compensation
Often associated with amphotericity, auto-compensation occurs when the incorporation of electrically active impurities of one type leads to the automatic incorporation of electrically active defects of the opposite type. As an example, the doping of melt-grown GaAs with donors always leads to inadvertent incorporation of acceptors at about one tenth of the donor concentration

B

Back End of Line (BEOL)
A term generally describing the final points in the IC fabrication process line, such as the interconnection and packaging fabrication steps. These process steps generally require lower thermal budgets

Background-Limited Performance (BLIP)
Describes the best signal-to-noise performance that can be achieved with a detector. In this condition the only noise stems from the random arrival rate of photons and is, therefore, at a theoretical minimum

Backscattered Electrons
Incident electrons that have interacted with atomic nuclei and scattered backwards with comparable incident energy. They may be used for sample imaging with contrast dependent on the local averaged atomic number

Baldereschi Concept
Mean-value k-points in the Brillouin zone are defined such that the value that any given periodic function of the wavevector assumes at these special k-points is an excellent approximation to the average value of the same function throughout the Brillouin zone

Ballistic-Electron-Emission Spectroscopy (BEEM) Technique
A method to determine the barrier heights of Schottky contacts by tunnel injection of almost monoenergetic electrons from a biased metal tip into metal–semiconductor contacts that pass the metal films as ballistic electrons and are collected by the semiconductors, provided their energy is high enough to overcome the interfacial barriers

Band Alignment
Type I: electrons and holes are confined within the same layer. Type II: electrons and holes are confined in different (adjacent) layers

Band Filling
An effect that takes place when a semiconductor is illuminated with light at a frequency within the absorbing region. Upon absorption of the incident photons, the electrons undergo a transition from an occupied band (VB) to the empty band (CB), thereby partially populating the empty band

Band Gap (BG) or Energy Gap
The forbidden energy gap between the valence band, normally filled with electrons, and the conduction band, normally empty of electrons. The band gap energy measures the energy difference between the top of the valence band and the bottom of the conduction band. The optical BG is the photon energy above which the semiconductor absorbs energy of incident EM radiation. It may be direct or indirect, depending on the type of electron transition from VB to CB the radiation induces; in Si, the optical BG is indirect with $E_g = 1.17\,\text{eV}$ at 4.2 K

Band Gap Engineering
The science of tailoring the semiconductor band gap to control the electrical and optical properties of the material

Band Offset
The valence and conduction-band offsets describe the corresponding band-edge discontinuities at intimate and abrupt semiconductor–semiconductor interfaces

Barrier Height
The barrier height of a Schottky contact is the energy distance right at the metal–semiconductor interface from the Fermi level to the majority-carrier band-edge, i. e., the conduction-band maximum in n-type and the valence-band maximum in p-type contacts

Beam Effective Pressure Ratio
In molecular beam epitaxy, this is the ratio between the partial pressures of the various components of the molecular source beams

Birefringent Crystals
Crystals such as calcite that are optically anisotropic, which leads to an incident light beam becoming separated into ordinary (o-) and extraordinary (e-) waves with orthogonal polarizations; incident light becomes doubly refracted because these two waves experience different refractive indices n_o and n_e

Blended Junction
A mixed junction that contains donor/acceptor molecular contacts acting as photocarrier generation sites

Blocking Temperature
Temperature where the magnetic exchange coupling between an antiferromagnet and a ferromagnet vanishes

Bolometer
A sensitive detector of terahertz radiation, in many circumstances the most sensitive. The resistance of a detector material, typically Si or Ge, is measured to evaluate the energy deposited. Performance improves on cooling, so operation is often at cryogenic temperatures, e.g., at the boiling point of helium, 4.2 K. The bolometer exhibits a relatively fast response (μs) for a thermal detector

Boltzmann Equation (BE)
Generally an integro-differential equation that governs the carriers' distribution function. At equilibrium the BE is automatically satisfied by a Fermi distribution. The generalized BE is a version revised to include quantum effects in the carrier's scattering and interaction with electromagnetic fields. Methods for the solution of the BE include the relaxation time and variational, displaced Maxwellian and Monte Carlo methods

Boule
There are many terms used for large crystals. These include boule, ingot, and crystal. The top of the crystal is called the seed end, the first to solidify, etc. The bottom is called the tail or tang end

Bragg's Law
Can be summarized by the defining equation $\lambda = 2d_{hkl}\sin\theta$, which must be satisfied for constructive interference to occur, corresponding to maxima in the positions of diffracted intensity

Bravais Lattice
A repetitive arrangement of points in space such that the environment of each point is identically similar to that of every other point. 14 such arrangements are used to describe crystal systems

Bremsstrahlung
Electromagnetic radiation produced by the rapid

deceleration of an electron in the vicinity of an atomic nucleus

Bridgman Growth
Technique developed in order to grow compounds where one component is volatile at the growth temperature. The growth is carried out by physically withdrawing the melt from its furnace, while the melt crystallizes on a seed. The volatile component is replenished from a reservoir in a separate furnace. The arrangement of the equipment can be horizontal or vertical, the former offering the possibility of viewing the top of the solid/liquid interface and giving visual feedback of the growth process

Bridgman Technique
Crystals are produced by progressively freezing a melt from one end, usually via crucible movement. A wide range of materials is made using this simple technique and its many variants

Brillouin Zone (BZ)
A primitive cell in the reciprocal-space lattice, which proves to be body-centered cubic (bcc) for Si. BZ provides the domain of wavevectors $\boldsymbol{k}$, which relate to the electron momentum in crystal $\boldsymbol{p} = \hbar\boldsymbol{k}$.

Broadband Sensitizer
Typically a species that is co-doped along with rare-earth ions into a glass host to increase the pumping efficiency or radiative efficiency of the rare-earth ions. Various sensitizers have been demonstrated, including silicon nanoclusters, silver ions, and other rare-earth ions (such as in the sensitization of erbium by ytterbium)

Brooks–Herring (BH) Formula
Celebrated equation for the carrier's momentum relaxation time due to scattering by ionized impurities

Bulk Modulus
Ratio of the compressive or tensile force applied to a substance per unit surface area to the change in volume of the substance per unit volume

C

Calamitic
Denotes the rod-like shape of the molecules forming the majority of liquid-crystal phases that, therefore, are known as calamitic liquid crystals

Capacitance Equivalent Thickness (CET)
Theoretical thickness of SiO_2 that would be required to achieve the same capacitance density as an alternative dielectric material. This is extracted directly from, for example, the accumulation region of a *C–V* curve and has no further correction

Capacitance–Voltage (*C–V*)
C–V measurements of a MOS capacitor can be used to determine the oxide thickness and the amount of electrically active defects (fixed charge, trapped charge, mobile charge, and interface trapped charge) and mobile contaminants in the oxide. It can also be used to determine the oxide thickness and the profile of dopants at the semiconductor surface. The *C–V* technique is a method to determine barrier heights of Schottky contacts from the variation of their capacitance as a function of applied voltage. It is a common characterization technique used in wafer fabrication facilities

Carrier Concentration
Number of current carriers per cm^3. It can be electron concentration n and/or hole concentration p. The intrinsic carrier concentration (n_i) is $n = p$ in a perfect crystal semiconductor that has not been doped (i. e., an intrinsic semiconductor). The value of n_i for Si is approximated to $1.08 \times 10^{10}\ cm^{-3}$ at 300 K

Carrier Mobility
Measure of the effect of applying an electric field to charge carriers (electrons or holes). It is the additional velocity acquired by the carriers (drift velocity) divided by the electric field

Carrier Range
Determines how far an injected carrier moves per unit field before becoming immobilized in a deep trap. Sometimes, it is calculated as the product of the drift mobility and lifetime

Carrier Velocity, Microscopic
Gradient of the specific band spectrum $E_s(\boldsymbol{k})$ with respect to $\boldsymbol{p}$

Cascaded Second-Order Materials
Materials that exhibit second-order nonlinear properties that act as an effective third-order nonlinear material

Cathodoluminescence
Luminescence stimulated by energetic electrons

Cation
Electropositive atomic component in a compound, e.g., Ga in GaAs, Cd in CdS

Cd-Based Compound Semiconductor
Semiconductor that contains the element Cd

Centro-Symmetric
A crystal structure in which a center of inversion exists is called centro-symmetric. There are 32 crystal classes, among which 11 are centro-symmetric

Cermets
Standard form of thick films requiring exposure to a high-temperature firing cycle, typically in the range 700−1000 °C

Chalcogenide Glasses
Glasses that are amorphous alloys containing S, Se, and/or Te. Typical examples include Se, GeS_2, $GeSe_2$, As_2S_3, As_2Se_3, and As_2Te_3. By intermixing these and other binary chalcogenide glasses, a wide variety of multicomponent glasses can be formed. Further, a wide range of nonstoichiometric compositions is possible. Several compositions have become standard industrial materials, including $Ge_{33}As_{12}Se_{55}$, and $Ge_{28}Sb_{12}Se_{60}$. Chalcogenide glasses are characterized by narrow band gaps and good transparency in the mid to far-infrared wavelength range

Chemical Vapor Deposition (CVD)
Technique for the growth of thin solid films on a crystalline substrate as the result of thermochemical vapor-phase reactions. This reaction occurs above

a solid surface, e.g., a single crystal diamond, which causes deposition onto that surface. In the CVD technique, thin films are obtained under equilibrium conditions. The gas-phase reaction is activated by temperature in order to create condensable species that lead to film growth. Different CVD techniques have been developed (see *Mercury-Sensitized Photo-CVD*, *Hot-Wire CVD*, *MOCVD*, *MPCVD*, and *PECDV*). All techniques used for the production of diamonds require a means of activating gas-phase carbon-containing precursor molecules, usually CH_4, diluted by hydrogen

Chiral Nematic
The liquid-crystal phase in which the director is twisted into a helical arrangement; the phase is, therefore, not superimposable on its mirror image and so exhibits macroscopic chirality. The chirality of the phase originates from that of the constituent molecules. The chiral nematic is sometimes referred to as a cholesteric phase

Chromatic Aberration
A blurring of resolution due to the differing focal lengths of a lens when acting on electrons of differing energy

Cleave and Stain
A rapid, although destructive, technique for assessing semiconductor structures. Particular chemical mixtures affect the surfaces of different semiconductor types in different ways, giving rise to effects that can be seen in a simple optical microscope. The sample is first cleaved to produce a clean surface, preferably at an angle that tends to magnify the scale of the structure, and then stained

Coefficient of Thermal Expansion (CTE)
Fractional increase in the length of a macroscopic material per unit temperature increase. The units of CTE are $1/\mathrm{K}$

Coercive Field
The electric field required to reduce the polarization in a ferroelectric material to zero in a fully saturated ferroelectric hysteresis loop

Coercive Squareness Parameter (S^*)
The slope of the major hysteresis curve of a ferromagnetic recording layer normalized to the ratio of the remanent magnetization to the coercive field $S^* = 1 - \left[M_\mathrm{r}/CH_\mathrm{c}\big/(\mathrm{d}M/\mathrm{d}H)_{H=-H_\mathrm{c}}\right]$

Coherence Length (ξ)
The shortest distance within which a considerable change of the Cooper-pair density is possible

Cohesive Energy
The energy required to dissociate the atoms of a solid into isolated atomic constituents

Columnar Phase
A liquid-crystal phase in which the constituent molecules, usually disc-like, are stacked into columns; these are arranged parallel to each other in either a hexagonal or rectangular array

Complementary Metal Oxide Semiconductor (CMOS)
An *MOS* device technology consisting of nMOS and pMOS transistor pairs

Complex
A bound state between two or more adjacent, like or unlike species, e.g., an antisite and a lattice vacancy at a nearest-neighbor site on the same sublattice such as $\mathrm{As_{Ga}V_{Ga}}$

Compositional Uniformity
A critical parameter in all ternary alloy systems that determines a material's usefulness in device applications

Compound Semiconductor
Semiconductor crystals composed of two or more atomic elements from different groups of the periodic chart

Compressive Strain
Type of strain obtained when a strained $\mathrm{Si_{1-x}Ge_x}$ layer is grown on a silicon substrate

Compton Scattering
Transfer of energy from a photon to an electron, leading to the scattering of a longer-wavelength photon

Concentration Quenching
The reduction in luminescence efficiency and luminescence lifetime of a laser glass when the rare-earth dopant concentration is high. Quenching is due to interactions between closely spaced rare-earth ions at high concentrations. These interactions create new pathways, other than the desired radiative decay, for the ions to relax to the ground state after they have been raised to a desired lasing level by pump energy

Conduction Band (CB)
Band of energies allowed for electrons in a semiconductor, which is empty in undoped semiconductors. There are many CBs, separated by band gaps. However, the term CB usually refers to the conduction band that has the lowest energies, or closest to the valence band (VB). It is separated from the VB by a band gap

Conductive Layer
The layer on which the charge-generation layer and charge-transport layer are coated. It is connected to the ground in the electrophotographic process

Conductivity
Product of the number of electrons or holes per unit volume, the electronic charge, and the mobility of the carrier $\sigma = ne\mu_\mathrm{e,h}$

Conductivity Mass
Used for electron or hole conductivity (mobility) calculations

Configuration Coordinate Model
Model of interaction between the electronic system of a luminescent center and the vibrational system of the surrounding atoms and ions

Converse Piezoelectric Effect
The generation of mechanical strain in a material in response to an applied electric field, where the strain is linearly proportional to the applied field

Conversion Efficiency
Ratio of incident photon energy and output electricity energy

Cooper Pairs
Bound electron pairs formed of electrons with opposite spins and momenta (in their ground state with zero electric field)

Corona Discharge
When a high voltage of 5−10 kV is applied to sharp edges or fine wires of metals, the gas such as air surrounding the metal is ionized and discharged. If the voltage is DC, ions having the same polarity as the DC voltage move toward a grounded electrode

Correlation Factor (f)
The jumps of a diffusing particle are generally not random but correlated, i. e., successive jumps are not independent of each other; f (always ≤ 1) is the ratio of the jump rate of correlated jumps to the jump rate if the jumps were random

Cotterell Atmosphere
The stress field around a dislocation can be reduced if defects congregate around it. This results in defect atmospheres, called Cotterell atmospheres, which are commonly observed even when the dislocation itself would not provide an observable signature. EL2 defects aggregate in this way in SI substrate material

Coupled Granular Continuous (CGC) Recording Media
In perpendicular recording media: a recording surface consisting of two layers, the first being granular and the second consisting of multilayers, typically of cobalt and palladium

Critical Field (B_c)
The magnetic field sufficient to destroy the superconducting state in a type I superconductor

Critical Fluctuations
Variations of the superconducting order parameter $\psi = n_C^{1/2} \exp(i\varphi)$ (where n_C is the Cooper-pair density) around the thermal equilibrium values

Critical Thickness
Maximum thickness of a strained layer that can be grown on a substrate before relaxation of the layer occurs

Critical (Transition) Temperature (T_c)
The T_c is a characteristic property of the superconductor in question. At T_c the resistance drops abruptly to an unmeasurably small value

Cryopanel
Liquid-nitrogen-cooled surface within the growth chamber arranged to minimize contamination of the wafer surface by impurities

Crystal-Originated Particle (COP)
A small aggregate or cluster of vacancies. Usually, it is of octahedral shape with (111) facets, but in the case of nitrogen doping, it can also adopt the form of a platelet or a rod-like cluster. The usual size is 70−200 nm

Current in the Plane (CIP)
Magnetoresistive device with the sense current flowing in the plane of the films

Current Perpendicular to the Plane (CPP)
Magnetoresistive device with the sense current flowing perpendicular to the plane of the films

Czochralski and Liquid-Encapsulated Czochralski (LEC) Growth
Czochralski growth pulls a crystal from the melt. The seed is held in a holder vertically above the melt and rotated. After partial immersion into the melt, it is slowly withdrawn and the crystal forms on the seed. The growth takes place because of the increased heat loss from the crystal compared to the melt. Rotation of the seed and growing crystal leads naturally to a near-cylindrical crystal from which circular substrates can be sawn. The GaAs melt must be protected against As loss if this method is to be used. Boric oxide (see *Liquid Encapsulation*) liquifies on the surface of the melt and, to some extent, on the crystal, and acts as a barrier against As loss. The boric oxide also insulates the melt from the crucible; reducing the rates at which impurities from the crucible are introduced into the melt and aiding withdrawal of the exhausted melt from the crucible after growth

D

Dangling Bonds (Broken Bonds)
Point-like defects that are produced by breaking covalent bonds. The dangling bond may have an unpaired electron, two paired electrons, or no electrons. These dangling bonds were denoted as D^0, D^-, and D^+ by Mott; the superscript represents the charge state

Dark Decay
Rate of decrease of the surface potential of a charged organic photoconductor

Data Rate
Rate at which the digital data, recorded or read back in a digital storage device, is transferred. Characterized by Mbyte/s or MB/s, where a byte (B) is 8 bits

Debye Length
A measure of the distance beyond which the electric field due to a point charge (e.g., an ionized dopant atom) is screened by the free carriers (electrons or holes) and becomes increasingly negligible

Deep-Energy-Level Impurities
Doping impurities or other impurities whose energy level lies toward the center of the band gap

Deep-Level Transient Spectroscopy (DLTS)
DLTS is a useful technique for measuring concentrations of deep levels in conducting semiconductors. Deep levels within the depletion layer of a pn junction are filled by putting the junction into forward bias. On applying a reverse bias, these levels are allowed to empty thermally. This produces a change in junction capacitance consisting of a series

of exponentials. From the temperature spectra generated for different time windows of the transient, Arrhenius plots can be made to determine the energy level (eV) and capture the cross-section (cm^2) of the defects

Defect Chemistry
Representation of interactions between defect species and free carriers (dopant, impurity, native defect, electrons, and holes) in terms of chemical-style equations from which the concentrations of the various species concerned can be obtained

Deformation Potential
Effective electric potential describing interaction of carriers with the lattice deformation irrespective of what is the cause of the deformation. It describes the phenomenology of the interaction between carriers and phonons of deformation types

Deformation Potential Parameters
There are two and four such parameters for electrons and holes in Si, respectively. For holes, the parameter set includes one describing interaction with anti-phase deformation of atoms in the nonequivalent primitive cells' positions

Density-Functional Theory (DFT)
DFT is a quantum-mechanical approach to many-body electronic structure calculations of molecular and condensed-matter systems. The many-electron wavefunction is written in terms of the electron density. The major problem with DFT is that the exact functionals for exchange and correlation are not known. The widespread approximations are the local-density approximation (LDA) and GW approximation (GWA). The LDA assumes that the functionals for exchange and correlation depend only on the density at the coordinate where the functional is evaluated. LDA DFT calculations dramatically underestimate the gaps of sp^3-bonded semiconductors and insulators. In GW approximation, the exchange and correlation is expressed as the product of a single-particle propagator G and a screened interaction W. GWA DFT calculations yield the gaps of sp^3-bonded semiconductors and insulators generally to within a few tenths of an eV

Density of States (DOS)
The density of states is the number of energy states per unit energy interval at the energy E. It stands for the density distribution of allowed electronic energies in a material and is widely used with respect to this distribution across the band gap of disordered semiconductors

Density-of-States Mass
Used for electron or hole density-of-states calculations

Depletion
The condition in which, for an MIS structure, the applied voltage to the gate V_G electrode results in a reduction of the majority carriers near the surface of the underlying semiconductor. This region is referred to as the depletion layer. Charge remaining in this near-surface region is due to ionized dopants present in the channel region. This condition occurs for applied voltages between the flat band and threshold voltages, i. e., $V_{FB} < V_G < V_{thr}$

Depth of Field
The distance along an optic axis that an object can be moved without noticeably reducing the resolution

Depth of Focus
Maximum spacing between an imaging screen and a photographic plate (or CCD) that allows a recorded image to be retained in focus

Depth Profiling
Monitoring of signal intensity as a function of a variable that can be related to distance normal to the surface, cf., the compositional depth profile. Note: the signal intensity is usually measured as a function of sputtering time

Depth Resolution
Depth range over which a signal changes by a specified quantity when reconstructing the profile of an ideally sharp interface between two media or a delta layer in one medium

Detection Efficiency
Percentage of radiation incident on a detector system that is actually detected

Detective Quantum Efficiency (DQE)
The ratio of the square of the signal-to-noise ratio at the output of the detector to that at the input. The relative increase in image noise due to an imaging system as a function of the spatial frequency, f, is expressed quantitatively as DQE(f), which represents the signal-to-noise transfer efficiency for different frequencies of information in an image

Detectivity
A signal-to-noise parameter, normalized to area and bandwidth, often used for photoconductive detectors or single-element photovoltaic detectors

Devitrification
The transition of a glassy material to its lower-energy crystalline state. This process is usually driven by thermal energy, such as if the material is held at some characteristic temperature above its glass-transition temperature. The difference between the crystallization temperature and the glass-transition temperature for a particular glass is one measure of its stability

Diamond Structure
A structure in which each atom lies at the center of a tetrahedron surrounded by four nearest neighbors located at the points of the tetrahedron. In such a structure diffusion is isotropic

Dielectric Constant/Susceptibility (DC/DS)
Basic material optical tensor property that linearly connects the electric displacement to the electric-field vector inside the material. For Si it is scalar. At optical frequency DC/DS is a complex quantity

Dielectric Materials
A class of materials that are insulators or nonconductors where charge imparted to one part of the material is not communicated to any other part

Diffraction
The process of scattering and reconstruction of radiation in specific directions as a consequence of interaction with a periodic structure, e.g., light interacting with a grating or X-rays interacting with a crystal lattice

Diffusion Coefficient or Diffusivity
The macroscopic parameter that characterizes the jump rate or jump frequency of a species at the atomic level. It is normally obtained by matching experimental profiles to solutions of the appropriate diffusion equation, e.g., Fick's second law. The diffusion coefficient measurement techniques include Haynes–Shockley, time-of-flight, and noise-measurement-based methods

Diffusion Length
Measure of the spatial extent of a diffusion region. Typically, the concentration of a diffusant can be expected to fall by more than a factor of 10^3 over three diffusion lengths

Digital Printing
Technologies in which the image is comprised of a pattern of pixels. In electrophotographic digital printing the electrostatic latent image is written on a charged photoreceptor using a computer-controlled laser or LED bar

Dipping
A *Liquid-phase epitaxy* technique in which the substrate is lowered into the melt in a vertical orientation

Dirac Fermion
Like electrons in graphene. These electrons behave massless or as relativistic particles like photons and they obey the Dirac equation

Direct Piezoelectric Effect
The generation of an electric polarization change or a charge separation in a material in response to an applied stress, where the polarization change is linearly proportional to the applied stress

Director
Symmetry axis for properties such as the refractive index or dielectric tensor of a liquid-crystal phase; at a molecular level it is commonly associated with the preferred orientation of the unique axis of the constituent molecules, either rod like or disc like

Discotic
Indicates the disc-like shape of the anisotropic molecules that constitute a class of liquid-crystal phases known as discotic liquid crystals

Disorder Model
Derived from Monte Carlo simulation studies, this model describes electronic transport in random media in terms of disorder-induced fluctuations of both hopping site energy and relative orientation

Distribution Function
A function of $\boldsymbol{p}$ (impulse) and the carrier's coordinate, which obeys *Boltzmann's equation* and presents a semiclassical probability description of a carrier's statistics and dynamics. At equilibrium, it converts to the Fermi function over the carrier's energy, which in turn reduces to a Maxwell–Boltzmann distribution for lightly doped semiconductors

Donor
An impurity in a material that is capable of inducing electrical conduction in that material by transferring an electron to the conduction band. A donor's level is the energy E_d of an electron in an orbit near the donor atom. The donor concentration is the number of donors per cm^3, referring to either the total concentration N_d or the concentration of ionized donors N_d^+

Donor-Acceptor Sensitization
Carrier generation via charge transfer (CT) exciton

Dopant
An atom from a different group of the periodic table from the host atom that substitutes for it. It can be a donor, acceptor, or amphoteric and is deliberately incorporated to give n- or p-type conductivity

Doped Semiconductor
A semiconductor that contains donors and/or acceptors. *Lightly* and *moderately-doped* semiconductors contain impurities with energy levels that are well separated from the CB and VB. *Heavily-doped* semiconductors (n^+- or p^+-type) contain so many dopants that their energy levels merge with CB or/and VB, which leads to metallic conduction at low temperatures. For n-Si, the critical donor concentration of this, Mott's insulator–metal transition, proves to be about 3×10^{18} cm^{-3}. Fermi–Dirac statistics must be used for carriers in n^+- or p^+-type semiconductors, but in n^+- or p^+-Si, full degeneracy of carriers is met only well below room temperature

Double-Negative Medium
A material composite or medium that exhibits simultaneously negative dielectric permittivity and magnetic permeability

Drift Mobility
Velocity per unit of applied field imparted to a charge carrier by the electric field. In the presence of carrier traps it is reduced from the free-carrier velocity by the fraction of time a carrier spends in the trapping centers

Drude Formula
Phenomenological equation for the free-carrier contribution to the *dielectric constant/susceptibility*. It contains two parameters: the effective plasma frequency and the relaxation time

Drum Photoreceptors
Devices where the layers are coated sequentially on a metal drum substrate

Dual-Layer Organic Photoconductor
A photoconductor with an architecture where the charge-generation and charge-transport functions are carried out in separate layers

Dynamic Coercivity
The change of coercivity of small ferromagnetic particles with switching time under thermal excitation

E

E Layer
Thin ferromagnetic cobalt alloy film added to one or two sides of the Ru layer in a synthetic ferrimagnetic media to increase the exchange coupling

E′ Center
A kind of unpaired-electron dangling bond in oxide glasses. In SiO_2, an Si atom that is bonded to three O atoms may have one E′ center. The center, which may be produced by radiation, gives ESR signals and optical absorption at $\hbar\omega \approx 6\,\mathrm{eV}$

Easy Axis
Direction in a ferromagnetic sample along which the magnetization is oriented in the absence of an external magnetic field. In thin films, the easy axis is the direction in a substrate surface along which the director tends to align; it is determined by the nature of the surface treatment

Edge-Defined Film-Growth Technique
Shaped crystals, including tubes, sheets, etc., are grown through a die placed on the melt surface

Edge-on Orientation
Parallel orientation of the π–π stack to the substrate

Effective Mass
Generally a set of parameters describing the dynamics of the current carriers, which may deviate drastically from the free-electron mass. For Si, it includes the effective mass tensor of electrons in the CB (m_t, m_t, m_l) and the masses of heavy (m_1), light (m_2) of spin orbital split-off (m_3) holes in the valence band. Generally, $m_{1,2}$ depend on the direction of hole's momentum. Different averages over this direction define masses that enter the conductivity and the density of states

Einstein Relation
An equation connecting the mobility μ, (speed per unit electric field) of a charged particle to its diffusivity D. Specifically $\mu = qD/(k_\mathrm{B}T)$, where q is the electric charge on the particle. This expression is only valid for nondegenerate material

Elastic Compliance Constant (*C*)
Defined from Hooke's law by $X = Ce$, where X is the stress and e is the strain. It has units of pressure

Elastic Deformation
Deformation of a body in which the applied stress is small enough that the object retains its original dimensions once the stress is released

Elastic Stiffness Constant (*S*)
Defined from Hooke's law by $e = SX$, where e is the strain and X is the stress. It has units of inverse pressure

Elasticity
A property of liquid crystals that causes the directors to be uniformly aligned parallel to each other. Deviations from this uniform ground state require the addition of elastic energy to the liquid crystal

Electrically Conductive Adhesives (ECA)
Composite materials consisting of a dielectric curable polymer and metallic conductive particles. The polymer is an adhesive material that chemically reacts with metals and other polymers to form a bond. The metallic particles in the ECA form a network in the cured joint, which forms an electrical conduction path

Electroabsorption
Generic term for all effects of changing the absorption coefficient upon applying a strong electric field. For Si, the dominant is the change of the free carriers' absorption in IR

Electroabsorption Modulator (EAM)
Light transmittance modulator based on the *Quantum-Confined Stark Effect* (QCSE)

Electron
Elementary particle having a negative charge of 1.602×10^{-19} C and rest mass m_0 equal to 9.109×10^{-31} kg

Electron Affinity
The energy distance from the conduction-band minimum to the vacuum level at the semiconductor surface

Electron-Beam Lithography
Method for micro and nanoscale fabrication where a pattern on a polymeric layer (resist) is exposed to electron irradiation

Electron Level Number 2 (EL2)
The deep levels in most semiconductors were labeled according to their observation during DLTS measurements. EL levels are donors (they are neutral when they have their electron). HL levels are acceptors (they are neutral when they have their hole). EL2 is the most important deep donor level (actually, a double donor) and is either the As anti-site defect or contains the As anti-site as a component

Electron–Phonon Scattering
In Si, in addition to scattering by deformation phonons, scattering may also occur with short-wavelength (inter-valley) phonons

Electronic Wavefunctions
Solutions (or eigenfunctions) of the Schrödinger equation which present the electron's position in an atom. The electronic wavefunctions have to be infinite, continuous, and single-valued and their square modulus presents the probability distribution of an electron in the atom

Electro-Optic Effect
A change in refractive index upon application of a strong electric field. This is linked to electro-absorption through the Kramers–Kronig relation. It manifests itself as a change in *birefringence* in response to an applied electric field

Electrophotography
Printing technology in which charged marking particles are developed on an organic photoreceptor with an image pattern of surface charge, the electrostatic latent image, and the subsequent transfer of these particles to a receiver

Electroplating
Process used to deposit a material on a conducting

substrate in a chemical bath containing metal ions using electrical current

Eley–Rideal Reaction
A surface-reaction scheme in which an arriving chemical species reacts on the surface without any surface diffusion process. It is considered the counterpart of a Langmuir–Hinshelwood reaction

Ellipsometry
Method of defining the optical constants by measuring the ratio of reflectance for s- and p-polarized radiation, and the relative phase shift between the two, at large incidence angles

Elliptical Dichroism
See *Asymmetric Transmission*

Emission Device
An emission device creates photons from extra electron–hole pairs through a process called radiative recombination. The electron–hole pairs are often created by forward-biasing a pn junction

Encapsulation
A dielectric polymer material that flows over and covers the electronic components in a package to provide mechanical and electrical protection to the packaged devices

Energy-Band Spectrum
Dispersion relation $E_s(\boldsymbol{k})$ of allowed electron energy to $\boldsymbol{k}$ within the *Brillouin zone*. As a function of small deviation of $\boldsymbol{k}$ from the extrema points of $E_s(\boldsymbol{k})$ it proves to be anisotropic parabolic for the conduction band and non-parabolic for valence band, forming ellipsoid valleys and warped spheres, respectively, in the *Brillouin zone*

Energy Gap
See *Band gap*

Epitaxial Layer (or Epilayer)
Thin crystalline layer on a single-crystal substrate with orientation and lattice structure determined by the substrate crystallography

Epitaxy
(Greek; arrange upon). This is the growth of one crystal (the epitaxial layer) on the surface of another (the substrate) and where the orientation of the grown layer is the same as the substrate. If the layer is of the same material as the substrate, the growth is homoepitaxial. If the layer and substrate are different, the growth is heteroepitaxial. Usually, epitaxy can only be performed where there is a close match between the lattice constants, although it is possible for a layer of quite different lattice constants to be grown if there is a change in crystal orientation between the layer and the substrate. The term epitaxy was introduced to describe the importance of having parallelism between two lattice planes with similar networks of closely similar spacing

Equivalent Oxide Thickness (EOT, t_{eq})
The theoretical thickness of SiO_2 that would be required to achieve the same capacitance density as an alternative dielectric material. This is often determined by quantum-mechanical modeling of the *C–V* curve for thin dielectric layers (where tunneling effects result in substantial *leakage* current) and/or where poly-Si gate electrodes (resulting in depletion effects) are present

Etch Pit Density (EPD)
The usual way in which substrate suppliers quote dislocation densities. EPD is found by etching the surface in molten alkalis, like KOH, and viewing the resulting pitted surface under an optical microscope. Each pit represents a single dislocation ending at the surface

Ewald Sphere
Geometrical construction used to illustrate the relationship between a diffraction pattern and the reciprocal lattice of a diffracting crystal

Exchange Field (H_{ua})
Parameter characterizing the strength of the coupling of a ferromagnetic to an antiferromagnetic film. Measured in Oe or A/m

Exciton
An electron–hole pair, which behaves like a H atom, with a binding energy of 10−100 meV and lifetime of ≈ 1 ns. The wavefunction, in principle, extends over the whole crystal

Exhaustion Concentration
$N = |N_d - N_a|$ is the maximum attainable carrier concentration in a doped semiconductor

Extrinsic Optical Activity
A rare instance of the optical activity phenomenon. It occurs in the presence of chirality, which is drawn extrinsically, i.e., from the experimental arrangement that includes not only molecules (or metamolecules) but also the wave propagation direction. The molecules (metamolecules) must be oriented and can be structurally achiral

F

Face-On Orientation
Perpendicular orientation of the π–π stack to the substrate

Fermi–Dirac Distribution
A distribution function that describes the number of particles over energy states in a material. The function is written as: $\bar{n}_i = 1/\left(e^{(\varepsilon_i-\mu)/(k_B T)} + 1\right)$, where k_B is Boltzmann's constant, T is the temperature, ε_i is the energy at state i, and μ is the chemical potential

Fermi Level or Fermi Energy
For a metal, the Fermi level is defined as that energy below which its one-electron levels are occupied and above which they are empty in the ground state, i. e., at zero temperature. In the context of semiconductors, the term Fermi level is a synonym for their electrochemical potential

Ferroelastic
A material that exhibits a phase transition from a non-strained to a strained state, generating a spontaneous strain at the transition, and in which the spontaneous strain can be switched between two or more stable states by the application of a stress

Ferroelectric
A polar dielectric in which the polarization can be switched between two or more stable states by the application of an electric field

Ferroelectric Domains
Adjacent regions in a ferroelectric crystal that have their spontaneous polarization vectors inclined to one another

Ferroelectric Hysteresis
The loop that is produced when the polarization in a ferroelectric material is plotted as a function of applied electric field

Ferroelectric Relaxor
A ferroelectric that shows a broad peak in relative permittivity at the paraelectric-to-ferroelectric phase transition and in which the temperature of the peak is strongly dependent upon measuring frequency

Fick's First Law of Diffusion
For diffusion parallel to the x-axis the diffusant flux (atoms per unit area per second) is equal to $-D\partial c/\partial x$, where c is the diffusant concentration that defines the diffusivity D

Fick's Second Law of Diffusion
Based on the first law, this law gives the rate of build up of the diffusant concentration at a given depth as $\partial c/\partial t = \partial(D\partial c/\partial x)/\partial x$

Field-Effect Transistor (FET)
A transistor where the current between two electrodes (the drain and source) is modulated by the electric field from a third electrode (the gate)

Field Emission
Electron emission from a metal or semiconductor into vacuum under the influence of a strong electric field

Figure-of-Merit (FOM)
Measure of the performance of an integrated circuit technology. The time delay associated with signal propagation is a common metric

Fill Factor (FF)
Measure of the maximum power that can be obtained from a photovoltaic solar cell compared with the product $I_{sc} \cdot V_{oc}$

Firing
One of the key stages of the thick-film production process. It is usually undertaken in a continuous-belt furnace at temperatures of up to 1000 °C

Flat band Voltage (V_{FB})
The voltage applied across an MIS device at which there exists no charge in the semiconductor. As a result, the valence and conduction band structure of the semiconductor are flat. This condition occurs when this voltage equals the work function difference between the metal electrode and the semiconductor under ideal conditions. The presence of charges in the insulator or at the interface, due to defects, modifies the voltage required to achieve the flat band condition

Flexoelectricity
Generation of a macroscopic electrical polarization in a nematic liquid crystal when the director distribution is deformed from its uniform state of alignment. The inverse of this effect is the deformation of the director distribution when an electric field is applied to a nematic; the magnitude of the deformation is linear in the field

Flip Chip
Integrated-circuit-level interconnect that can be used to replace wirebond interconnects. The flip chip is a solder bump on an area array on the chip surface that routes the power, ground, and signals from the integrated circuit to the bumps. The metallization on the IC surface is called the under-bump metallurgy and forms a metallurgical bond with the solder. The flip-chip die is joined to the package by placing the die face down on the matching bond pads on a substrate and reflowing the solder to form an electrical, thermal, and mechanical interconnect

Floating Gate Current
Very small gate currents ($\approx$ fA or less) can be measured using the floating gate technique in which the drain current of a MOS transistor is measured after the gate bias has been removed. Then, by using the decay in the drain current as a function of time and the measured drain current versus gate voltage characteristics and the oxide capacitance, the gate current can be calculated

Floating Zone (FZ) Technique
This technique uses a solid feed rod that is melted at its lower end by a high-frequency coil. The melt flows through a central hole of the coil down to the growing crystal below the coil. The FZ technique does not need a crucible. Hence, the melt is not contaminated by other materials. The crystals that are grown according to the FZ technique are called FZ crystals

Fluence
The total, time-integrated, flux of particles (electrons, protons, ions, etc.) that reach a unit area of sample. It can be used to represent the total number of ions implanted into a surface and is sometimes called the dose

Fluorescence
Luminescence with a lifetime $\lesssim 10$ ns

Fluoride Glasses
Multicomponent glasses, typically based on fluorides of zirconium, barium, lead, gallium, lanthanum, aluminum, and sodium. They have a wide transparency range, from ultraviolet to mid-infrared wavelengths. They also have low characteristic phonon energies and can dissolve large concentrations of rare-earth ions. For these reasons, they are extremely popular as hosts for rare-earth-doped amplifiers and lasers operating in the UV/visable and mid-infrared regions

Flux Lines (Vortices)
Regions in which magnetic flux enters a type II superconductor in the mixed state. Screening currents flow around each of the flux lines. Due to the repulsive vortex–vortex interactions a hexagonal flux line lattice is formed

Flux Pinning
An effect caused by defects. The energy for the formation of the normal cores of the flux lines is reduced in regions with a reduced Cooper-pair density

Flying Height
Spacing between the bottom of the slider and the top of the recording medium in a disk drive

Focal Plane Array (FPA)
Focal plane array consisting of an HgCdTe sensor on a silicon ROIC

Forced Convection
Process of forced flow of fluid

Fourier-Transform Infrared (FTIR) Spectrometry
A means of performing infrared absorption measurements with great speed and precision, based on an optical interferometer

Four-Point Probe
A popular technique used to measure the resistivity (Ω cm) of a semiconductor sample. Four collinear probes are equally spaced, a current is applied through the outermost probes, and the voltage difference across the two inner probes is measured. From the voltage and current values, the resistivity can be calculated assuming that the contact and spreading resistances between the probes and the semiconductor and the resistance of the probes are negligible in comparison to the resistance of the semiconductor

Franck–Condon Principle
Absorption and emission of light takes place in such a short time that the atomic coordinates in ground and excited states are unchanged; i. e., transitions are vertical on a configuration coordinate model

Free Convection
The process of movement of fluid through a temperature gradient

Free Film
Ferromagnetic film in a magnetoresistive or spin-valve read head in which the magnetization can respond easily to an external magnetic field

Frenkel Exciton
Strongly bound electron–hole pairs

Friedel's Law
A law that states that the intensity of an *hkl* reflection in a diffraction pattern is equal to the intensity in the opposite $\overline{hkl}$ reflection

Front End of Line (FEOL)
A term generally describing the initial points in the integrated-circuit fabrication process line, such as the transistor fabrication steps. These processes generally require higher thermal budgets

G

Gate Oxide Integrity (GOI)
GOI measures the breakdown stability of the oxide layer of a MOS capacitor when a voltage is applied to the capacitor. There are different types of gate oxide integrity tests. Most of these are standardized. This may currently be the best method to detect the density of very small voids

Geminate Recombination
The primary recombination of a correlated (quasi-bound) hole–electron pair immediately following photoexcitation

Germanene
A single layer of sp^2 bonded Ge atoms in a honeycomb shape

Ghosting
The change of X-ray sensitivity of the X-ray image detector as a result of previous exposure to radiation. In the presence of ghosting, a shadow impression of a previously acquired image is visible in subsequent uniform exposures

Giant-Magnetoresistive (GMR) Effect
Change in the resistivity of a stack of ferromagnetic films coupled by thin non-ferromagnetic films when the orientation of the magnetization of the films adjacent to the non-ferromagnetic film is varied. The change in resistivity with low temperatures and high magnetic fields was observed to be as large as 50%; hence the adjective *giant*

Glass Ceramics
Glass that contains crystalline particles or regions. Such materials may be transparent or smoggy

Glass Frit
Binding matrix within the thick film. This binds the active particles together and also bonds the thick film to the substrate

Glass-Transition Temperature
The approximate temperature at which a material changes from a supercooled liquid to an amorphous solid, or vice versa. The transition is marked by an abrupt but continuous change in slope of the specific volume and enthalpy versus temperature curves. Viscosity varies rapidly near the glass-transition temperature, which is also sometimes called the softening temperature

Glow Discharge Spectrometry (GDS)
A method in which a spectrometer is used to measure relevant intensities emitted from a glow discharge generated at a surface. This is a general term that encompasses glow discharge optical-emission spectrometry (GDOES) and glow discharge mass spectrometry (GDMS). GDOES is a method in which an optical-emission spectrometer is used to measure the wavelength and intensity of light emitted from a glow discharge generated at a surface. GDMS is a method in which a mass spectrometer is used to measure the mass-to-charge quotient and abundance of ions from a glow discharge generated at a surface

Gradient Freeze Growth
This technique is similar to the *Bridgman method* except that the melt is not physically removed from the furnace. Instead, the temperature gradient along the melt is controlled electrically so that solidification commences at the seed and progresses until the melt is

exhausted. Once again, both horizontal and vertical arrangements are possible, and often the reservoir is replaced by liquid encapsulation to impede As loss from the melt. This is now the favored technique for the growth of most GaAs substrate material

Grading
Grading is the gradual change from one semiconductor to another, exemplified by the gradual change from GaAs to GaAsP that is necessary in some LEDs

Grätzel Cell
Named after its inventor, Michael Grätzel, this is a dye-sensitized cell using a porous TiO_2 substrate to collect the photo-generated charge

GRINSCH Laser
A graded-refractive-index separate-confinement heterojunction laser (see *Separate Confinement Heterojunction (SCH) Laser*). Optical confinement is optimized by grading the composition of the cladding layers

Gunn Diode
(See *Negative Differential Resistance*). This is a device based on low-doped, n-type GaAs and relies on the NDR effect. It is used for the generation of low-power microwave currents

H

Hall Effect
The deflection of a charged particle moving in a magnetic field that is perpendicular to its motion. The deflection is due to the Lorentz force on the charged particle and it causes the charges to accumulate on one side of the sample. The voltage measured at right angles to the current flow is called the Hall voltage. The Hall effect can be used to characterize the mobility (cm^2/Vs), resistivity (Ω cm), type of carrier, and carrier density (cm^{-3}) of a semiconductor sample. The *Hall mobility* is the product of conductivity and the Hall constant (the transverse-electric-field Hall field divided by the product of the current density and the magnetic induction) of a conductor or semiconductor; a measure of the mobility of the electrons or holes in a semiconductor. The *Hall coefficient* and *factor* relate μ^{H} linearly to σ and μ, respectively

Hard Axis
Direction in a ferromagnetic sample at right angles to the easy axis

Hard Magnetic Bias Film (also Longitudinal Bias Film)
Permanent magnetic film abutted to the free film in a magnetoresistive or spin-valve head to eliminate magnetic domains

Head Field Slope Parameter (Q)
The maximum slope of the magnetic field of a write head in the direction of the magnetization of the recording layer normalized to the coercive field (H_c) divided by the magnetic spacing (d), $Q = \frac{\mathrm{d}H_x/\mathrm{d}x}{H_c/d}$

Heat Capacity
The amount of heat required to change the temperature of a substance temperature by 1 degree, with units of energy per degree

Heteroepitaxy
The growth of a layer of markedly different composition from the substrate, i. e., the epitaxial layer and the substrate are made from different materials, e.g., the diamond growth on an iridium single-crystal substrate. Another example is the growth of $Ga_xIn_{1-x}P_yAs_{1-y}$ layers lattice-matched to InP. The heteroepitaxial growth techniques are chemical-vapor deposition, liquid-phase epitaxy, and molecular-beam epitaxy

Heterojunction Bipolar Transistor (HBT)
A modification to the standard bipolar transistor where a heterojunction is used to control the carrier flow at the base

Heterojunction Laser
A semiconductor laser where both electrical and optical confinement exploit the properties of heterojunctions is a heterojunction laser. If the active region is sandwiched between two heterojunctions, the laser is termed a double heterojunction

Heterojunctions and Heterostructures
A heterojunction is the junction between different materials (e.g., GaAs and AlGaAs with different band gaps). Such a junction exhibits several properties that may be very useful to device manufacturers. These include changes in the energies of the valence or conduction bands, or both, and changes in the optical properties. A heterostructure is a semiconductor structure where the properties of heterojunctions are exploited

Heterostructures
Multilayer structures that have regions of different composition within them; used in advanced IR device types

Hexagonal As
The natural form of elemental arsenic

HgZnTe
An alternative Hg-based ternary system to MCT

Hierarchical Nanostructuring
The combination of introduced defects/impurities in a thermoelectric material with different characteristic length scales and meant to scatter phonons with a wide range of wavelengths

High Impedance Surface
A planar metamaterial, which acts as a ground plane that does not support propagating surface waves (see *Magnetic Mirror*.)

High Index Contrast
Waveguides or devices fabricated using two or more materials that have very different refractive indexes. High index contrast is the basis for the confinement of light to very small cross-sectional-area waveguides or very-small-volume optical cavities, either using total internal reflection or photonic band gap effects. High

index contrast is thus the basis for increased density of optical integrated circuits

High-Electron-Mobility Transistor (HEMT)
A modification of the *metal semiconductor field-effect transistor (MESFET)* where the separation of the electrons in the channel from the ionized donors is achieved by using a heterojunction. This results in higher electron mobility and thus greater speed and lower noise

Highest Occupied Molecular Orbital (HOMO)
The highest energy molecular orbital of an atom or molecule that contains an electron. If the atom or molecule were to lose an electron, it would most likely lose it from this orbital

High-Temperature Solution Growth
Slow cooling, or top-seeded (similar to Czochralski) growth of materials in a solution designed to reduce the freezing point of the desired phase below a critical temperature

Homeotropic Alignment
Alignment of the director perpendicular to the surface of a substrate

Homoepitaxy
Epitaxial growth of a layer of the same composition as the substrate

Homojunction
Junction between materials of the same chemical type but of different electrical properties. For example, a p-n junction in GaAs is a homojunction

Hopping Conduction
Electronic transport that localized electrons hop from site to site with the assistance of phonons

Hot-Carrier Phenomena
Group of effects associated with the carrier transport in high electric fields. For Si, it includes the nonlinear relation between drift velocity and electric field with saturation (n-Si) and near saturation (p-Si), anisotropy of drift velocity regarding the electric orientation relative to the crystallographic axes (the Sasaki–Shibuya effect), and the diffusion's anisotropy as regards the density gradients along and perpendicular to the strong electric field

Hot-Wire Chemical Vapor Deposition (CVD)
The same technique as catalytic CVD, which is used for thin-film growth through decomposition of source gas materials utilizing catalytic reaction on the heated filament

Huygens's Principle
Concept used to explain the propagation of a wave (and diffraction processes), in that every point on a primary wavefront acts as a source of spherical wavelets, the envelope function of which acts to reconstruct the primary wavefront a short time later

Hybrid Circuits
Circuits consisting of electronic elements made from differing enabling technologies such as thick-film, thin-film, monolithic silicon, etc.

Hybridization
In molecules, the orbitals occupied by the electrons are seldom pure s or pure p orbitals. Instead, they are mixed or hybrids, formed by combining the s and p orbitals. Examples are sp (linear), sp^2 (trigonal and planar), and sp^3 (tetragonal)

Hydrothermal Growth
Growth that takes place in aqueous-based solutions at high temperatures and high pressures

Hyperbolic Metamaterial
An artificial material composite characterized by a hyperbolic dispersion of the supported (propagating) waves. Such dispersion is typically found in composites of extreme anisotropy, where the sign of dielectric permittivity is reversed for the directions parallel and orthogonal to the anisotropy axis

Hysteresis
The state of the magnetization in response to an external magnetic field depends on the initial state of the magnetization

I

***I/V* Technique (Current–Voltage Technique)**
The *I/V* technique is a method to determine barrier heights and ideality factors of Schottky contacts from their current–voltage characteristics

Ideality Factor
The ideality factor of Schottky contacts characterizes the variation of their barrier heights as a function of applied voltage

Image Lag
Lag is the carryover of image charge generated by previous X-ray exposures into subsequent image frames. The residual signal fractions following a pulsed X-ray irradiation are referred to as image lag

Impact Ionization
Process of generation of electron–hole pairs by carriers moving in an electric field higher than the breakdown threshold, which for Si is of the order of 3×10^5 V/cm

Improper (or Extrinsic) Ferroelectric
A ferroelectric material in which spontaneous polarization is not the primary order parameter. Frequently, the primary order parameter is the spontaneous strain associated with a ferroelastic phase transition

Impurity
A foreign atom unintentionally present in the semiconductor, incorporated either during growth or processing

Inelastic Scattering
The result of a collision between a photon or electron and the nuclei or electrons of a material, such that there is a net change in the internal energies of the system and in the sum of the kinetic energies before and after the collision

Infrared-Laser-Absorption Spectroscopy (IRLAS)
Technique to determine the density of gas-phase molecules and chemical species using infrared absorption spectra

Infrared-Reflection-Absorption Spectroscopy (IRRAS)
Technique to determine the surface-bonding

configuration on the film-growing surface using light absorption during reflection of an infrared beam

In-Situ Monitoring
Tools to help in understanding layer growth kinetics and to provide monitors suitable for feedback control in epitaxial growth systems

Integrated Circuit (IC)
Combination of active and passive circuit elements to enable computational logic or analog operations

Integrated Detector Cooler Assembly (IDCA)
A commonly used infrared detector scheme in which the detector is mounted directly on the cold finger of a cryocooler (often based on Stirling cycle engines). The detector and cold finger are then enclosed in a vacuum vessel with a transparent window and optical baffles

Integrated Optics/Photonics
The manufacture of photonic elements and circuits on a planar substrate, typically using thin-film deposition, lithography, and etching steps. Typically, the substrate is a glass or semiconductor wafer, and the photonic elements are guided-wave devices

Interconnect
The system of metal conducting lines and contacts among IC components. The categories include local (between neighboring devices), intermediate (between neighboring circuit elements), and global (across the IC chip)

Interdiffused Multilayer Process (IMP)
The process of obtaining a uniform alloy composition of $Cd_xHg_{1-x}Te$ by the growth of alternate layers of the binary compounds that are thin enough to completely interdiffuse within the time of growth

Interface Trap Density
The density (cm^{-3}) of positive or negative charges located at the silicon/silicon dioxide interface, due to defects induced by oxidation, structural defects, impurities, or other defects caused by bond-breaking mechanisms such as radiation or hot carriers. These states are typically in electrical communication with the charges in the channel of a MOSFET and results in stretch-out of the capacitance–voltage characteristics. They also affect the turn-on and turn-off characteristics of a MOS transistor

Interface-Induced Gap States
Because of the quantum-mechanical tunnel effect, the wavefunctions of electrons tail across semiconductor interfaces in energy regions where occupied states overlap a band gap. These evanescent waves are the continuum of the *intrinsic* interface-induced gap states

Intergrain Critical Current Density
Macroscopic transport critical current density, which can be much smaller than the intragrain j_c because of the weak-link behavior of large-angle grain boundaries in the cuprate superconductors

Intermetallic Compounds
Substances that form between pad metallizations and the active components of molten solder (typically Sn). For Cu metallization, Sn reacts to form Cu_3Sn and Cu_6Sn_5 intermetallics. For Ni, Sn reacts to form Ni_3Sn_4. These intermetallics typically form at the interface of the solder and the metallization and are usually more brittle than the solder or the metallization

Intermetallic Reactions And Phases
Two or more metals can react chemically to produce compounds. These are revealed in phase diagrams as labeled, vertical lines. The compound is often referred to as an intermetallic phase

Internal Photoemission Yield Spectroscopy (IPEYS) Technique
IPEYS is a method to determine barrier heights of Schottky contacts and band offsets of semiconductor heterostructures by photoinjection of hot electrons over the energy barriers at metal/semiconductor and semiconductor/semiconductor interfaces, respectively, as a function of the photon energy of the exciting light

Interstitial
A site lying between regular lattice sites that can be occupied by dopant, impurity, or host atoms. The latter case is known as a self-interstitial. It generally behaves as a donor

Intersubband (ISB) Transition
Transition between confined states within the conduction or valence bands

Intragrain Critical Current Density j_c
The value of critical current density j_c within a single grain, limited only by the pinning properties

Intrinsic Point Defect
Is the general term for either a vacancy or a Si interstitial in the Si matrix

Inversion
Condition in which, for a MIS structure, the applied voltage to the gate electrode V_G results in an increase of minority carriers near the surface of the underlying semiconductor. This region is referred to as an inversion layer. When present in a MISFET, this condition results in a conducting channel between the source and drain regions. This condition occurs for applied voltages beyond the threshold voltage, i. e. $V_{th} < V_G$

Inversion Symmetry
A system in which the laws of physics are unchanged by the operation of inversion

Ion Implantation
A method of introducing impurities into the surface of a solid. The impurity atoms to be introduced are ionized and accelerated by a high voltage, up to 10^6 V in some instances. They penetrate the surface to a depth dependent on their energy. Unfortunately, this disrupts the crystal lattice by introducing irradiation damage, which in turn must be repaired by annealing at elevated temperatures before the electrical activity of the implanted atoms can be obtained

Ionization
Process of forming an electrically charged atom (ion)

Ionization Energy
The energy distance from the valence-band maximum to the vacuum level at semiconductor surfaces

Irreversibility Line
Situated well below the upper critical field, this line separates a region without pinning and zero j_c, above the irreversibility field, from a region with pinned vortices ($j_c > 0$)

Irvin Curve
The experimental resistivity ρ of Si at room temperature versus N, the free-carrier concentration

Isostructural Solid Solution
A type of alloy solid solution in which an atom substituted in a lattice has the same valence electron structure as the atom it replaces

K

Kikuchi Lines
Diffuse background of lines in a diffraction pattern formed by the elastic scattering of incoherently scattered electrons

Knudsen Cell
High-vacuum oven from which molecular beams can be produced in molecular-beam epitaxy. These ovens are usually fitted with temperature sensors. Shutters mounted in front allow the beams to be turned on and off, as required

Kramers–Kronig Relations (KKR)
Dual integral relations between the real and imaginary parts of a one-sided Fourier transform of causal functions of time, which includes all physical admittances. KKR allow us to define the optical constants based on KKR relations for amplitude and phase of complex reflection coefficients from an opaque sample of the material under measurement

Kyropoulos Technique
This is similar to the Czochralski method but the extent of pulling is limited, producing crystals of greater diameter but shorter length

L

Langmuir–Blodgett Film Deposition
Method to build up multilayer structures of organic materials by the transfer of a monolayer floating on a water surface to a solid substrate as the latter is raised and lowered through the monolayer/water interface. Deposition modes include Y-type (monolayer transfer from the water surface to the substrate on both the upward and downward motion of the latter through the monolayer/air interface), X-type (film transfer only on downward motion of the substrate), and Z-type (film transfer only on the upstroke)

Laporte's Rule
Transitions between states of the same parity are forbidden

Laser Crystallization
A useful method to fabricate polycrystalline semiconductors showing high electron mobility by irradiating laser beams on amorphous semiconductor thin films

Laser-Induced Fluorescence (LIF)
Technique to determine the density of gas-phase atoms and molecules whose optically excited states are emissive states

Lattice-Matching
Growth of quantum wells on a substrate with the same, or very similar, lattice constant

Law of Mass Action
An equation linking the concentrations of the various reacting species in a chemical-type equation, derived from defect chemistry analysis

Lifetime
Minority-carrier lifetime is one of the electrical parameters determining the performance of IR devices; it can be affected by impurities and/or dopants

Light-Induced Defect Creation
The defect creation associated with illumination of samples. This is normally observed in amorphous materials having a flexible network

Lineage
A reduction in the total dislocation strain energy can be obtained either by polygonization or by dislocations arranging themselves in linear arrays. The latter results in the creation of a small-angle grain boundary or lineage. The reader will find examples of this in most books on materials science

Liquid Crystal
State of matter with properties characteristic of both a liquid and a crystal; that is, it flows like a liquid and certain properties are anisotropic like those of a crystal. At the molecular level the phase has long-range orientational order and some element of translational disorder in the long range

Liquid-Encapsulated Czochralski Technique
An inert layer, usually B_2O_3, is used, which floats on the top of the melt to prevent loss of volatile components; this is used for As and P-containing compounds

Liquid-Phase Epitaxy (LPE)
LPE is epitaxial growth of a layer by dissolving the required material in a liquid. On cooling, the material becomes supersaturated and forms a solid film. If this film is deposited on a substrate, the growth can be epitaxial

Localized Vibrational Mode (LVM) Absorption
A form of mid to far-infrared absorption resulting from simple vibrations of light atoms in a heavier lattice. Such vibrations do not couple well to lattice vibrations and result in sharp absorption lines, whose frequencies are directly related to mass and whose strengths are proportional to concentration. LVM absorption is particularly suited for measuring carbon concentrations in SI GaAs

Long Wave (LW)
This term refers to the transparent atmospheric window between the wavelengths 7.5 and 14 μm (sometimes called the thermal band). The edges of

this window are influenced by the water-vapor content and atmospheric conditions

Longitudinal Magnetic Recording
Recording system consisting of a write head and a read head and a recording medium where the magnetization is parallel to the surface of the recording medium

Long-Range Disorder
Defined by the absence of the long-range order, i. e., the lack of periodic arrangements of atoms

Low-Dimensional Structure
Heterostructure with dimensions comparable to the wavelength of an electron or hole, typically less than 100 Å, so that quantum effects are important

Lowest Unoccupied Molecular Orbital (LUMO)
This refers to the lowest-energy molecular orbital of an atom or molecule that does not contain an electron. If the atom or molecule were to accept an electron, it would be most likely to do it with this orbital

Low-Frequency Noise
Electrical measurement of the current or voltage noise spectral densities of a semiconductor component, typically at frequencies from ≈ 1 Hz to ≈ 100 kHz. The units are in A^2/Hz for current noise spectral density and V^2/Hz for voltage noise spectral density. It is a sensitive electrical technique that can be used to probe microscopic defects in semiconductor components and as a gauge of their reliability

Low-Temperature Co-fired Ceramic (LTCC)
A composite material structure made of alumina bonded at a temperature below the sintering temperature of alumina using a glass binder. In a substrate form, LTCC can be co-fired (melting the glass and bonding the alumina while capturing the metal interconnects and lines) at lower temperatures, so it is possible to use higher conductivity materials like Cu or Ag conductors in the ceramic

Low-Temperature Solution Growth
Normally used for water-soluble materials, with growth being progressed by either slow cooling or solvent evaporation

L-Pit
See *A-Swirl*

Luminescence
In general terms, this is the emission of light by a luminescent material (also called phosphor) due to conversion of a certain type of energy into electromagnetic radiation over and above thermal radiation. The luminescence is the light emitted by nonthermal sources in contrast with the emission of radiation from a heated object, which is called incandescence. In accordance with the source of energy the luminescence may be the photoluminescence (excited by external illumination), electroluminescence (induced by the passage of electrical current), cathodoluminescence (excited by irradiation with electrons), triboluminescence (excited by mechanical treatment, e.g., grinding), chemiluminescence (emitted during chemical reaction), bioluminescence (appearing as a result of biological processes), thermoluminescence (cased by temperature rise), and so on

M

Magnetic Annealing
Thermal process involving a ferromagnetic sample in a magnetic field in which the sample has induced a direction of easy magnetization (easy axis)

Magnetic Mirror
A planar metamaterial, which acts as a mirror that preserves, rather than reverses, the phase of electromagnetic waves upon reflection

Magnetic Random-Access Memory (MRAM)
Digital memory device composed of ferromagnetic thin films coupled to current-carrying conductors used to establish the direction of the magnetization in the ferromagnetic films. The resistance of the coupled films is different when the direction of magnetization in the two ferromagnetic films changes

Magnetic Spacing
Spacing between the bottom of the slider and the center of the ferromagnetic storage layer in a recording medium characterized by the magnetic spacing parameter d

Magnetic Transition
In digital magnetic recording, the region between opposite states of the magnetization in the recording medium. The length of the transition (l) is $l = \pi a$, where a is the transition parameter

Magnetic Tunneling Junction (MTJ)
Magnetic random-access memory (MRAM) in which the magnetic state of the ferromagnetic sample is sensed by the change in resistance of electron current tunneling through a thin insulating layer when the magnetization in the two electrodes coupled to the ferromagnetic sample changes from parallel to antiparallel orientation

Magnetoresistance
Change in resistance of a ferromagnetic material due to changes in the orientation of the magnetization with respect to an induced easy axis. It is sometimes referred to as anisotropic magnetoresistance (AMR) and is characterized by the change in resistivity with magnetic field normalized to the nominal resistivity $\Delta\rho/\rho$

Magnetoresistivity (MR)
The change of resistivity ρ in the applied magnetic field. For Si, MR is a diverse tensor reflecting the symmetry of the CB and VB

Majority Carriers
Electrons in an n-type and holes in a p-type semiconductor. Majority-carrier mobility is the mobility of electrons (holes) in n-type (p-type) material

Masking Effect
Severe suppression of the magnitude of photocurrent due to the development of a dead region in front of the

active region, in which the incident solar light is absorbed and no photocurrent is generated

Media Flux
Product of the remanent magnetization and the thickness of a magnetic recording film

Media Noise
Noise voltage at the terminals of a read head caused by fluctuations in the magnetization of the recording media

Medium Wave (MW)
The term for the transparent atmospheric window between the wavelengths 3.3 and 5 μm. There is a CO_2 absorption band around 4.25 μm, which divides the band into two with the better atmospheric transmission in the 3.3 to 4.2 μm band

Mercury Cadmium Telluride (MCT)
A narrow band gap II–VI semiconductor compound, $Hg_{1-x}Cd_xTe$, which is still the pre-eminent infrared (IR) material

Mercury-Sensitized Photo-CVD
Technique for preparing thin films by decomposing source gas materials by collision with photo-excited mercury atoms

Mesogen
A molecular compound or mixture of compounds capable of forming a liquid-crystal phase over a range of temperatures between the crystal and isotropic liquid phases

Meta-Atom
See *Metamolecule*

Metafilm
See *Metasurface*

Metal Semiconductor Field-Effect Transistor (MESFET)
Probably the simplest three-terminal device that can be fabricated on GaAs. The *metal semiconductor* refers to the gate structure, which is a simple Schottky barrier. The application of a reverse bias to this gate produces a depleted region that occludes part of the cross-sectional area of the channel, and thus modifies its conductivity

Metallic Nanocomposites
Composite materials made out of metallic quantum dots embedded in organic or glass hosts

Metal-Organic Chemical Vapor Deposition (MOCVD)
Technique for depositing thin films of compound semiconductors from a chemically reactive vapor phase where at least one of the components is a metal-organic

Metal-Organic Molecular-Beam Epitaxy (MOMBE)
Form of MBE where some or all beams are generated from external gas sources, in III–V MOMBE only the group III beams are from metal organic sources

Metalorganic Vapor-Phase Epitaxy (MOVPE)
A low-temperature growth technique using metal alkyls (and elemental Hg in the case of Hg-based ternary compounds) as the sources

Metal/Oxide/Semiconductor (MOS)
Structure that is one possible realization of *MIS*. The insulator has traditionally been SiO_2 (the oxide) due to its superior material and electrical properties, resulting in the abbreviation MOS

Metamaterial
Meta (a Greek preposition that can mean *after* or *beyond*) materials are artificial materials constructed with specific geometries to produce exotic behavior. Unconventional optical responses, such as a negative refractive index (leading to superlenses), result from these structures. Terahertz metamaterials are usually formed by patterning an electrical conductor, such as gold, on a substrate of an electrical insulator, such as plastic or undoped semiconductors. Metamaterials find further application in the terahertz spectral region by offering control of polarization and the possibility of *cloaking* or *invisibility*

Metamolecule
An elementary building block of a metamaterial, which is artificially built on a sub-wavelength scale from common constituent materials such as metals and dielectrics

Metastability
A term that refers to the nonequilibrium nature of glasses or amorphous solids. Amorphous solids have excess internal energy relative to the corresponding crystalline state or states of the same material. The method of manufacture, such as melt quenching, inhibits a transition to the lowest-energy crystalline state

Metasurface
A low-dimensional (planar) form of a metamaterial, represented by a thin metal film periodically patterned on a sub-wavelength scale. Although metasurfaces are many times thinner than the wavelength of light, they can strongly interact with electromagnetic waves, which they transmit, absorb, or reflect without diffraction, acting as media of zero dimension in the direction of light propagation

Microphotonics
Refers to the chip-scale manufacture of optical and photonic waveguide circuitry, using processing techniques borrowed from the microelectronics industry. Related to this is the need for high-density integrated optics, as facilitated by high-index-contrast waveguides and photonic crystals. By usual definition, microphotonics refers specifically to the monolithic manufacture of optical and photonic elements on silicon (CMOS) chips

Microwave Plasma Chemical Vapor Deposition (MPCVD) Technique
A technique for film deposition. This is now the most widely used technique for diamond growth. Energy is transferred by the microwaves to gas-phase electrons, which transfer their energy to the gas through collisions. The gas molecules dissociate, the active species are formed, and the deposition of diamond onto the substrate immersed in the plasma occurs

Miller's Rule
An empirical rule proposed by Miller in 1964, which

suggests that $\chi^{(2)}/\{\chi^{(1)}\}^3$ is nearly constant for all non-centrosymmetric crystals

Miniband
Interval of allowed energies for carriers in a superlattice, resulting from the delocalization and broadening of the quantum-well energy levels

Minimum Thermal Conductivity
The minimum value of lattice thermal conductivity of a solid, corresponding to a phonon mean free path on the order of one interatomic spacing

Minority Carriers
Electrons in a p-type and holes in an n-type semiconductor. Many electron devices work on the base of minority-carriers transport. Minority-carrier mobility is the mobility of electrons (holes) in p-type (n-type) material

Misfit Dislocations
These are dislocations introduced near the boundary of an epitaxial layer and the substrate, when a mismatch between the lattice constants exists. The density of these dislocations is proportional to the mismatch. Although they run parallel to the boundary, they can interact and penetrate much of the epitaxial layer; see *Threading Dislocations*

Mixed State
In type II superconductors, this state exists between the lower and upper critical fields. In the mixed state, superconducting and normal regions coexist. Magnetic flux enters the superconductor in the normal cores of the flux lines

Mobility
Parameter defined as the ratio of the carrier velocity (cm/s) to the electric field through which it is moving (V/cm). It is expressed in units of cm^2/Vs. At the microscopic level, it is related to the dominant scattering time and the effective mass of the carrier. The notion includes drift mobility μ, which corresponds to the motion parallel to the field and Hall mobility μ^H, which corresponds to motion perpendicular to the electric field, when a magnetic field is applied

Mobility Edge
The boundary between localized and delocalized states in a band

Mobility Gap
Energy separation between two mobility edges of the conduction and valence bands

Modulation Transfer Function (MTF)
MTF measures the efficiency of an imaging system such as a detector to resolve (transfer) different spatial frequencies of information in an image. In other words, MTF is the relative signal response of the system as a function spatial frequency

Molecular Electronics
Exploitation of organic materials for electronics and optoelectronics applications. Examples are displays based on liquid crystals and organic electroluminescent polymers. In the case of organic polymers, the area is sometimes referred to as plastic electronics

Molecular Reorientation
Nonlinear process in which the orientation of molecules in liquid changes upon illumination with intense light

Molecular-Beam Epitaxy (MBE)
Low-temperature growth technique for epitaxial films from atomic or molecular beams from thermal evaporation sources. It is carried out in ultra-high vacuum. Chemical Beam Epitaxy (CBE) is a form of molecular-beam epitaxy in which group III and group V beams are generated from external gas sources. Gas Source MBE (GSMBE) is a form of MBE where some or all beams are generated from external gas sources, in III–V GSMBE only the group V beams are from hydride sources

Molecularly Doped Polymers (MDP)
Formulations of charge-transport materials dissolved in a polymer matrix. CTLs of OPCs are typically MDPs

Monolayer (ML)
A single layer of atoms (in III–V MBE a layer of the compound or alloy, e.g. a layer of Ga + As)

Moore's Law
Prediction based upon the empirical observation by G. Moore that the minimum cost of manufacturing integrated circuits per component actually decreases with the increase in the number of IC components, and thus with greater circuit functionality and computing power. A corollary of this observation is that the density of integrated circuit transistors will double roughly every 1.5–2 years (see *Scaling*)

Moseley's Law
States that the square root of the frequency of characteristic X-rays, for certain elements, is linearly related to the atomic number

Multi-Crystalline Silicon (mc-Si)
Wafers of silicon, cheaper to produce than the single-crystal silicon wafers but with multiple single-crystal grains in each wafer

Multiple Quantum Well (MQW)
Structure composed of a stack of QWs separated by sufficiently thick barriers so that the electron wave functions and energy levels are localized to each QW

N

Nanowire
Quantum wire formed by seeded growth on a nano-patterned substrate

Narrow Band gap
Refers to a semiconductor with a forbidden energy gap of less than about 0.7 eV, making it suitable for detection in the infrared wavebands

Native Defect
A vacancy, a self-interstitial, an anti-site, or any complex of these

Negative Differential Resistance (NDR)
(See *Transferred Electron Effect*). In GaAs, electrons that undergo the transferred electron effect are excited from the primary conduction band minimum, where their effective mass is low, to subsidiary minima, where their effective mass is considerably greater. As a result, although they have greater kinetic energy, their drift velocity is lower. This, in turn, results in a reduction of current in the external circuit. The current–voltage characteristics show a reduction in current after a critical voltage is applied (although the ratio of current to voltage will always be positive). This is NDR. It can be shown that current flow in the NDR region is unstable and that charge tends to be transported in groups, resulting in oscillations in the current flow. This is the Gunn effect and is used in some microwave sources

Negative Index Medium
A medium characterized by the refractive index less than zero

Negative Refraction
An exotic optical effect characterized by anomalous change in the direction of light propagation upon refraction, when a beam of light bends the *wrong* way while crossing an interface between two media

Nematic Phase
The simplest of the liquid-crystal phases; it has long-range orientational order but is devoid of long-range translational order

Noise-Equivalent Temperature Difference (NETD)
A measure of sensitivity for a multiplexed infrared detector and the change in scene temperature that produces a signal equivalent to the rms noise level of the detector

Nonlinear Directional Coupler
A device where a part of one of two waveguides is made out of a nonlinear material. Changing the intensity of the incident light changes the effective path length experienced by the waveguided light, thus controlling coupling between the two waveguides

Nonlinear Fabry–Pérot Interferometer
A device that consists of two mirrors separated by a nonlinear material. As the refractive index of the nonlinear material changes with an increased level of illumination, the effective path length of the resonator is altered, changing the transmission properties of the device

Nonlinear Figure of Merit
A figure of merit that describes the applicability of a refractive nonlinear optical material in terms of nonlinear index change and effective absorption

Nonlinear Kramers–Kronig Transformations
A set of two relations that permits calculation of the real (imaginary) part of nonlinear response at a given optical frequency given the knowledge of the imaginary (real) nonlinear response at all other optical frequencies

Nonlinear Mach–Zehnder Modulator
A Mach–Zehnder modulator in which a part of the waveguide is made out of a nonlinear material. Changing the intensity of the incident light changes the effective path length experienced by the light at the nonlinear waveguide, controlling transmission

Nonlinear Material Figures of Merit
A set of metrics that allows us to quantify the applicability of a nonlinear optical material to optical switching

Nonlinear Optical Medium
A material whose properties change upon illumination with intense light

Nonlinear Optical Switching
Switching of optical signals using an illumination-dependent phase shift of nonlinear materials

Nonlinear Periodic Structure
A Bragg periodic structure in which at least one set of layers is nonlinear

Nonlinear Phase Shift
A difference between the phase shift experienced by intense light and a phase shift experienced by light which intensity approaches zero

Nonlinear Polarization
Polarization that does not experience purely linear dependence on the electric field

Non-Stoichiometry
Defined as the difference between the total anion and cation concentrations, hence anion- or cation-rich

n-Type Conductivity
A semiconductor material, with electrons as the majority charge carriers, that is formed by doping with donor atoms

Nucleation Temperature
Temperature where stable defect clusters/aggregates start to form as nuclei

Numerical Aperture
Figure of merit used to described the power of a lens, depending on the angle of collection of the lens and the refractive index of the medium in which the lens is situated

O

Onsager Model
A model that describes the probability of geminate recombination as a function of the applied field with the efficiency of geminate pair photogeneration and the initial pair-separation distance as parameters

Optical Absorption Coefficient
4π times extinction index, divided by the radiation wavelength

Optical Bistability
A phenomenon in which the instantaneous transmittance of the device depends both on the level of incident illumination and on the prior transmittance of the device. Such an element enables all-optical switches

Optical Constants
The real and imaginary parts of the square root of the *dielectric constant/susceptibility*, called refraction and extinction index, respectively; they define the optical properties of bulk material and mesoscale structures

Optical Emission Spectroscopy (OES)
A technique to determine the plasma parameter and reactions in the plasma by measuring the emission intensity of line spectra from the plasma

Optical Gap
(See *Band gap*)

Optical Limiter
A device in which the transmittance decreases with increased level of illumination

Optical Properties of a Sample
These include the apparent reflectance, transmittance, and absorbance

Optical Texture
Pattern observed for thin slabs of birefringent material between crossed polarizers under a microscope. It is used as a fingerprint to help identify the many different liquid-crystal phases

Optically Detected Magnetic Resonance
Magnetic resonance observed by optical means, particularly by detection of electron spin resonance by monitoring luminescence intensity and polarization, etc.

Order Parameter
The parameter that measures the extent of long-range order characteristics of a phase. The orientational order parameter is the most important for liquid crystals; by definition it vanishes in the isotropic phase

Organic Photoconductors (OPC)
Single material or a formulated blend of materials that have photoconductive characteristics

Organic Photoreceptors
Thin-film multilayer devices made from organic photoconductive materials. These devices are often called organic photoconductors

Oscillator Strength
Measure of the probability of a transition between levels; an oscillator strength of 0.01–1 is highly probable (*allowed*) and has a short lifetime (ns), one less than 0.001 is improbable (partially *forbidden*), and has a long lifetime (μs–ms)

Oxidation-Induced Stacking Fault
A stacking fault that shows up after wafer oxidation. During oxidation, Si interstitials are injected from the wafer surface into the bulk. They aggregate around oxygen precipitates of a critical size and thereby squeeze an additional lattice plane (stacking fault) between two regular lattice planes

Oxygen Precipitates
Aggregates of oxygen atoms that form in Czochralski-grown crystals due to their relatively large oxygen content. The silica crucible that holds the Si melt is slowly dissolved during the growth process, which introduces oxygen into the melt and, hence, into the growing crystal

P

Paraelectric
Non-polar phase that transforms into the ferroelectric phase at the Curie temperature

Partial Response Maximum-Likelihood (PRML) Recording Channel
Particular form of equalization used in digital recording channels. The detector is maximum likelihood

Passivation
Refers to the removal of electrical activity of a defect, often by trapping a mobile atom of the opposite electrical type (a donor may trap a mobile acceptor, for example). The defect now becomes a complex, consisting of the original defect and the new component in close proximity, but without electrical properties. Unlike compensation, ionized impurity scattering is reduced by passivation and carrier mobility is increased

Passive Matrix Addressing
Also known as simple multiplexing, passive matrix addressing is a technique for writing images onto liquid-crystal displays. The display elements are arranged as a matrix of rows and columns, and a series of voltage pulses are applied to each row in sequence. Individual pixels are activated by applying a voltage to the relevant column, such that the sum of the row and column voltages exceeds the desired switching or threshold voltage. Each pixel responds to the root mean square voltage applied during the line-address time, and unlike the active-matrix addressed displays, there is no charge-storage facility. Thus, the number of rows that can be addressed is limited

Peltier Effect
The absorption or evolution of heat at the junction of two different conductors

Penetration Depth (λ_L)
The characteristic length scale for the penetration of a magnetic field into the surface layer of a superconductor

Perfect Absorber
A planar metamaterial that completely absorbs incident electromagnetic energy (light)

Perfect Diamagnetism (Meissner Effect)
Of the superconducting state is reflected by the fact that a magnetic field is expelled from the interior of a field-cooled superconductor as soon as the superconducting state is reached

Permalloy
Alloy of nickel and iron with approximately 80% Ni and 20% Fe. Composition of alloy with zero magnetostriction

Permittivity (ε)
A measure of the polarizability of a dielectric material, which is frequency dependent. The relative permittivity of a material is often given by ε or ε_r, such as with the parallel-plate capacitor expression for capacitance: $C = \varepsilon\varepsilon_0 A/t$ (see *Dielectric Constant*)

Perpendicular Magnetic Recording
Recording system consisting of a write and read head and a recording medium where the magnetization is perpendicular to the surface of the recording medium

Phase-Separated Structure
Blended junction having routes for electrons and holes, which are formed by percolation due to self-assembly, while maintaining donor/acceptor molecular contacts in the bulk of the organic film

Phonon Glass Electron Crystal
A crystalline solid that exhibits the thermal properties of a glassy or amorphous material and the electric properties of a good crystal

Phonons
Quantized lattice vibrations in a solid

Phosphorescence
Luminescence with a lifetime $\gtrsim 10$ ns, typically involving a metastable state in the pumping cycle

Photo-Catalysis
An enhancement of the reaction rate on the surface by shining UV radiation onto the growing film

Photochromic
Temporal coloring induced by (UV) light illumination. For instance, photo-structural changes of Ag particles, which are dispersed in oxide glasses

Photoconductivity
The part of the conductivity that is caused by the absorption of light. Negative values of the photoconductivity are possible when optical excitation activates efficient carrier traps. A photoconductive device uses the change in resistance of a slab of semiconductor to measure the extra electron–hole pairs created by the absorption of photons

Photocurrent
The current that is generated as a result of the absorption of light

Photodarkening
Quasi-stable darkening induced by light illumination. In chalcogenide glasses, it occurs with a red-shift of optical absorption edge, which is induced through athermal photo-structural processes. The process, however, is speculative

Photoemission Spectroscopy (PES) Technique
PES is a method to determine the barrier heights of Schottky contacts and band offsets of semiconductor heterostructures from energy-distribution curves of electrons photoemitted from Schottky contacts and semiconductor heterostructures excited by ultraviolet light (UPS), soft X-rays (SXPS), and X-rays (XPS)

Photoinduced Degradation
A deterioration of semiconductor properties by prolonged light exposure

Photoinduced Discharge Characteristic (PIDC)
Plot of exposed potential versus exposure using either a continuous exposure or flash exposures of varying intensity

Photoinduced Effects
Changes in the properties of a glass induced by light, involving transitions between metastable states of the glass or changes in defect sites within the glass. Typically, a laser beam is used to locally modify the refractive index, density, absorption coefficient, etc., of the glass (see *Photodarkening*, *Photo-Induced Degradation*). These processes are widely used to pattern photonic structures such as Bragg gratings, waveguides, and refractive lenses into glasses

Photoisomerization
Light-induced change in the molecular structure of a molecule

Photoluminescence
Luminescence excited by the external illumination of the material. It should be distinguished from reflection and light scattering, which are also caused by external illumination and are not connected with thermal radiation of the material. The criterion to distinguish these phenomena is the characteristic decay time after the cessation of the incident light. While reflection and scattering would decay within $\approx 10^{-14}$ s (characteristic time of electromagnetic oscillations of incident light wave), the photoluminescence would persist at least more than $\approx 10^{-12}$ s (characteristic time of atomic vibrations of luminescent material)

Photon
A quantum of electromagnetic radiation

Photonic Glass
High-purity glasses in which impurities are controlled or suppressed at ppm levels. An example is the optical fiber glass developed at the end of the twentieth century, which is very recent when compared to the long history of artificial glasses of 5000 years

Photorefractive Beam Fanning
Spreading of an intense light beam due to the photorefractive properties of a material

Photosensitivity
of an organic photoconductor is a measure of the rate of photodischarge when exposed to light. Typical characterizations are the initial slope of the photodischarge (V/J) and the energy required for photodischarge to half the initial surface potential

Photo-Thermal Ionization Spectroscopy
A very sensitive technique for the measurement and identification of shallow impurities in semiconductors. The measurement is a two-stage process; the absorption of light excites the electron on un-ionized donors into a higher level and thermal energy now ionizes this electron into the conduction band, where it can be measured as a photo-current. Because the energy of the optical ionization is dependent on the donor type, the photocurrent spectrum as a function of photon energy, displays peaks that are characteristic of the donors present

Photovoltaic Device
This utilizes a p-n junction in a semiconductor to separate electron–hole pairs (created by the absorption of photons) to generate a voltage

Photovoltaic Module
A complete encapsulated device suitable for mounting on a roof or building façade. Modules will often

comprise a number of cells and the typical size is of the order of 1 m^2

Photovoltaic Solar Cell
A semiconductor device for the conversion of solar energy into electricity

Piezoelectric Polarization
Generation of electric polarization in certain dielectric crystals as a result of the application of mechanical stress

Piezoelectricity
Property of some crystalline materials, which produce an electric charge when subjected to an externally applied force (direct effect). They also deform when subjected to an external electric field (reverse effect)

Piezoresistivity
Property of certain materials, including thick-film resistors, whereby an externally applied force gives rise to a change in resistance

Pinned Film
Coupling of a ferromagnetic to an antiferromagnetic film resulting in the magnetization of the ferromagnetic film being constrained to a fixed direction. The strength of the coupling is characterized by an exchange field parameter (H_{ua}) measured in Oe or A/m

Pixie Dust
Thin layer of Ru as used in synthetic antiferromagnetic media (SAF)

Planar Chirality
Two-dimensional form of mirror asymmetry, which implies that a planar object or a pattern cannot be brought into congruence with its mirror image (obtained by reflection across a line in the object's plane) unless it is lifted off the plane. The simplest example of a planar chiral object is the Archimedean spiral

Planar Lightwave Circuit (PLC)
Industrially established processes for manufacturing integrated optics devices in silica-based glasses deposited on silicon wafers. Typically, the glass layers are deposited by chemical vapor deposition or flame hydrolysis. These technologies were developed mainly for applications in fiber optics and are widely used to manufacture wavelength multiplexers

Plasma-Enhanced Chemical Vapor Deposition (PECVD) Technique
Like CVD, this is a technique for thin-film deposition. In contrast to the CVD technique, in the PECVD technique the deposition of a film is a nonequilibrium process. In the PECVD technique, the gas-phase reactions, which are activated by temperature in the CVD technique, originate from an interaction of electrons with the gas molecules injected in the reaction chamber. In the plasma, the dissociation of the gas precursors is obtained by their collisions with energetic electrons. The products of the reactions interact with the substrate surface and lead to the deposition of a thin film. As the dissociation of the gases is produced by electron impacts, the reactor and substrate can be kept at room temperature

Plasmon
Quantized collective motions of electron gas in a metal. In a bulk metal, the plasmon propagates as a longitudinal wave, which may be probed by an electron beam. In a metal nanoparticle, a transversal surface plasmon can be excited by light waves

Plastic Deformation
Deformation of a body caused by an applied stress, which remains after the stress is removed

Plastic Electronics
Use of polymers in electronic and optoelectronic devices (see *Molecular Electronics*)

Point Defect
Smallest structural element, or imperfection, to cause departure from a perfect lattice structure, e.g., a dopant or impurity atom

Poisson's Equation
A differential equation relating the spatial gradient of the local electric field to the local space-charge density

Poisson's Ratio (ν)
The ratio of transverse contraction strain to longitudinal extension strain in the direction of stretching force. Tensile deformation is considered positive and compressive deformation is considered negative. The definition of Poisson's ratio contains a minus sign so that normal materials have a positive ratio

Polarity
Property of a physical system that has two points with different characteristics, such as one that has opposite charges or electric potentials

Polarization Ratio
The density of states at the Fermi energy in an energy band for electrons in a metal with spins parallel (antiparallel) to the local magnetization is $D(E_F^{\uparrow})$ and $D(E_F^{\downarrow})$, respectively. The polarization ratio is $P = \frac{D(E_F^{\uparrow}) - D(E_F^{\downarrow})}{D(E_F^{\uparrow}) + D(E_F^{\downarrow})}$

Poles of a Write Head
Structure in a write head that couples the magnetic fields to the write gap

Polycrystalline Solid
Polycrystalline material is not a single crystal as a whole, but is composed of many small crystals randomly oriented in different directions. The small crystals in polycrystalline solids are called grains. Theses grains have irregular shapes and orientations. A polycrystalline material has grain boundaries where differently oriented crystals meet. Polycrystalline silicon is produced in the form of a ribbon or thin film

Polydimethylsiloxane (PDMS)
$CH_3[Si(CH_3)_2O]_nSi(CH_3)_3$ is mainly used as silicon-based organic polymer

Polygonization
The motion of dislocations to form structures that resemble three-dimensional polygons. In this way, they minimize their strain energies. Readers are

referred to the many excellent books on materials science that discuss dislocations and their interactions

Polymer Thick Films
Thick films that are cured at much lower temperatures than their cermet counterparts. The binding matrix is a polymer material and the resulting films can be used in flexible circuits

Polymethyl-Methacrylate (PMMA)
$(CH_2{=}C(CH_3)COOCH_3)_n$ is an organic compound that is widely used to transfer graphene sheets

Positron Annihilation
Analytical technique that is sensitive to open defects, such as vacancies, in a crystal. The lifetime of positrons is greater when the total number of electrons is reduced, as happens at vacancies. Measuring positron lifetime can give semiquantitative estimates of the concentrations of these defects

Power Stabilizer
Optical device that can provide a fixed output intensity upon varying input intensity. The principle may be based on multi-photon absorption, which becomes efficient when light is more intense

Precursor
Volatile compound containing the element required for deposition onto the substrate

Process Integration
The carefully engineered combination of materials and processes to enable the fabrication of an integrated circuit technology. The careful consideration of materials properties, device physics, and electrical engineering principles is required for successful integration of materials into an IC chip

Proper Ferroelectric
Ferroelectric material in which the spontaneous polarization is the primary order parameter

Pseudomorphic Layers
True epitaxy only occurs when the lattice constants of the substrate (material A) and layer (material B) are equal. If they are different, misfit dislocations or even gross bending of the structure can occur. In pseudomorphic growth, the lattice constants can be very different. However, before misfit dislocations can be formed, either the growth of B is stopped or a layer of the substrate material, A, is grown on top of the structure. In either case, the structure is stabilized with material B being severely strained. Manufacturers use this technique to produce devices with advanced electrical or optoelectronic properties

p-Type Conductivity
Semiconductor material in which the dopants create holes as the majority charge carrier. It is formed by doping with acceptor atoms

Pulse Width
Width of a pulse in a digital magnetic recording read head when sensing a recorded transition. Characterized by the 50% width (PW_{50})

Pyroelectric Effect
Generation of an electric polarization change or a charge separation in a material in response to a change in material temperature

Pyrolytic Boron Nitride (pBN)
The favored crucible material for GaAs growth from the melt and for Knudsen cells for MBE. Pyrolytic refers to its high-temperature capabilities

Q

Quantum Cascade Laser (QCL)
The QCL is a source of radiation suited to the terahertz and infrared regions. *Quantum* refers to quantization of energy levels within a narrow region realized by interleaved energy wells and barriers formed by alternating layers of different semiconductor alloys. *Cascade* refers to the operation of a single electron tumbling down many potential steps, in contrast to other laser designs where many charge carriers encounter a single potential step. The *lasing* operation means that only one frequency (or a small range of frequencies) is available, and this is often built in during the manufacture, which is a limitation of lasers generally. QCLs promise future compact, robust, and reliable sources of terahertz radiation but presently suffer the drawback of having to be cooled to operate

Quantum Confined Stark Effect (QCSE)
The change in near-band-edge optical absorption in quantum wells that takes place when a high electric field is present

Quantum Dot (QD)
A semiconductor quantum heterostructure that is quantum confined in all three dimensions, for example, InAs islands embedded in a layer of GaAs and Ge islands in Si

Quantum-Dot Infrared Photodetector (QDIP)
Infrared detector based on intersubband transitions in quantum dots

Quantum Efficiency
Indicates either the probability that the absorption of a photon will result in the creation of a free electron and hole or the ratio of the number of luminescence photons to the number of stimulating photons. When determined by xerographic discharge of the photoreceptor, QE is a measure of the overall effectiveness of surface charge neutralization per absorbed (or incident) photon. It is sometimes called xerographic gain or supply efficiency. When determined by spectroscopic methods, such as fluorescence quenching, it is a measure of charge generation within the light-absorbing moiety

Quantum Size Effect
Quantum effect associated with quantization of the band and level. It is observed in two, one, and zero-dimensional systems whose size becomes small

Quantum Well (QW)
Low-dimensional semiconductor quantum heterostructure where a thin layer with a lower band gap is sandwiched between layers (barriers) with a higher band gap, leading to spatial confinement of carriers along one dimension for which the energy is

quantized in discrete levels. Carriers have free motion in the plane perpendicular to the confinement direction. May be repeated to produce a multiple quantum well (MQW)

Quantum Well Intersubband Photodetector (QWIP)
Long-wavelength (IR) detector based on light absorption through intersubband electron transitions in n-doped QWs

Quantum Wire
Semiconductor quantum heterostructure with two-dimensional confinement of carrier motion. Carriers have just one direction of free motion

Quasi-Fermi Levels
Levels that correspond to the energy positions the equilibrium Fermi level would need to have in the gap in order to produce carrier densities equivalent to the ones that are generated by the illumination

Quasi Particles
Unpaired electrons excited above the energy gap

R

Radiation Resistance
This is important for solar cells operating in space where high doses of gamma rays and high energy particles can significantly degrade the efficiency of the cells

Radio-Frequency (RF) Circuits
High-speed and microwave analog circuits that operate in the range 0.1−100 GHz

Reactive Sputtering
Technique for preparing thin films by sputtering with a mixture of reactive gas and conventional sputtering gas such as oxygen and argon

Reactor Cell
Chamber where the precursors react to deposit a film onto the substrate

Read Head
In magnetic recording the ferromagnetic device used to generate a voltage proportional to the state of magnetization in the recording medium

Readout Integrated Circuit (ROIC)
Commonly describes the silicon chip on which the detector material is mounted. The role of the ROIC is to integrate the signal, perform some signal processing and readout the array. Other terms are: multiplexer or mux

Reciprocal Lattice
Theoretical lattice constructed from a real lattice, such that any vector from the origin to a diffracted spot is normal to a particular plane in the real lattice, with reciprocal length of that plane spacing

Recombination
Process whereby non-equilibrium populations of electrons and holes return to their equilibrium values. Depending on whether the recombination rate is proportional to the excess carrier density or the square of that quantity, the processes can be referred to as linear and quadratic, or alternatively as a monomolecular and bimolecular recombination. Radiative recombination refers to the transfer of at least some of the energy of the excess carriers into photons. Non-radiative recombination involves only the production by lattice phonons (heat)

Reflectance-Transmittance (R–T) Method
Defines the optical constants, based on two equations connecting the former to the reflectance and transmittance of the slab sample of material under measurement

Reflection High-Energy Electron Diffraction (RHEED)
Glancing-angle electron diffraction technique, sensitive to surface reconstruction and morphology. Key in situ analytical technique in MBE

Reflectometer
Instrument used for reflecting light off the substrate to monitor the growth process. This is normally achieved through detecting the interference modulation of the light intensity from a growing film

Reflow
Process of heating a glass above its glass-transition temperature, to the point that its viscosity is sufficiently reduced to enable the material to flow. In combination with surface tension effects or other external forces, reflow is often exploited in the reshaping of optical devices

Remanent Magnetization
Magnetization that remains in a sample after the magnetic fields are removed

Remanent Polarization
Dielectric polarization that remains in a ferroelectric material after an electric field has been applied

Resistivity
Parameter of a semiconductor that depends on the free electron and hole densities (cm^{-3}) and their respective mobilities (cm^2/Vs), and is expressed in units of Ω cm. It is the reciprocal of conductivity and it depends strongly on temperature in a semiconductor

Resolution
The smallest separation of two points in an object that can be distinctly reproduced within an image

Resonant Nonlinear Response
The nonlinear response taking place in the absorbing spectral region

Responsivity
Signal term often measured using a two-temperature black body. Units can be V/W (used for photoconductors) or V/photon (often used for multiplexed photodiodes)

Retrograde Solidus
Describes the shape of the solidus when it shrinks as the temperature is reduced. In GaAs, the increased width of the solidus at high temperatures indicates that concentrations of Ga or As, in excess of stoichiometry, can exist in the solid crystal. These concentrations must reduce as the crystal cools because of the retrograde solidus

Rutherford Scattering
Elastic scattering of electrons due to an electrostatic interaction with the nucleus and surrounding electrons of an atom

S

Saturated Vapor Pressure (SVP)
Partial pressure of a substance in equilibrium with its liquid, so the partial pressure is determined solely by the temperature of the liquid and vapor

Saturation Intensity
Intensity at which the effective absorption decreases to a half of its initial value

Saturation Magnetization
Maximum magnetization that a ferromagnetic sample can attain in magnetic fields larger than the coercive field

Saturation of Absorption
Resonant nonlinear process in which absorption decreases with an increased level of illumination

Scaling
Calculated reduction of integrated circuit element dimensions according to physical and engineering principles and constraints, as well as economic considerations (See *Moore's Law*)

Scanning Probe Microscopy
Generic term given to microscopy techniques that use a scanned micro or nanoscale tip in immediate proximity to a surface to image topography or other physical features with almost atomic resolution. The main types are atomic force microscopy (AFM) and scanning tunneling microscopy (STM)

Schottky Contacts
Metal–semiconductor contacts are also named Schottky contacts

Schottky–Mott Rule
Equates the barrier heights of n-type (p-type) Schottky contacts with the difference of the metal work function and the electron affinity (ionization energy) of the semiconductor. The rule is incorrect since it does not consider the *intrinsic* interface-induced gap states

Schrödinger Equations
Partial differential equations in quantum mechanics, which describe how the quantum state of a quantum system changes with time

Screen
Mask used to define the desired thick-film pattern. It is usually made of stainless steel, polyester, or nylon

Screen Printing
Method by which thick films are deposited onto substrates. See *Antiferromagnetically Coupled Media (AFC)*

Secondary Electrons
Low-energy (< 50 eV) electrons that escape from the near sample surface, used to form topographic images

Secondary-Ion Mass Spectrometry (SIMS) and Glow Discharge Mass Spectrometry (GDMS)
Mass-spectrometric techniques that are well suited for the chemical analysis of semiconductors. In addition to being extremely sensitive to most impurities, they are quantitative and can give valuable information regarding the distributions of concentrations with depth. In SIMS, high-energy primary ions are focused on the sample surface. These sputter atoms from the material under investigation. Those that are charged are passed into a mass spectrometer for analysis. In GDMS, the sputtering is accomplished by a glow discharge. SIMS is, by convention, generally classified as dynamic, in that the material surface layers are continually removed as they are being measured, and static, in which the ion dose during measurement is restricted to less than 10^{16} ions/m^2 in order to retain the surface in an essentially undamaged state

Seebeck Effect
The voltage induced on a material due to a temperature difference across it

Segregation Coefficient (k)
In growth from the melt the incorporation of an impurity into the crystal depends on the equilibrium ratio of the solubilities in the melt and the solid. For a dilute solution, this ratio is given by the ratio of the equilibrium value of impurity concentration at the solidus and the liquidus at the growth temperature. This ratio is the segregation coefficient, k. For most impurities, k is less than unity and the growing crystal contains a lower impurity concentration than in the source melt. The exploitation of this fact has resulted in the purification technique of *zone refining*

Selection Rule
Quantum-mechanical rule based on wavefunction and operator symmetry that determines oscillator strength; typically set by parity or angular-momentum considerations

Selective Epitaxy
Growth of a single-crystal layer in a window without any deposition on the surrounding mask layer

Selenization
Process of forming the CIGS or CIS layer by exposing a copper, indium, and gallium precursor layer to Se to form the alloy

Self-Assembly
Spontaneous formation of a layer of organic material on a solid substrate surface. The process usually takes place by immersing the substrate into a solution of the organic compound. The driving force for self-assembly can be chemical and/or electrostatic attraction

Self-Defocusing
A nonlinear process associated with negative refractive nonlinearity that results in spatial spreading of intense light

Self-Focusing
A nonlinear process associated with positive refractive nonlinearity that results in focusing of intense light

Semiconductor Detector
Device that converts the incident photons directly into electrical pulse

Semiconductor Nanocrystals
Composite materials made out of semiconductor quantum dots embedded in organic or glass hosts

Semi-Insulating
A term given to semiconductors whose resistivity lies

above about $10^6\ \Omega$ cm. In GaAs, semi-insulating properties normally lie in the range of $10^7-10^8\ \Omega$ cm

Separate Confinement Heterojunction (SCH) Laser
In the SCH laser, the optical and electrical confinements are achieved separately by altering the thickness and the alloy composition of the cladding layers

Shallow-Energy-Level Dopant
Doping impurities whose energy level lies very close to the conduction or valence band for donors or acceptors

Shear Modulus
Sometimes also called rigidity, this relates stress and strain according to Hooke's law and is a measure of a material's resistance to shearing stress. The shear modulus, therefore, has units of pressure

Sheet Resistance
Resistivity of a semiconductor sample divided by its thickness, measured in $\Omega/\square$. It is commonly used by integrated circuit designers when designing resistances by specifying the number of squares required to give a certain value of resistance

Shields
Soft ferromagnetic films used to direct the flux from a recording layer away from sensor films in a read head

Short Wave (SW)
Often used for wavelengths between 1.0 and 3 μm. The atmosphere is transparent in relatively narrow bands within the SW region, the most common one is between 2.0 and 2.25 μm, although 1.5 μm is also important, as this is the wavelength for eye-safe lasers

Short-Range Atomic Structure
Atomic bonding structures within a scale of ≈ 0.5 nm, which are characterized by coordination number (the number of nearest-neighbor atoms), bond length, and bond angle. It is demonstrated for such simple glasses as SiO_2 that the short-range structure is nearly the same with that in a corresponding crystal

Signal Decay Rate
Rate (R) at which the amplitude of a signal read back from a recording system decays with time

$$R = 100\frac{\frac{A(t)}{A(t_0)}-1}{\log(t/t_0)}$$

Signal Processing in the Element (SPRITE)
A device used in the UK common module camera. It relies on a strip of CMT with a high voltage bias to drift photon generated holes at the same speed as the image is scanned, thus resulting in an amplified signal at the end of the strip

Silicene
A single layer of sp^2-bonded Si atoms in a honeycomb shape

Single-Layer Organic Photoconductor
A photoconductor with an architecture where the charge-generation and charge-transport functions are combined into a single layer

Slider
In a disk drive, the structure used to support an air bearing over the recording medium on the write and read heads

Sliding Boat
Liquid-phase epitaxy technique in which the substrate is slid under the melt in a horizontal orientation

Slope Parameter
(In perpendicular magnetic recording) The slope of the magnetization curve at the coercive field times

$$\alpha = 4\pi\left(\frac{dM}{dH}\right)_{H=-H_c}$$

Slush
A homogeneous charge of a ternary is held across the liquidus–solidus gap with the lower end solid, the upper end liquid, and the central section in a slushy state, during recrystallization

Smectic Phase
Liquid-crystal phase with some long-range translational order in addition to the long-range orientational order of a nematic. The constituent rod-like molecules are arranged in layers giving translational order in one dimension

Soft Underlayer (SUL)
Magnetically soft (low-coercivity) film underneath the recording layer in a perpendicular magnetic recording system. Used as a low-reluctance path for the flux from the write head

Solar Cell
Semiconductor device that converts the energy of sunlight into electric energy. Also called photovoltaic cell

Solid/Liquid/Gaseous Phase Equilibria
These phase equilibria are essential for understanding both the various growth techniques used and the post-growth heat treatments, i. e., cool-down after growth and subsequent annealing stages

Solid-State Recrystallization (SSR)
Solid-state recrystallization is a growth technique that produces a homogeneous but polycrystalline charge that is subsequently recrystallized in the solid state to produce multi-grained material

Solidus
In an equilibrium phase diagram, the solidus is the line below which all the components are solid

Solitonic Propagation of Pulses
Propagation of pulses characterized by a lack of temporal pulse spreading

Spatial Frequency
Spatial frequency is the reciprocal of a spatial dimension (e.g., height x or width y in a two-dimensional image) similar to the temporal frequency, which is the reciprocal of time. However, the unit of spatial frequency is preferably expressed as line pairs/mm (lp/mm), instead of cycles/mm

Specific Heat (Capacity)
Amount of heat required to change a unit mass (or unit quantity, such as mole) of a substance by 1 degree in temperature

Spherical Aberration
A blurring of resolution due to the spread of path lengths of rays traveling from an object to the image

plane, arising from a variation of the focal length of a lens as a function of distance from the center of the lens

Spin Valve
Read head that uses the giant magnetoresistive effect in three metallic films: a ferromagnetic free film, a non-magnetic spacer film, and a ferromagnetic pinned film

Spontaneous Polarization
The electric polarization that a substance possesses in the absence of an external electric field

Sputter Depth Profile
Compositional depth profile obtained when the surface composition is measured as material is removed by sputtering. Note: in some analytical methods such as SIMS, the sputtering is often accomplished by the ion beam used for analysis, but in other methods an ion beam may need to be added

Sputtering
Process in which atoms and ions are ejected from the sample as a result of particle bombardment

Squeegee
Device used to transfer the thick-film paste through a screen and onto the substrate

Stepanov Technique
Crystals are pulled from a crucible containing a crystal-shaped aperture

Stoichiometry
State of a perfect compound where the ratio of the numbers of atoms of the elements is a simple fraction. In GaAs, for example, stoichiometry exists when the number of As and the number of Ga atoms are equal

Stokes Shift
Difference in energy between the maximum of the emission spectrum and the maximum of the excitation spectrum

Strain (ε)
Defined in elementary form as the change in length divided by the original length; it can be thought of as the movement of one corner of a cubic box from its initial position under stress

Strained Quantum Well
QW layer grown on a substrate with a different lattice constant, resulting in a significant strain of the QW layer lattice

Stress
Force per unit area provided either by gravity or by the flow of viscous fluid

Strong Anchoring
Strong anchoring of the director at a surface parallel to the easy axis is said to occur when an applied field is unable to alter the orientation of the director at the surface but can in the bulk

Structural Relaxation
Essentially an aging effect associated with glasses. Because glasses are metastable materials with random network structures, they are inherently subject to short or long-term changes in material properties. Often, structural relaxation is manifested by a change in specific volume (densification) at fixed temperature versus time. The rate of such changes is extremely sensitive to the difference between the glass transition temperature and the observation temperature. Structural relaxation can be induced rapidly by an annealing step, in which the glass is heated near its glass-transition temperature for some period of time

Substitutional Impurities
Impurities that replace the crystal's base atom at that base atom's lattice position

Substrate
Base material onto which a film is deposited. Examples of typical substrates include materials such as alumina, beryllia, aluminum nitride, silicon, insulated steels, and various plastics. In epitaxial processes, substrates are normally near lattice-matched pieces of material of a sufficiently large area for use in the various epitaxial growth processes, normally from similar, i. e., common-cation, ternary systems, e.g., CdZnTe for MCT

Superconductors of Type I
Characterized by the fact that normal and superconducting regions cannot coexist in these materials. The value of $\kappa = \lambda_L/\xi$ is smaller than $1/\sqrt{2}$

Superconductors of Type II
Type II superconductors, for which $\kappa = \lambda_L/\xi > 1/\sqrt{2}$, show perfect diamagnetism only below the lower critical field B_{c1}. Between the lower and the upper critical fields type II superconductors are in the mixed state. Magnetic flux enters the superconductor via the normal conducting cores of the flux lines

Supercooling Temperature
Difference between the glass-transition temperature and the in-use temperature for a glass-based device. For a large (small) supercooling temperature, the structural relaxation rate is low (high)

Superlattice (SL)
Structure of repeated QWs with thin barriers allowing the coupling of wavefunctions from adjacent QWs and the subsequent delocalization of the energy levels

Superlattice Avalanche Photodiode (SL-APD)
Avalanche photodiode architecture where an SL is incorporated in the carrier multiplication region for the purpose of reducing the dark current as well as the excess noise factor

Superparamagnetism
In small ferromagnetic particles the magnetization can flip from one state to the opposite state under thermal excitation. This behavior is similar to a paramagnet. The resistance of the particle to switching is characterized by a thermal stability factor $= \frac{K_u V}{k_B T}$, where K_u is the uniaxial anisotropy factor for the particle, V is the particle volume, k_B is the Boltzmann constant, and T is the absolute temperature

Supersaturation
Ratio between the species concentration and its

solubility limit/equilibrium concentration at a given temperature when this ratio is larger than 1

Superstrate
Describes the thin-film configuration where the glass substrate acts as the window for solar radiation and, therefore, needs a TCO layer before the photovoltaic structure can be deposited

Surface Passivation
Semiconductor surfaces are often electrically active and appear to be covered with a high density of deep-level states. These can greatly affect the properties of a device. Fortunately, it is often possible to treat the surface to substantially reduce this density to values that do not affect device operation. This is surface passivation. In GaAs, passivation is often accomplished by covering the surface with a layer of GaAlAs (see *Passivation*)

Surface-Mount Devices (SMDs)
Electronic components that are attached to the surface of a circuit board as opposed to having through-hole connections. They are a characteristic feature of a hybrid circuit

Susceptor
This is normally made of high-density graphite and is used to transfer the heat from the heater (possibly RF coupling straight into the susceptor) to the substrate

Synthetic Antiferromagnet
Sequence of films: antiferromagnet/ferromagnet/ruthenium/ferromagnet. An example is MnFe/Co/Ru/Co. With a thin Ru film, the final ferromagnetic film is antiferromagnetically coupled to the first ferromagnetic film. The coupling between the two ferromagnetic films is characterized by the exchange coupling parameter J_{12} ($\mathrm{erg/cm^3}$ or $\mathrm{J/m^3}$)

Synthetic Ferrimagnetic Media (SFM)
Magnetic recording media in which there are two ferromagnetic layers of unequal thickness coupled by a thin layer of ruthenium (Ru), sometimes referred to as *Pixie Dust*. The two layers are antiferromagnetically coupled. The coupling between the ferromagnetic films is characterized in terms of the antiferromagnetic coupling between the thinner layer (layer 2) and Ru, and the media flux from that layer by an exchange magnetic field $H_{\mathrm{ex}} = J_{\mathrm{ex},2}/(M_2 t_2)$

T

T Nonlinear Figure of Merit
Figure of merit that describes the applicability of a refractive nonlinear optical material in terms of the Kerr coefficient and two-photon absorption

Tail States
Localized states in the tail of band, i. e., conduction band and valence band, which generally have an exponential density-of-states function

Tandem Cell
Cells connecting more than two unit cells

Tauc Gap
Tauc discovered that in many chalcogenide glasses, such as As_2S_3, optical absorption spectra α around the fundamental edge can be fitted as $\alpha\hbar\omega \propto (\hbar\omega - E_g)^2$, where E_g is called the Tauc gap. The energy is often used as a measure of the optical band gap, while its theoretical interpretation is not conclusive

Technology Node
Minimum half-pitch of metal interconnect is most representative of the process capability enabling high-density (low cost/function) integrated circuits and is selected to define an ITRS technology node. For each node, this defining metal half-pitch is taken from whatever product has the minimum value. Historically, DRAMs have had leadership on metal pitch, but this could potentially shift to another product in the future. Other parameters are also important for characterizing IC technology. For example, in the case of microprocessors (MPUs), physical bottom gate length is most representative of the leading-edge technology level required for maximum performance. Each technology node step represents the creation of significant technology progress in a metal half-pitch – approximately 70% of the preceding node, 50% of two preceding nodes

Temperature Coefficient of Resistance (TCR)
Denotes the sensitivity of a resistor material to changes in temperature. It is usual to quote TCR in terms of ppm per °C

Tensile Strain
Type of strain obtained when a strained Si layer is grown on a relaxed $Si_{1-x}Ge_x$ layer

Terahertz
One terahertz is the frequency of precisely 10^{12}, or one trillion, cycles per second. As such one terahertz corresponds to one thousand gigahertz and to one million megahertz: $1\,\mathrm{THz} = 10^{12}\,\mathrm{Hz} = 10^3\,\mathrm{GHz} = 10^6\,\mathrm{MHz}$. A photon of frequency one terahertz has an energy of $\approx 4.14\,\mathrm{meV}$ and, in free space, a wavelength of exactly $299.792458\,\mu\mathrm{m}$ and spatial frequency of approximately $33.36\,\mathrm{cm^{-1}}$. Physically, therefore, terahertz may be described as meV or *submillimeter* radiation, characterized in *wavenumbers* by tens of $\mathrm{cm^{-1}}$. More loosely, *terahertz* is taken to refer to a range of frequencies around $\approx 1\,\mathrm{Thz}$, a convenient spectral range being the two decades from 0.1 to 10 THz. In this sense, *terahertz* means $10^{12\pm1}$ Hz

Ternary and Quaternary Alloys
These are alloys containing three or four components, respectively. GaAlAs and GaAlAsP are examples

Thermal Budget
Term describing the temperature–time product associated with an IC annealing process step. Material stability and morphology are typically very sensitive to the annealing temperature for a period of time

Thermal Conductivity or Heat Conductivity
Heat flow across a surface per unit area per unit time, divided by the negative of the rate of change of temperature with distance in a direction perpendicular to the surface

Thermal Expansion Coefficient
Fractional change in length or volume of a material for a unit change in temperature

Thermal Nonlinearities
Nonlinear effect associated with heating of the materials by intense light

Thermally Stimulated Current (TSC) Spectroscopy
TSC is a useful technique for assessing concentrations of deep levels in high-resistivity semiconductors. These levels are filled optically with light above the band gap energy. They are emptied sequentially, with the sample under bias, as the sample temperature is raised. Defects of different ionization energies produce separate peaks in the graph of current as a function of temperature

Thermistors
Thermally sensitive resistors that exhibit a change in resistance when the temperature is altered. The most common forms have a negative TCR, meaning that the resistance decreases as the temperature increases

Thermoelectric Figure of Merit
A measure of the efficiency of a material for the conversion of heat to electricity, determined by its Seebeck coefficient and electrical and thermal conductivities

Thermoelectric Power Factor
The "electrical" part of the thermoelectric figure of merit, consisting of the product of the square of the Seebeck coefficient and the electrical conductivity

Thermoelectricity
The collective effects involving the conversion of a temperature difference to electricity and vice versa

Thermomechanical Fatigue
Occurs when materials with different CTE are joined and used in an environment that experiences cyclic temperature fluctuations resulting in imposed cyclic strain, which results in damage to the joined materials

Thermophotovoltaic (TPV)
Absorption of solar radiation and re-emission of infrared radiation that is then converted into electricity by absorption in a narrow-band gap cell

Thermotropic
Liquid crystals are those where the transition from one phase to another occurs on changing the temperature

Thick Film
Layer deposited onto a substrate by the process of screen printing

Threading Dislocations
This term is applied to dislocations that are formed at an interface and that thread their way into the epitaxial layer

Threshold Voltage (V_{th})
Voltage at which an inversion layer forms in the semiconductor substrate of an MIS structure. For applied voltages beyond this threshold, the transistor turns on, i. e., a conducting channel is established between the transistor source and drain, as observed by the increase in drive current

Time Division Multiplexing
A multiplexing technique used in modern optical networks that allows close spacing in time of bits in a single channel

Time-Domain Charge Measurement (TDCM)
TDCM is a rapid technique for the non-contacting measurement of electrical resistivity in SI materials. Its speed and high spatial resolution allows TDCM to be used in a mapping system

Tipping
A *liquid-phase epitaxy* technique in which the melt is moved over the substrate in a tipping furnace

Transferred Electron Effect
This is a somewhat unusual effect where conduction-band electrons can exist in alternative conduction-band states. In GaAs, the lowest conduction-band minimum corresponds to zero electron momentum. The application of an electric field can excite electrons into subsidiary minima of somewhat greater energy from where they will relax back to their original states. If the effective mass of the electrons in the subsidiary minima is greater than that in the primary minimum, negative differential resistance can be realized. This effect can be found in GaAs, InP, and certain alloys

Transformation Optics
A mathematical technique aimed at designing complex electromagnetic media (metamaterials) that allow one to control the flow of light in a desired manner. The transformation optics approach enables the recreation of an electromagnetic phenomenon in a warping virtual space as an optical illusion in the physical Euclidian space

Transistor
A three-terminal device in which the current flow between two terminals (called the source and drain regions) is controlled by the voltage applied between a third terminal (the gate electrode) and one of the two terminals

Transit Time
Refers to the difference in time between the moment a charge carrier is generated at one end of the sample and its arrival at the other end. It is the primary result in time-of-flight experiments. In xerography, the transit time is the time for a photoinjected carrier to traverse the charge-transfer layer

Transparent Conducting Oxide (TCO)
Thin layer of highly conducting material that is used as the front contact in thin-film solar cells. The requirement of high optical transmission over a wide spectral range is important to allow as much sunlight through as possible to the absorber layer

Transverse Magnetic Bias Film
In a magnetoresistive head, an antiferromagnetic film coupled to the pinned film to maintain the magnetization of the pinned film in a direction transverse to the magnetization of the recording film

Trapping Centers
Irregular sites in the photoconductor with localized

electronic energy levels inside the band gap. Such sites will trap charge carriers and thus inhibit electrical conduction

Traveling Heater Method
A molten zone is made to migrate through a homogeneous solid source material

Trimming
Process by which the value of thick-film elements can be adjusted. Usually achieved by using a laser or air abrasive jet

Tunneling Conduction
Process for charge conduction where the charge carriers pass through an energy barrier by quantum-mechanical tunneling

Tunneling Magnetoresistance (TMR)
With two ferromagnetic films coupled by a thin insulating layer, electrons can tunnel through the insulating layer and the magnetoresistance coefficient is given in terms of the polarization ratio for the two contacts as $\frac{\Delta R}{R} = \frac{2P_1P_2}{1-P_1P_2}$.

Twin Crystal
Crystal having two or more crystals or crystal sections that, when regularly positioned, are in reverse position relative to the other sections

Two-Photon Absorption
Nonlinear process in which a simultaneous absorption of two photons results in an electronic transition

Type I Band Alignment
Electrons and holes are confined within the same layer

Type II Band Alignment
Electrons and holes are confined in different (adjacent) layers

U

Ultrafast Nonlinear Response
Nonlinear response taking place in the non-absorbing spectral region

Ultra-High Vacuum (UHV)
A vacuum better than 10^{-9} Torr

Underfill
Dielectric composite organic material that is bonded between the chip and substrate of a flip-chip device to help mechanically interlock the chip to the substrate. The underfill material is typically a silica-filled anhydride resin polymer

Uniaxial Anisotropy
Magnetic anisotropy along one direction in a ferromagnetic material and characterized by a uniaxial anisotropy parameter K_u (erg/cm^3 or J/m^3)

Uniform Planar Alignment
of the director is when the director is parallel to the surface and to a particular direction in the surface. This has also been referred to as homogeneous alignment

Unipolar Avalanche Photodiode (UAPD)
QWIP detector employing avalanche multiplication of only one type of carrier via intra-QW impact ionization by carrier–carrier scattering

Upper Critical Field (B_{c2})
The highest magnetic field allowing the existence of the superconducting mixed state in the type II superconductor considered

V

Vacancy
Regular lattice site from which the host atom is missing. Anion (cation) vacancies are generally donors (acceptors)

Valence Band
Highest range or band of energies in a semiconductor where electrons are normally present at zero temperature. When electrons are promoted from this band, holes are left behind that contribute to the electrical conductivity

Vapor Growth Techniques
Growth takes place via a vapor phase: particularly useful for high-melting-point materials, or those with high partial pressures of one constituent, e.g. ZnS

Vapor-Phase Epitaxy (VPE)
A form of epitaxy where the components of the layer to be grown are transported to the substrate as a vapor. Decomposition of these components to produce the layer occurs because the substrate is heated, often on a support called a susceptor

Vegard's Law
States that the lattice parameter of an alloy material is given by the concentration weighted average of the constituents

Vehicle
Liquid component of the thick-film paste. Typically contains a resin dissolved in a solvent. It is removed during the drying and firing processes

Verneuil Technique
Rapid growth method used for many high-melting-point materials, mainly oxides

Vertical Superlattice Structure
An ideal nanostructure which enables the efficient dissociation of photogenerated excitons at the donor/acceptor interfaces within the exciton diffusion length and the transport of electrons and holes to the respective electrodes

Vertical Transport
Mechanisms of carrier transport parallel to the growth direction in a semiconductor quantum-confined structure

Vertical-Cavity Surface-Emitting Laser (VCSEL)
QW diode laser emitting through its top semiconductor surface

Vertical-Gradient Freeze Technique
Similar to the Bridgman method but freezing is controlled by moving a temperature gradient along a stationary crucible

V/I Boundary
Denotes the spatial location of the transition from the vacancy-dominated region to the Si interstitial dominated region and vice versa

VI Recombination
Mutual annihilation between vacancies and Si interstitials

Virtual Gap States
The virtual gap states are the *solutions* of Schrödinger's equation for complex wave vectors in the energy gap. The continuum of these *virtual* states is the source of *real* impurity states in the bulk, as well as of surface states and the interface-induced gap states at *real* surfaces and interfaces, respectively, provided the corresponding boundary conditions are considered

Void
See *Crystal-Originated Particle (COP)*

W

W Nonlinear Figure of Merit
Figure of merit that describes the applicability of a refractive nonlinear optical material in terms of nonlinear index change and linear absorption

Wannier Exciton
Weakly bound electron–hole pairs

Wavelength Division Multiplexing
A multiplexing technique used in modern optical networks that involves sending many signals in parallel at closely spaced wavelengths along the same fiber

Weak Links
Large-angle grain boundaries in cuprate superconductors, which act as barriers for the supercurrents

Web Photoreceptors
Devices where the layers are coated on an insulating polymeric support and subsequently fashioned into a loop

Wire Bonding
Process where a thin wire (can be less than 25 μm) made of Au or Al is bonded to the surface of an integrated circuit and then to a pad or a leadframe in the package

Work Function
The energy (usually measured in electron volts) needed to remove an electron from the Fermi level in a solid to outside the surface

Write Efficiency
Ratio of the magnetic field times the write gap to the magnetomotive force (turns times current) in the write coil

Write Gap
Region in a write head that generates the magnetic fields that couple to the recording medium. Characterized by the gap length g

Write Head
In magnetic recording, the ferromagnetic device used to generate magnetic fields from current that can switch the state of the magnetization in the recording medium. In disk drives, the ferromagnetic device is made using thin films

Wrong Bond
Homo-polar bonds in stoichiometric glasses, such as As–As in As_2S_3. The defective bond nominally does not exist in the corresponding crystal. However, specifically in covalent glasses such as As_2S_3, the bond exists with a concentration of $\approx 1\%$, which depends upon preparation methods, and so forth

Wurtzite Structure
Comprises two interpenetrating close-packed hexagonal lattices, one for cations and the other for anions. Each anion (cation) has four cation (anion) nearest neighbors. In principle, diffusion should be anisotropic but the meagre evidence available indicates only slight effects

X

Xerographic Discharge
Time-dependent decrease in the surface potential with exposure of a charged OPC

Xerographic Gain
or quantum efficiency of supply is the fractional number of surface charges neutralized per absorbed photon

Xerography
Name coined by Chester Carlson for electrophotography using dry powder marking particles

X-Ray
The form of energetic electromagnetic radiation of the wavelength, ≈ 0.1 nm

X-Ray Diffraction
A highly sensitive technique for measuring the lattice constant of crystalline solids

X-Ray Photoelectron Spectroscopy (XPS)
A method in which an electron spectrometer is used to measure the energy distribution of photoelectrons and Auger electrons emitted from a surface irradiated by X-ray photons

X-Ray Sensitivity
The X-ray sensitivity of a photoconductive detector is the collected charge per unit area per unit exposure of radiation

X-Ray Topography
A method for sampling the diffraction condition of X-rays from a surface in order to observe changes in lattice constant. The X-rays are scanned over the surface and an image is built up from changes in the fraction diffracted at a particular angle. Changes in lattice constant resulting from strain or changes in composition can be imaged in this way

Y

Yield Strength
Applied stress (in pounds per square inch or psi in the English system, megapascals or MPa in the metric system), under which an object experiences plastic deformation

Young's Modulus
Ratio of a simple tension stress applied to a material to the resulting strain parallel to the tension. Therefore, Young's modulus has units of pressure

Z

Zincblende Structure
A crystal structure that has two interpenetrating face-centered cubic (fcc) lattices, one for the cations and the other for anions. Each anion (cation) has four anion (cation) nearest neighbors. Diffusion is isotropic

Zone Refining
Technique used to repeatedly pass zones of molten material through a solid bar in order to purify it, either for use directly in applications or to produce pure starting materials, e.g., elements for compound semiconductors

Subject Index

π band 1461
π^* band 1461
π bond 1461
π-electron delocalization 1299
π–π stacking 1335
2-D electron gas (2DEG) 40
2DEG electronic 104
2DEG heterostructure 40

A

Abbe number 79, 81, 1084, 1461
abrasive trimming 715
absolute constant photocurrent method (ACPM) 157
absorptance 159
absorption 74, 900
– rate 901
absorption coefficient 222, 1461
– amorphous semiconductor 562
– direct 684
– effective 1083
– indirect 684
– Tauc 684
– Urbach 685
AC conductivity 221
AC Josephson effect 1232
accelerated crucible rotation technique (ACRT) 274, 343, 346, 1461
accelerometer 718
acceptor 1461
– -bound exciton (ABE) 769
– concentration 1461
– dopant 137
– doped material 257
– doped oxide defect diagram 257
– doping 359
– impurity 250
– level 1461
– molecule 1335
AC-conductivity, relationship to imaginary susceptibility 221
accumulation 1461
acoustic-phonon scattering 507
ACRT/THM 276
activation energy 198, 252, 446, 1117
active
– component 714
– material 1461
– region 1333
active matrix
– addressing 952, 1461
– array (AMA) 1125
– flat-panel imager (AMFPI) 1125, 1126
– organic light-emitting diode (AMOLED) 1111
actuator 718, 1170, 1207
Adam-Gibbs (AG) 441
adhesive
– interconnects, rework 1326
– interlayer 981
– isotropic 1326
advanced Compton telescope (ACT) 863
Advanced Storage Technology Consortium (ASTC) 1220
aerogel 640
Ag 1081
Aharonov–Bohm effect 1167
air mass (AM) number 1461
air-bearing 1186
Al_2O_3 635
AlGaAs 7
AlGaN 7, 11
AlInGaP 7
alkali halide 273
all-optical switch 1461
alloy
– composition interpolation scheme 726
– disorder scattering 41
– group III–V 729
– group I–VII 729
– semiconductor optical spectra 738
AlN (aluminum nitride) 829, 844
– electron transport 841
– free-exciton energy 780
– material parameter 833
– mechanical property 749
– optical phonon frequency 750
– phonon dispersion spectrum 750
– thick film technology 710
alternating current (AC) 196, 221, 490, 612, 660
alumina
– materials 710
– self-ordered porous 1021
– thick film substrate 710
aluminum nitride 764
ambipolar diffusion length 160
ammonia sensor 1283
amorphous
– -crystalline boundary 1150
– GST 1156
– magnet 86
– metal oxides 1461
– network 1461
– organic semiconductor 1117
– oxide semiconductor 1117
– phase GST alloy 1152
– polymer 434
– semiconducting film 1111
– silicon (a-Si:H) 558
– silicon-germanium alloy 163
– -to-crystalline transition 674
amorphous film 670, 674
– electronic properties 1112
– metallic 675
amorphous selenium
– avalanche detector structure 1131
– carrier multiplication 1131
– effective hole drift mobility 1132
– hole impact ionization coefficient 1132
– impact ionization coefficient 1131
– multilayer pin-type detector 1130
amorphous semiconductor 59, 151, 557, 970, 1086, 1118, 1461
– absorption coefficient 561
– conduction band 558
– electrical conductivity 565
– electrical properties 565
– electronic state 558
– Hall effect 565
– light-induced phenomena 567

– optical properties 561
– photoluminescence 561
– photoluminescence spectra 64
– stationary photoconductivity 210
– structural properties 560
– thermoelectric power 565
– valence band 558
amorphous silicon (a-Si:H) 155, 578, 1102, 1334, 1462
– alloy 158
– dangling-bond-defect 573
– dangling-bond-defect density 580
– device applications 582
– device-grade 574
– drift mobility 566
– electrical conductivity 566
– electronic material properties 1112
– electronic transport properties 1112
– growth 574, 577
– growth process 573
– hydrogenated 573
– optical properties 563
– sputtered 1118
– structural model 558
– TFT 1111
amperometric sensor 259
– charge transfer 1291
– sensitivity 1291
amphiphilic 1462
amphoteric dopants 1462
amphotericity 1462
amplifier, optical 898
amplitude reflectance 683
AMR effect 1193
analytical electron microscope (AEM) 385, 1462
Anderson model 197
Anderson's criteria 1113
angle of deviation 75
angular dispersive x-ray diffraction (ADX) 748
anharmonicity 1388
anion 1462
anisotropic
– conductivity 1327
– electrically conductive adhesives (AECA) 1326
– magnetoresistance (AMR) 98, 1192
anisotropy
– axis 89
– biaxial 95
– dielectric 941
– parameter 1239
– uniaxial 95
annealing 139, 632, 1462
anodic oxide film 663
anodisation 663
anomalous Hall effect (AHE) 105
anti-curl layer 981
anti-ferroelectric 229, 595
antiferromagnetic
– film pinning 1194
– insulator 1235
– spin configuration 1198
antiferromagnetically coupled (AFC) 1202, 1203
– media 1462
antiferromagnetism 96
antimonide 282
anti-reflection coating 680, 683, 684
anti-site defect 1462
apparent band gap narrowing 529, 1462
application magnetic 107
areal density 1191, 1462
array package 1314
– physical property of material 1315
array via-hole type structure 888
arrayed-waveguide grating (AWG) 1068
Arrhenius
– equation 1061
– relation 134, 1462
– temperature dependence, inorganic glass 254
Arrhenius–Néel model 1202
arsenic selenide 163
arsenic-based material 314
artificial
– design 1335
– magnetic conductor 1367
– magnetism 1354
As_2Te_3 164
as-deposited film 676
a-Se avalanche detector 1131
a-SiN_x:H film 1112
aspect ratio 1020
association energy 249
astigmatism 1462
astroid 1216
A-swirl 112, 1462
asymmetric transmission 1373
asymmetric-AC sputtering 660
atom diffusion 133
atomic
– force microscopy (AFM) 171, 317, 458, 646, 1178, 1270, 1288
– layer deposition (ALD) 6, 634
– mixing 415, 423
– vibration 431
atomic layer epitaxy (ALE) 310, 365, 371
– growth rate 375
atomically flat surfaces 318
attempt frequency 1202
attempt-to-escape frequency 162
attenuated total reflection (ATR) 579, 1462
audio–video equipment 97
Auger
– depth profiling 418
– electron 414
– hardware 415
– instrument calibration 414
– map 415
– quantitative analysis 414
– spectrum 414
Auger electron spectroscopy (AES) 413, 1462
– surface sensitivity 414
Auger recombination 876, 911
– nonradiative 910
Aurivillius crystal structure 597
auto exhaust sensor 258
autocatalytic (AC) model 445
auto-compensation 1462
avalanche
– breakdown 467, 929
– generation 514
– photodetector (APD) 928
avalanche photodiode (APD) 353, 1050
– unipolar 1052
average
– grain size 32
– hopping time 214
– positron lifetime 790
AV-MRAM 108
Avogadro number 728, 1259
Avrami exponent 443
azo compound
– electrophotographic 986
– photosensitivity 986
– pigment 986

B

B_2I complex 126
Babinet principle 1366
back end of line (BEOL) 639, 1462
background-limited
– detectors 870
– device performance 875
– performance (BLIP) 1463
backscattered electron (BSE) 389, 1463
Baldereschi concept 1463
ballistic
– conductance 1167
– electron 38
– electron emission 834
– SWNT 1167
– transport 1169
ballistic-electron-emission microscopy (BEEM) 176, 179, 1463
band
– alignment 1463
– diagram 624
– filling 1463
– mobility 1114
– offset 623, 1463
– structure 1099, 1102
– tail states 208
band gap 486, 632, 634, 1463
– bowing 929
– engineering 524, 1037, 1463
– narrowing 58
– structure, graded 137
band gap energy 53, 726
– cubic group III–V ternaries 735
– group III–V binaries 733, 735
– group III–V quaternaries 735
– group III–V ternaries 733
– temperature-insensitive 543
band-edge modulated films 569
band-pass filter 684
band-splitting 905
band-stop filter 684
band-structure, MUTIS structure 188
bandtailing 58
band-to-band
– absorption 56, 65
– tunneling 877, 927
bandwidth 875
barium strontium titanate (BST) 637
Barkhausen noise 90
barrier energy 233
barrier height 1463
– BEEM 179
– effective 177
– flat-band 179
– IPEYS 180
– slope parameters 182
basal stacking fault (BSF) 776
battery, high energy density 261
BCS theory 1230
beam
– deflection 1119
– effective pressure ratio 1463
– quality 919
Beer–Lambert law 222
Beilby overlayer 72
bend deformation 942
bend elastic constant 943
Bernal stacked graphene 1177
beryllia, thick film technology 710
BI complex 126
biaxial elastic modulus 1112
BICUVOX 261
biexciton 779
bilayer graphene (BLG) 1177
bi-layered metamolecule 1359
BIMEVOX 254
bimolecular recombination 978
binary
– alloy, solid solution 26
– endpoint 731
– parameters 726
bioluminescence 997
bio-sensors, nanotube based 1170
bipolar
– junction transistor (BJT) 474, 497, 1178
– magnetic junction 104
– transistor 473, 523
biquadratic exchange 86, 99
birefringence 948, 960
birefringent crystals 1463
bisazo compound 986
bis-polycarbonate (Lexan) 211
bistable switching 1271
bit aspect ratio (BAR) 1215
bit error rate (BER) 1213
bit-patterned 1185
black-body distribution 1098
black-body emission 997
blended junction 1330, 1333
Bloch oscillation 1046
Bloch wave 20
blocking layer 982
blocking temperature 1195, 1463
blue LED 325
Bode plot 243
Bohr magneton 88
Boling's relation 1084
bolometer 1345
Boltzman–Matano analysis 140
Boltzmann
– constant 57
– equation (BE) 1463
– transport equation 29, 37, 491, 830, 831
bombarding particles primary 423
bond
– angle 625
– arrangement 627
– energy hierarchy 1155
– failure 1315
– length 625
– switching 568
– variation 627
borate 289
boro-phosphosilicate glass (BPSG) 1069
borosilicate crown 80
Bose–Einstein distribution 57, 773
bottom shield 1214
boule 1463
bound exciton (BE) 769
– GaN 774
boundary effect 456
bovine serum albumin (BSA) 1296
bowing parameter 733
– group III–V ternaries 733
Bragg
– condition 914
– grating 1014
– mirrors 915
Bragg's law 1463
branch point 176
branch-point energy 182
Bravais lattice 1463
break junction technique 1273
breakdown 624
Bremsstrahlung 1463
Brewster angle 71
Bridgman 273, 379
– crystal 347
– growth 286, 378, 1464
– method 274
– process 285, 345
– technique 281, 283
Brillouin scattering 1087

Brillouin zone (BZ) 485, 1001, 1014, 1464
broadband sensitizer 1464
broken bonds 1466
Brooks–Herring (BH) formula 1464
Brownmillalite 1408
Brunauer, Emmett and Teller (BET) 1298
Bruno quantum interference model 100
B-swirl 115
bubble technology 98
buckministerfullerene 1263
bucky materials 1170
buffer chamber 358
bulk
– concentration 255
– crystal growth 269, 280
– crystal growth, vapor phase 375
– crystal structure 669
– defects 474, 477
– electron 685
– film, dielectric constant 620
– growth 344
– heterojunction 1333
– heterostructure device 922
– hole concentration 462
– -limited conduction 694, 696
– material conductivity 685
– modulus 730, 1464
– nanostructured thermoelectric 1387
– resistivity 454, 455
– SiO_2 627
– SiO_2, point defect 628
– space charge 976
bulk semiconductor
– electron transport 831
– spreading resistance 458
bulk wurtzite GaN
– band structure 834
– material parameter 833
– valley occupancy 836
– velocity-field characteristic 835
buried heterostructure (BH) laser 908
buried junction 883
Burstein–Moss shift 781
burst-illumination LIDAR (BIL) 891
BWO 1343

C

C_{60} 1164
cadmium chalcogenide 671
cadmium compound film 690
– morphological characteristics 678
cadmium zinc telluride 1133
calamitic 1464
calcination 602
canonical ensemble
– configuration entropy fluctuations 231
– thermal fluctuations 231
cap layer growth 349
capacitance
– charge storage 219
– definition 220
– equivalent thickness (CET) 623, 1464
– flatband 468
– method 466, 468
– silicon surface 468
– transducer 1295
capacitance–voltage (C–V) 1464
– –frequency (C–V–f) characteristic 1114
– measurement 464
capacitor dielectric 594, 619, 636
– DRAM 637
– ferroelectric material 638
– non-volatile memory applications 639
– scaling 636
– thick film 713
capture coefficient 162
carbon 1173
– doping 126
– monoxide (CO) 1282
– overcoat 1191, 1199
carbon nanotube
– catalysts 1165
– dimensionality 1164
– electronic structure 1165
– field-emission display 1169
– interconnect 1169
– semiconductor 1165
– symetries 1165
Carlson, Chester F. 968
Carnot efficiency 1227
carpet cloak 1362
carrier
– capture 1046
– concentration 691, 1392, 1464
– concentration, direct determination 463
– confinement 907
– density 248, 1114
– depletion current 1137
– distribution function 491
– drift mobility 564
– drift mobility calculation 214
– effective mass 734
– leakage 912
– lifetime 152, 351, 877
– range 1464
– recombination 910
– relaxation 1047
– scattering 739
– temperature 37
– transport 151, 507
– trap 152
– tunneling 460
– velocity, microscopic 1464
carrier mobility 152, 622, 734, 1464
– diffusion picture 213
– effective 1116
– modulation 1118
– nanotube 1164
– time-dependent 213
carrier–carrier scattering 493
carrier–lattice scattering 492
cascaded second-order materials 1464
cast recrystallize anneal (CRA) 283, 345
catalyst 279
catastrophic breakdown 631
catastrophic optical damage (COD) 918
cathodoluminescence (CL) 386, 775, 997, 1464
cathodoluminescence/correlated TEM investigation 408
cation 1464
– self-diffusion 145
cavitand compound 1303
Cd_3As_2
– morphological characteristics 677
Cd-based compound semiconductor 1464
CdHgTe growth 317
Cd-rich film 653
CdS 1081
– film 676, 698
CdSe film 692, 693
CdTe:Sn 168

CdTe-based compound 316
central processing unit (CPU) 1186
centro-symmetric 1464
ceramic 98, 1016
– actuator, piezoelectric 608
– capacitor 595
– fabrication 601
– laser ablation 606
– materials, thick film technology 710
– q-DC behaviour 241
cermet (ceramic/metallics) 708, 1464
cermet thick film 712
– resistors 718
chalcogenide glass 155, 557, 561, 1088, 1464
– band gap 564
– dark conductivity 564, 567
– drift mobility 564
– optical property 564
– PL spectra 564
chalcogenide-based superlattice 1160
characteristic temperature 913
characterization techniques 454
characterizing functional activity 403
charge
– area development 969
– collection 152
– density difference (CDD) 1152
– pumping (CP) 469, 472
– storage, real susceptibility 221
– transfer (CT) exciton 1332
– transfer complex 977, 1166
– trapping 972
charge carrier 1234
– drift mobility 213
– effective mass 1239
– mobility 196, 198
– relaxation kinetic 213
charge generation 977
– amorphous material 978
– bisbenzimidazole perylene 985
– CT complex 985
– fluorenone bisazo 985
– impact ionization 513
– intrinsic 987
– layer (CGL) 981, 983
– Onsager model 978
– PVK–TNF 985
– sensitized 987
– thiapyrylium dye 985
– titanyl phthalocyanine (TiOPc) 985
– trisazo pigment 985
– UFSq 985
charge generation material
– aggregate 986
– chlorodiane blue 986
– fluorenone bisazo pigment 986
– hydroxygallium phthalocyanine (HOGaPc) 987
– metal-free phthalocyanine (H_2Pc) 987
– perylene diimide 986
– squarylium pigment 988
– titanyl phthalocyanine (TiOPc) 987
charge transport 193, 211, 972
– disordered materials 193
– limited range 219
– localized states 203
– phenomena 513
charge transport layer (CTL) 970, 981, 982, 988, 991
– electron transport 988
– glass-transition temperature 988
– inhomogeneous 982
– organic photoreceptor 988
charge-blocking layer 991
– electron transport 983
charge-coupled device (CCD) 395, 454, 863, 876, 1005
charge-generation layer (CGL) 970, 982, 983, 985, 991
– bisazo compound 986
– material 984
– perylene 986
– photoreceptor 983
– phthalocyanine 987
– trisazo compound 986
chemical
– annealing 578
– bath deposition (CBD) 1105
– beam epitaxy (CBE) 334
– diffusion 136
– mechanical polishing (CMP) 640
– self-diffusion 136, 145
– sensing 1282
– shift 416
– solution deposition (CSD) 605
– transformation 441
– vapor deposition (CVD) 139, 156, 295, 372, 404, 523, 573, 636, 646, 663, 856, 1024, 1062, 1102, 1165, 1175, 1285, 1464
– vapor transport (CVT) 284, 365
chemical sensor
– polymer 1297
– thick film 719
chemically assisted ion beam etching (CAIBE) 405
chemiluminescence 997
chip scale optic (CSO) 13
chip scale packaging (CSP) 13
chiral
– dopant 937
– metamaterial 1353, 1358
– metamolecule 1359
– nanotube diameter 1165
– nematics 937, 1465
– smectic C phase 938
chirality 107, 937, 1358
chromatic aberration 1465
circuit
– element interconnection 617
– fanout 618
– inverter 618
– NAND gate 618
– response delay time 617
– semiconductor 462
circular dichroism 1358
cleanliness 329
cleave and stain 1465
close-space sublimation (CSS) 855, 1105
closure domain 93
cluster 138
CMOS
– device 1267
– gate stack 622
– soft error 1321
– technology 617, 1321
coating
– anti-reflection 680, 683
– conformal 717
Coats–Redfern–Sestak plot 446
coclose-spaced sublimation (CCSS) 858
Co-Cr-Pt alloy 1208
code rate (CR) 1213
codeposited layer 1333
codoping 788
– with nitrogen 123
coefficient of thermal expansion (CTE) 1112, 1315, 1465
coercive field 90, 98, 591, 1192, 1465

coercive squareness parameter (S^*) 1465
coherence length 1233, 1239, 1465
– cuprate superconductor 1240
– in-plane 1239
– iron-based superconductor 1252
– MgB_2 1247
– out-of-plane 1239
cohesive energy 1465
Cole–Cole function 225
collinear probe 457
colloidal quantum dot 1054
color glass 1089
columnar
– grain structure 676
– phase 1465
– phase, rectangular 939
commercial crystal 280
compact fluorescent lamps (CFL) 1006
comparison of LPE techniques 350
complementary characterization
– electronic material 406
– optolectronic material 406
complementary metal-oxide-semiconductor (CMOS) 6, 454, 523, 616, 620, 876, 1164, 1267, 1284, 1321, 1345, 1465
complex 1465
– Fermi wavevector 99
– heat capacity 447
– perovskite 595
compliance constant 729
composite material, ionic conductivity 255
compositional interdiffusion (CID) 137, 144
compositional uniformity 347, 348, 351, 1465
compound annual reduction rate (CARR) 616
compound semiconductor 7, 665, 1465
– crystal properties 367
– MOCVD technique 373
– quality of epilayers 373
compressive strain 142, 904, 1465
compressive stress 120
Compton scattering 387, 1465
computer modeling 153
computing power 616
concentration quenching 1074, 1465
conditional glass former 1065
conducting polymer 1281, 1297
– conductivity 1298
conduction 42
– electron 435, 436
– mean free path 21
– metallic-type 1113
– quantized 42
– tail states 59
conduction band (CB) 56, 900, 1083, 1465
– amorphous semiconductor 558
– deformation 737
– density of states 558
– effective mass 488
– potential energy 199
conduction mechanism
– bulk-limited 690
– electrode-limited 690
conduction–hole–conduction–conduction (CHCC) 910
conduction–hole–spin–hole (CHSH) 911
conductive layer 982, 1465
conductivity 19, 1391, 1465
– anisotropic 1327
– electrical 566, 686
– film thickness 686
– ionic 253
– ionic-electronic 247
– layer-by-layer 1167
– mixture rule 34
– oxygen ion 255
– relaxation time 20
– temperature dependence 208
conductivity mass 740, 1465
– group III–V binary 736
conductor 1326
– ionic 248
– thick film 604
configuration coordinate model 1465
configuration entropy fluctuation 237
conformal coating 717
conjugated polymer 211, 974
constant current (CC) 1298
constant photocurrent method (CPM) 157, 559
constant ratio (CR) cycle 442
contact resistance 459
– direct measurement 460
– model 460
– ratio 188
– semiconductor 459
container-free LPE (CFLPE) 313
continuous Czochralski method (CCZ) 305
continuous wave (CW) 799, 898, 1073
continuous-charging Czochralski (CCZ) 305
continuous-granular-composite 1214
continuum theory 943
– chiral nematic 944
convection control 272
conventional DSC 440
– total heat flow 447
conventional TEM (CTEM) 397
convergent beam electron diffraction (CBED) 396
converse piezoelectric effect 1465
conversion efficiency 1380, 1466
Cooper pair 1229, 1466
– coherence length 1233
– collective wave function 1231
– density 1233, 1244
coordination number 1407
copper thick film 711
core multishell (CMS) 800
core-level photoemission spectroscopy (PES) 180
core-level shift, adatom induced 181
corona charging 970, 982
corona discharge 1466
corona-discharge poling 1088
corotron 969
correlation factor (f) 1466
corrosion propensity 1195
cost per transistor 616
Cotterell atmosphere 1466
Coulomb
– blockade 1167
– gap 204
– interaction 62, 99
coupled granular continuous (CGC) 1212
– recording media 1466
coupling constant 1196
coupling field 1196
coverage ratio 1286
criteria for precursors 353
critical
– angle 75
– current 1230

– current definition 1241
– dopant concentration 127
– field 1466
– field, metallic superconductor 1229
– fluctuations 1466
– magnetic field 1208
– thickness 1466
critical (transition) temperature 1466
critical current density 1241
– anisotropy 1241
– iron-based superconductor 1250
– MgB_2 film 1246
– polycrystalline cuprate superconductor 1242
critical temperature 1226
– dependence on hole concentration 1238
– epitaxial strain 1245
– iron-based superconductor 1226, 1248, 1251
– metal 1226
– molecular superconductor 1226
– oxide 1226
– ultrathin film 1244
cross Kelvin resistor (CKR) 460
cross-luminescence 1010
crosstalk 884
– long range 881
cross-track 1211
cryopanel 330, 332, 1466
crystal
– clear project 1010
– Debye heat capacity 426
– density 727
– field splitting 999
– in R&D 280
– ionic 51
– key parameters 440
– neckingless method 306
– originated particle (COP) 115, 1466
– phonon concentration 427
– properties 367
– structure 669
crystal growth 270
– hydrothermal method 377
– piezoelectric 601
– striations 302
– transport agent 375
crystal structure
– iron-based superconductor 1248
– MgB_2 1246
crystalline enthalpy 439
crystalline material
– non-centrosymmetric 589
– properties 589
crystalline polymer
– thermal conductivity 434
crystalline semiconductor 62, 565
– Hall effect 565
– optical absorption 62
– thermoelectric power 565
crystalline silicon 562, 1100
– intrinsic mobility 501
– phonon modes 492
– room-temperature mobility 501
– structural model 558
crystallization
– kinetics 443
– temperature 425
c-Si 560
μc-Si:H
– dangling-bond-defect 573
– dangling-bond-defect density 580, 582
– device applications 582
– device-grade 574
– epitaxial-like crystal growth 579
– film 579
– formation 577
– growth 573, 577, 579
CT exciton 1332
cubic anisotropy 95
cubic III–V binary 737, 738
– compliance constant 729
– elastic stiffness 729
cubic III–V ternary 735
cubic perovskite 1406
$CuGaSe_2$ 164
CuO_2 planes 1234
cuprate superconductor 1234, 1249, 1250
– carrier concentration 1242
– characteristic length scales 1239
– common features 1234
– critical current density 1241
– critical temperature 1238
– crystal structure 1236
– CuO_2 planes 1235
– doping 1234
– energy gap 1236
– film 1245
– grain boundaries 1243
– hole concentration 1238
– in-plane strain 1245
– lattice parameters 1237
– pseudogap 1237
– upper critical fields 1239
Curie temperature 90, 229, 591, 1188
Curie Weiss temperature 592
current confinement 908
current noise measurement, high-impedance devices 477
current perpendicular to the plane (CPP) 1466
current-in-plane (CIP) 1192, 1466
current-perpendicular-to-plane (CPP) 1192
current–voltage (C–V)
– characteristic 1099, 1231
– measurement 470
– measurement, low-frequency 469
Czochralski (CZ) 112, 271, 297, 495
– growth with an applied magnetic field (MCZ) 304
– method 297, 495
– pulling procedure 299
– technique 4
Czochralski (CZ) crystal
– growth condition 301
– seed-end portion 301
Czochralski (CZ) silicon 301
– carbon 303
– crystal 300
– crystal growth sequence 300
– crystal impurity 301
– dislocation-free 305
– doping 303
– impurity 303
– oxygen 303
– property 299
– striation 302
– swirl 302
Czochralski and liquid encapsulated Czochralski (LEC) growth 1466

D

D–A heterojunction 1330
damping relaxation time 225
dangling bond 580, 628, 1466
– a-Si:H 580
– defect 573, 580
– light-induced creation 568
DAP transition 777
dark conductivity 210, 567

dark current 886
– activation 152
– in x-ray detectors 1136, 1139
dark decay 979, 1466
– characteristic 979
– photoreceptor 979
– rate 971
data rate 1466
Davydov splitting 1288
DBR laser 914
DC conductivity temperature dependence 198
DC hopping 696
DC sputtering 606
de Broglie wavelength 1039
dead region 1333
Debye
– frequency 426, 1231
– heat capacity 426
– length 1283, 1466
– molar heat capacity 427
– theory 425, 427
Debye response 225, 239
– exponential decay 225
Debye temperature 24, 427, 428, 755, 1383
– III–V binaries 731
– phonon concentration 433
decay rate 1202
decoration 137
deep localized states 209
deep trap 1129
deep-energy-level impurities 1466
deep-level transient spectroscopy (DLTS) 469, 559, 789, 1114, 1466
– averaging techniques 478
– electrical characterization 477
defect 624
– chemistry 1467
– equilibria 257
– free silicon 119
– free zone 116
– hierarchical structuring 1387
– interaction 255
– orientation 1089
defective
– solid 254
deformation potential 736, 1467
– parameters 1467
deformational phonon 491
degenerate semiconductor 1393
degradation in overwrite 1204
DEH 986
delocalised states 196
demagnetization energy 96
demagnetizing factor 1210
demagnetizing field 1202, 1210
dendritic structure 345
Dennard scaling 618
dense wavelength-division multiplexing (DWDM) 915
density functional theory (DFT) 88, 114, 190, 1389
density of states (DOS) 25, 42, 59, 61, 152, 155, 195, 435, 486, 735, 736, 1113, 1114, 1137, 1176, 1197, 1467
– conduction band tail 1114
– mass 1467
– normal state 1230
– single-electron 1230
– vibrational 426
density of tail states 1114
density size distribution of voids 117
density-functional theory (DFT) 1467
depletion 1467
– effect 623
– region 691, 877
depletion layer 465, 1243
– doping profile 466
deposition
– parameter 674
– rate 653
deposition method 645
– chemical deposition 662
– physical deposition 647
– thin film 647
– vacuum evaporation 648
depth
– of field 1467
– of focus 1467
– profiling 1467
– resolution 1467
designer metasurface 1370
destabilisation 1156
detailed balance 162
detection efficiency 1467
detection phase sensitive 1005
detective quantum efficiency (DQE) 1142, 1467
detectivity 1467
detector 897
– background-limited performance 874
– cooldown time 874
– diffusion-limited 357
– indium antimonide (InSb) 868
– internal gain 928
– materials engineering (DME) 347
– multiple quantum well (MQW) 869
– photoconductive 868, 871, 874
– polycrystalline PbS 868
– white noise current 874
– zero bias resistance 874
deuterated triglycerine sulfate (DTGS) 612
device
– characteristics 615
– degradation 138
– dimensions 617
– optical 1086
– performance 573
– photovoltaic 720, 874
– status 352
– structure 138, 882
– technology 133
– under test (DUT) 474
device-grade a-Si:H 573
device-grade μc-Si:H 573
devitrification 1065, 1467
DFT (density-functional theory) 88, 114, 190, 1389, 1467
– simulation 1154
diamagnetic 97
diamagnetism 96, 97, 1354
diameter control 272
diamond structure 1467
diamond thin film 156
diamond-like lattice structure 524
dibit 1201
dichlorosilane (SiH_2Cl_2) 295
Dieke diagram 999
dielectric 601, 615
– band gap 182
– capacitor 594, 619, 636
– contrast 1016
– DC conductivity 240
– dispersion 225
– film 675
– loss 231
– materials properties 619
– modulus 243
– paste thick film 713
– permittivity 74, 636, 941, 949
– permittivity relative 619
– quantum mechanical tunneling 619
– relaxation 234

– reliability 631
– response function 226
– scaling 618
– spacer layer 636
– susceptibility 221
– thick film 604
– thickness 636
dielectric constant 48, 491, 617
– conductivity contribution 493
– cubic III–V binaries 738
– effective 32
– optical 737
– static 501
– substrate 710
– wurtzite III–V binaries 738
dielectric constant/susceptibility (DC/DS) 1467
dielectric increment 231, 235
– correlation length 236
– dynamic scaling 236
– temperature dependence 231
– thermally activated 230
dielectric material 617, 1467
– energy barrier 624
– microelectronic device 615
– polarizability 619
dielectric relaxation
– self similar scaling 240
– time 169
dielectric response
– basic feature 222
– basic macroscopic definition 220
– correlated dynamics 243
– element 241
– low frequency 240
– physical concept 219
dielectrics 275
dielectric–semiconductor interface 629
diethylzinc (DEZ) 373
differential heat flow 440
differential quantum efficiency 909
differential scanning calorimeter (DSC) 425, 440
differential scattering 1197
– cross section 21
diffraction 1468
– electron 401
– limit 1357
– loss 1017
diffusant
– concentration 139, 140
– model 140
– profiling 139
diffusion
– anneal 139
– coefficient 139, 513
– current 877
– dopant 143
– hot-carrier 513
– isoconcentration 136
– length 873, 1333, 1468
– mechanisms 134
– recombination enhanced 138
– semiconductor 133
– short-circuit path 138
– source 139
– tracer concentration flux 136
– transient enhanced 137
– velocity 877
diffusion coefficient 134, 1468
– measurement 139
diffusion-induced disorder 137
diffusion-limited detectors 357
diffusivities of vacancies and Si interstitials 114
digital copying 967
digital packaging 1313, 1315
– thermal performance 1326
digital printing 967, 969, 1468
digital signal processor (DSP) 85
digital versatile disk (DVD) 557, 898, 1158
dihydrogen phosphate (KDP) 598
dilute magnetic semiconductor (DMS) 104, 106, 286
dilute nitrides 917
dimensional scaling
– capacitor dielectric 636
diode 177, 454, 714
– sputtering 658
dip coating 982
dipolar fluctuation 223
dipole
– disorder model 975
– field orientation 228
– fluctuation 225, 229
– glass 229, 235
– oscillator model 53
– permanent 243
dipole density fluctuation 228, 242
– relaxation rate 231
dipole moment 220
– effective 230
dipole relaxation 239
– fractal time processes 239
– self energy description 239
dip-pen nanolithography (DPN) 1270
dipping 1468
– boat 311
– cycle 666
– reactors 350
– techniques 349
Dirac cone 1159
direct
– alloy growth (DAG) 354
– -bandgap energy 743
– bandgap semiconductor 327
– chip attach (DCA) 1321
– current (DC) 196, 219, 455, 486, 567, 606, 660, 718, 1218
– gap semiconductors 562
– photo CVD 573
– piezoelectric effect 1468
director 935, 940, 948, 1468
– distribution 935
– orientation, optical properties 950
– reorientation, threshold voltage 949
discharged area development 969
disc-like molecule 939
discotic 1468
– liquid crystal 940
– nematic 939
dislocation 670
– density 755
– misfit 525
– threading 525
disorder
– model 976, 989, 1468
– potential 197
– static 61
disordered conductor 251
disordered ionic conduction 193
disordered material 193, 1117
– charge transport 198
– DC conductivity 197
– electrical conductivity 200
– electron mobility 207
– electron transport 201
– extended states 198
– hopping charge transport 200
– transport phenomena 195
disordered organic material 212
– carrier mobility 212
– charge transport 211
– hopping conductivity 212
disordered semiconductor 151, 1112
– AC conductivity 196

– density of band carrier 1115
– energy spectrum 195
– hopping conduction 201
– localized states 200
– materials 194
– transport properties 196
disordered solid 975
– electrical conduction 200
disordered sublattice 254
disordered systems
– doped crystal 194
– transport phenomena 194
– transport theory 197
dispersion 48
– ionic crystal 51
– relation 51, 52
– semiconductors 51
– semiempirical single oscillator single oscillator-based 52
dispersive transport 169
display device 941
Dissado–Hill function 226
– equation of motion 226
– fractional time power laws 226
dissimilar material integration 1075
dissipation power 221
distributed Bragg reflector (DBR) 797, 913, 1030, 1048
distributed feedback (DFB) 547, 899
– laser 914, 1052
distribution function 1468
DNA chip 1275
DNA origami 1275
doctor blading 603
domain wall (DW) 92, 101
donor 1468
– molecule 1335
donor–acceptor
– charge-transfer 984
– pair (DAP) 379, 777
– photoreceptor 984
– sensitization 1331, 1332
donor-bound exciton (DBE) 769
dopant 311, 351, 1468
– concentration 136, 138
– density 250
– diffusion 136, 137, 143, 145
doped semiconductor 1468
doping 357, 1337
– concentration 466, 839
– efficiency 359
– in GaAs 335
– in MCT 317
– in SiC 313
– profiling 466
Doppler broadening 790
DOS effective mass 486
DOS spectroscopy 167
double crystal x-ray diffraction (DCXRD) 546
double exchange 1218
double heterostructure (DH) 9, 549, 898
double quantum well (DQW) 549
double Schottky barrier 1243
double split-ring resonator (DSSR) 1355
double-channel planar buried heterostructure (DCPBH) 908
double-layer charging 1170
double-layer heterojunction (DLHJ) 316, 349, 883
– structure 359
double-layer planar heterostructure (DLPH) 883
double-negative medium 1355
drain depletion region 473
drain resistance 470
DRAM planar capacitor 637
drift measurement 495, 503
drift mobility 20, 169, 971, 972, 1468
– concentration dependence 215
– effective 1129
– temperature dependence 212
– time-of-flight 972
drift velocity
– electric field relationship 508
drift-diffusion relation 490
drive current 621
Drude
– approximation (DA) 518
– formula 493, 1468
– model 55
drum photoreceptor 981, 1468
– insulating material 983
drying 602
DSC cell
– heat-flux 440
dual beam photoconductivity (DBP) 158
dual layer organic photoconductor 1468
dual polarization interferometric (DPI) 1303
dual-band device 355
dual-in-line (DIL) 715
dual-in-line-package (DiP) 715, 1312
dual-layer photoreceptor 981
dual-waveband detector (DWB) 888
Dulong–Petit rule (DP) 427
Durham technique 284
DVD 557
d-wave symmetry 1235
dye-sensitized solar cell 172, 1266
dynamic
– coercivity 1204, 1468
– impedance 876
– random-access memory (DRAM) 606, 608, 620, 636, 1186, 1267
– resistance 881
– SIMS (DSIMS) 418
dynamical mean field theory (DMFT) 102
Dzyaloshinski–Moriya exchange interaction 96

E

E layer 1469
E′ center 1088, 1469
easy axis 94, 95, 1216, 1469
ECA metal particle 1325
ECB cell transmission curve 949
eddy current 1192
edge filter 684
edge-defined film growth 273, 1469
edge-defined film-fed growth (EFG) 287
edge-emitting laser
– elliptical beam profile 915
edge-on orientation 1335
effective
– dielectric constant 32
– dipole moment 230
– hole mass 735
– mass 1395, 1469
– media approximation (EMA) 32
– mobility 1116, 1117
– resistivity 23
– scattering 21
effusion cell 332
effusion sources 329
Einstein coefficient 901
Einstein relation 134, 790, 1469
Ekman flow 347
EL(2) 1469
elastic
– compliance constant (*C*) 1469

– constants 942
– deformation 1469
– energy density 942
– stiffness 729
– stiffness constant (S) 1469
elastic properties 729
– semiconductor alloys 731
elasticity 1469
elastoresistance 1118, 1119
– coefficient 1118
electric dipole (ED)
– extrinsic 183
– intrinsic 183
electric torque 946
electrical conductivity 257, 566, 686, 1260
– amorphous semiconductors 565
– metallic film 686
– nanotube 1164
– silicon 495
– size-dependent 688
– temperature-dependence 566
electrical measurement techniques 454
electrical properties 557, 685
– bulk SiO_2 628
– chalcogenide glasses 566
– electrical conductivity 566
– Hall coefficient of a-Si:H 566
– thin film 646, 685
electrical resistivity 27
electrically
– conductive adhesives (ECA) 1324, 1469
– controlled birefringence (ECB) 948
– detected magnetic resonance (EDMR) 171
– erasable programmable read-only memory (EEPROM) 637
electrically active impurities
– point defect 477
electroabsorption 68, 1469
– modulator 923
– modulator (EAM) 1052
electrochemical etching 1021
electrochemomechanics 263
electrochromic element 262
electrochromic window 258, 261
electrode-limited conduction 697
electrodeposition (ED) 662, 855
electroding 603
electrodynamics 1229
electroluminescence (EL) 549, 997
electrolyte solid oxide 249
electrolytic domain 257
– boundary 257
electromagnetic compatibility 91
electromechanical coupling 609
electron 1469
– affinity 1469
– beam lithography (EBL) 1017, 1019, 1469
– beam-induced current (EBIC) 385, 790
– cyclotron resonance (ECR) 665, 745
– diffraction 393
– energy analyser 415
– energy loss spectroscopy (EELS) 10, 385
– injection 691
– mobility 198, 1174
– motion 20
– paramagnetic resonance (EPR) 628, 789
– saturated drift velocity 510
– saturation velocity 1174
– scattering 878
– spectroscopy for chemical analysis (ESCA) 386
– spectrum 488
– spin resonance (ESR) 171, 559, 584, 628
– wavefunction 1086
electron avalanche photodiode (e-APD) 867, 869, 889
– device 869
electron device
– high-frequency 829
electron drift velocity 508, 837
– AlN 844
– doping dependence 840
– GaAs 844
– GaN 844
– InN 844
electron effective mass
– density-of-states 736
– III–V quaternaries 736
electron transport 829–831, 848
– AlN 841
– InN 842
– material parameter 832
electron transport material
– BCMF 977
– DPQ 977
– PTS 977
– TNF 977
electron-beam evaporation 652
electron-beam poling 1088
electronegativity, Miedema 183
electroneutrality equation 256
electron–hole pair (EHP) 1126, 1127
– creation energy 1130
Electronic Industries Alliance (EIA) 607
electronic material 309, 829
– C_{60} based 1170
– Debye heat capacity 428
– heat capacity 428
– silicon dioxide 625
– thermal conductivity 431
– thermal properties 425
electronic nose sensor system 1305
electronic optical nose 1295
electronic packaging 1311
– classes of material 1313
– encapsulation 1323
– interconnect 1318
– levels 1312
– material 1313, 1315
– solder 1318
electronic transport devices 726
electronic/optoelectronic materials 270
electron–lattice 1002
electron-momentum parameter 792
electron–nuclear double resonance (ENDOR) 559
electron–phonon
– coupling 563, 565
– interaction 492, 494, 1230
– inter-valley scattering 503
– scattering 1469
electron-transport charge mobility 989
electron-transport material
– charge mobility 989
electro-optic
– coefficient 595
– detection 1345
– device 600, 947
– effect 926, 1027, 1469
electrophotographic cycle
– printing 980
– process parameter 972
electrophotography 967, 968, 1469
electroplating 663, 1469
electrospinning 1301
electrostatic flux density 220
electrostatic latent image 968

elemental maps 415
Eley–Rideal reaction 1470
ellipsometric angles 70
ellipsometry 70, 73, 78, 1470
elliptical dichroism 1375
Elovich model 1292
emergency lighting 1007
emerging technology
– OLED 991
– organic photovoltaic devices (OPV) 991
– organic transistor 991
emission
– cross-section 1073
– device 1470
– stimulated 1073
empirical tight-binding approximation (ETB) 182
enabler
– metastability 1067
encapsulation 1470
– electronic package 1323
– material 1323
– mold compound 1324
– processing 1324
– temperature 1324
– wire sweep 1324
end-member perovskite 594
endothermic glass-transformation enthalpy 440
endpoint materials 740
endurance 1218
energy
– band diagram 460
– barrier 1201
– conversion 260
– dispersive x-ray (EDX) 385, 748
– dispersive x-ray spectroscopy (EDS) 389
– dissipation 224
– transfer coefficient 655
energy gap 1463, 1470
– iron-based superconductor 1251
– MgB_2 1246
energy relaxation
– charge carriers 211
energy-band spectrum 1470
energy-loss relaxation 209
enthalpy 426, 438
– crystalline 439
– exothermic 440
– fusion 438, 440
– relaxation 439
enzyme-linked immunosorbent assay (ELISA) 1296
epilayer
– ellipsometry 353
– laser reflectometry 353
– material 904
– monitoring 353
epitaxial
– CdTe layer morphology 677
– crystal growth 309
– film 1244
– GaN 408
– growth 671
– lateral overgrowth (ELO) 315, 793
– layer (or epilayer) 1470
– layer overgrowth (ELOG) 311, 744
– strain effects 1245
– substrates 269
epitaxy 310, 1470
equilibrium
– band bending 40
– concentration 113
– population 230
equipartition of energy 427
equivalent oxide thickness (EOT) 623, 1470
erasable programmable read-only memory (EEPROM) 638
erbium-doped fiber amplifier (EDFA) 918, 922, 1072
ESCALAB chamber 358
etch pit density (EPD) 359, 379, 1470
etching
– dry etching (RIE) 1062
– electrochemical 1021
– reactive ion etching (RIE) 1069
evaporation
– coefficient 649
– of compounds 652
– point source 650
– process 651
– reactive 653
– source 650
– surface source 650
evaporator temperature 671
evolution of the void density size distribution 121
Ewald sphere 1470
EXAFS (extended x-ray absorption fine structure) 560
exchange
– bias 86
– -coupled-composite (ECC) 1214
– field (H_{ua}) 1470
– length 92
– stiffness 101
excited state absorption (ESA) 1074
exciton 47, 156, 1043, 1331, 1470
– Bohr radius 773
– bound 769
– diffusion 1332
– diffusion length 1334
– dipole moment 1089
– dissociation 1334
– Frenkel-type 1331
– polariton (EP) 773
– recombination 977
– two-level system 1089
– Wannier-type 1331
excitonic absorption 65
excitonic PV 1107
excitonic radiative lifetime 780
exciton–phonon interaction 374
exhaust sensor 258
exhaustion concentration 1470
exothermic enthalpy 440
exotic QW 921
exponential DOS 1117
extended defect 792
extended graphics array (XGA) 952
extended states 195, 1113, 1114
extended x-ray absorption fine structure (EXAFS) 560, 1151
extinction coefficient 48
extraordinary magnetoresistance (EMR) 1220
extraordinary refractive index
– effective 948
extrinsic chirality 1372
extrinsic doping 348, 877
extrinsic ferroelectric Curie temperature 591
extrinsic optical activity 1371
Eyring rate law 233

F

fabrication 990
– facility (FAB) 1267
– nanoscale patterning method 1022
– photoreceptor 990

Fabry–Pérot (FP)
– cavity 905, 906
– interference filter 684
– laser 914
face-centered cubic (fcc) 607, 746, 1151
face-on orientation 1335
Faraday
– effect 68
– induction 1354
– law 1186
fast
– Fourier transformation (FFT) 475
– ion conductor (FIC) 251
– switching display 938
fatigue, thermomechanical 1318
FeCoNi, saturation magnetization 1190
FEL 1343
Fermi
– energy 25, 201, 489, 1113, 1230, 1470
– integral 1382
– level 1470
– surface 42
– velocity 686, 1176
Fermi–Dirac distribution 180, 491, 901, 1114, 1382
ferrimagnet 96
ferrite 98
ferroelastic 598, 1470
ferroelectric 229, 1471
– characteristic properties 591
– correlated dynamics 229
– correlation length 229
– Curie point 596
– device electroding 603
– dielectric response 235
– domain 591, 1471
– extrinsic 591
– hysteresis 591, 1471
– improper 591, 598
– laser ablation 606
– oxide 595, 606
– oxide thin film 606
– properties 590
– relaxor 1471
– sputtering 606
– static scaling 229
– system 592
– theoretical model 592
– thin film 606
ferroelectric application 591
– capacitor 607
– dielectric 607
– piezoelectric 608
– pyroelectric material 611
ferroelectric ceramic 230, 606
– fabrication technology 601
– poling 604
ferroelectric material 638
– Aurivillius compound 597
– Czochralski growth 600
– fabrication 600, 601
– ferroelectric oxide 593
– flux growth 600
– illmenite 596
– lead germanate ($Pb_5Ge_3O_{11}$) 598
– oxide 598
– perovskite ferroelectric 593
– phosphate 598
– polycrystalline ceramic 601
– polymeric 599
– properties 589
– reliability 639
– single crystal 600
– solution 600
– thick-film 604
– thin film 605
– triglycine sulphate 598
– tungsten bronze 597
ferroelectric random access memory (FRAM or FeRAM) 608, 638
ferroelectric to store information
– dynamic random-access memory 608
ferromagnetic (FM) 1198
– resonance 94
– spin configuration 1198
ferromagnetism 96, 1354
FET sensor 1285
fiber Bragg grating (FBG) 1065, 1083
Fick's first law of diffusion 1471
Fick's second law of diffusion 1471
field effect transistor (FET) 620, 726, 1471
field emission 1471
field factors 33
field-dependent barrier lowering 1138
field-effect mobility 1115
field-effect transistor (FET) 279, 474, 612, 617, 1168, 1177, 1263, 1285
field-emission display 1169
figure of merit (FOM) 612, 618, 1086, 1098, 1471
fill factor (FF) 854, 1471
film
– anodic oxide 663
– as-deposited 676
– a-SiN_x:H 1112
– CdS 676, 698
– CdSe 692, 693
– crystal structure 672
– in-plane strain 1245
– insulating 690, 697
– morphology 678
– oxide 659
– phthalocyanine 693
– Poisson's ratios 1120
– thickness 647, 674, 676
film-growing surface 575
filter
– band-pass 684
– band-stop 684
– design 684
– edge 684
– semiconductor-doped 1089
fin field effect transistor (FinFET) 6, 620
firing 1471
first Brillouin zone (FBZ) 1176
first-order transition 941
fixed charge 630
flame hydrolysis deposition (FHD) 1062
flame oxidation deposition (FOD) 1062
flash evaporation 653
flash memory element
– nitrided SiO_2 638
flat lens 1357
flat panel display 582
flat surface in homoepitaxy 310
flatband capacitance 468
flatband voltage (V_{FB}) 478, 630, 633, 1471
flexible substrate 710, 1111, 1112
flexoelectric coefficients 945
– measurement 963
flexoelectric effects 964
flexoelectric polarisation 955
flexoelectricity 944, 955, 1471
flip chip 1471
– interconnect 1319
– organic package 1321
– solder deposition technique 1320
– substrate 1322
– technology 1321, 1323
– underfill material 1323

float zone 275
floating gate current 1471
floating gate technique 471
– semiconductor parameter 471
floating-zone (FZ) 112, 270, 297, 495
floating-zone (FZ) method 297, 495, 1471
– properties 299
– silicon 298, 299
flow pattern defect (FPD) 115
fluctuation 273
fluence 1471
fluorescence 997, 1005, 1471
– quenching 978
fluoride glasses 1471
fluorinated silicate glass (FSG) 640
fluorine-doped tin oxide (FTO) 680
fluorite 250
flux
– lines (vortices) 1471
– method 287
– pinning 1472
– quantum 1231, 1232
flying height 1212, 1472
focal plane array (FPA) 316, 343, 358, 544, 873
– large-area 343
focused ion beam (FIB) 10, 405, 1019
forced convection 1472
foreign substrates 356
forming-gas anneal 630
forward biasing 476
Fourier transform (FT) 159, 167
Fourier's heat conduction law 431
Fourier-transform infrared (FTIR) 159, 347, 351, 612, 1283
– spectrometry 347, 1472
Fourier-transform photocurrent spectroscopy (FTPS) 159
four-point probe 1472
– technique 454, 456
Fowler–Nordheim 1215
Franck–Condon principle 1004, 1472
Frank–Read growth spiral 317
Frank–Turnbull mechanism 135
Frank–van der Merwe (FM)
– growth 374
Franz–Keldysh effect 68, 520
free
– convection 1472
– electron model 26
– film 1472
– layer 1193
free carrier 47
– absorption (FCA) 54, 56
– density 254
– plasma effect 914
Fréedericksz transition 949
freezing temperature 1061
Frenkel
– defect pair 249
– disorder 253, 257
– exciton 63, 1331
– formation energy 250
– pair 135
frequency
– dispersion 633
– selective surface (FSS) 1365
– synthesis 91
Fresnel
– coefficients 680
– equations 71
– reflectance spectrum 515
– reflection coefficient 73
– zone plate 1346
Friedel's law 1472
Fröhlich scattering 772
front end of line (FEOL) 1472
Fuchs–Namba model 689
Fuchs–Sondheimer
– equation 31
– formalism 689
– model 688
fuel cell
– monolithic 260
– solid oxide 260
full width at half maximum (FWHM) 359, 373, 392, 563, 859, 1179
fullerenes 1164
– transition temperature 1226
fundamental properties 680
fusion enthalpy 438, 440
fusion temperature 440

G

GaAs/AlGaAs 334
GaAsBi 921
gain spectrum 907
gain-guiding 907
gallium arsenide (GaAs) 837, 844
– device-quality 334
– solar cell 1102
gallium nitride (GaN) 12, 758, 829, 837, 844
– alloy 326
– average electron energy 835
– bound exciton 774
– defect 789
– donor-acceptor transition 776
– electron mobility 758
– free exciton 769
– mechanical property 749
– nanostructure 794
– n-type doping 787
– optical phonon frequency 750
– photoluminescence spectra 771
– p-type doping 787
– wurtzide 835
galvanomagnetic effect 496
galvanomagnetic measurement 495
garnet 98, 318
gas phase diffusion 584
gas precursor 373
gas sensor 1284
– impedance-based paste 719
– inorganic material 1283
– potentiometric 258
– semiconductor 1283
gas source, MBE 333
GaSb substrates 361
gas-source molecular beam epitaxy (GSMBE) 333, 545
gate
– capacitance 621, 1115
– electrode work function 624
– leakage currents 623
– oxide integrity (GOI) 112, 1472
gate dielectric 619, 620, 633
– thickness scaling 618
gate dielectric material 631
– reliability 624
gate-first process 634
gate-last process 634
gate-substrate capacitance 478
Gaussian DOS 213
Gaussian units 1189
Gay–Berne mesogen 935
Ge 279
– dopant diffusion 141
$Ge_2Sb_2Te_5$
– electronic excitation effect 1157
– metastable structure 1151
– molecular dynamics 1154
– non-thermal process 1157
$Ge_2Sb_5Te_5$ 164
Geiger mode operation 1050

geminate recombination 978, 1472
GEN III detector 886
generalized Drude approximation (GDA) 493
generic phase diagram cuprate superconductor 1235
geometry factor 874
Ge-on-insulator application 313
germanium (Ge) 4, 523, 1103
Ge-Sb-Te (GST)
– alloy 1151
– phase-change alloy 1150
GeTe 164
– amorphous phase 1152
– crystalline phase 1152
– structure 1152
ghosting 1472
giant magnetoresistance (GMR) 86, 1185, 1472
Gibbs free energy 231, 438
Gilbert damping 1217
Ginzburg–Landau
– parameter 1233, 1239
– theory 1232
Gladstone–Dale coefficient 51
glass 47, 1082
– aging 448
– aluminium oxide 1069
– aluminosilicate 1074
– bandgap energy 1085
– borofluorogallium, BIG glass (BaF2-IF3-GaF3) 1060
– ceramics 1089
– chalcogenide 1071
– enthalpy 439
– fast ionic conductivity 254
– fluoride 1069
– fluorozirconate (ZBLAN) 1060, 1069
– formation 234
– former 1061
– glass-transformation kinetics 447
– heavy-metal oxide 1070
– homogeneous 1083
– ionic transport 254
– key parameter 440
– laser glasses 1071
– matrix, electrical conductivity 193
– metallic 436
– nanoparticle dispersed 1089
– oxide 51, 430
– particle-embedded 1089
– phosphosilicate 1074
– poling method 1088
– silicon-based 1062, 1068
– substrate 1112
– surface, hydrophilic 1286
– Tellurium-based chalcogenide (TeX,TeXAs) 1060
– thermal conductivity 433
– third-order nonlinearity 1083
– transition region 448
glass ceramics 1472
glass frit 1472
glass transition 234, 441, 964, 1062
– activation energy 443
– polymer 429
– temperature 254, 425, 429, 439, 975, 988, 1324
– temperature (T_g) 1061
glasses 251
glass-forming liquid 438
glass-forming material 238
glass-transformation kinetics 447
glass-transition temperature 1472
glassy carbon (GC) 1285
glassy film
– high-index 1070
– low index 1068
– medium-index 1069
glassy metals 436
glow curve 1003
glow discharge
– mass spectrometry (GDMS) 416, 417, 778, 861, 1486
– optical emission spectroscopy (GDOES) 413
– spectrometry (GDS) 1472
– spectroscopy (GDOES) 413, 416
– sputtering 659
GMR
– effect 1192, 1193, 1196
– ratio 1193
– sensor 1220
Golay cell 1342
gold alloy thick film 712
gold conductor
– thick film 711
Goldschmidt criteria 593
gradient freeze growth 1472
gradient freezing (GF) 379
grading 1473
grain boundary 30, 1242
– conductivity 688
– scattering 688
grain size 1187
grain-growth process 345
granular
– medium 1212
– polysilicon 296
– recording layer (GRL) 1214
graphene 190
– bandstructure 1176
– characterization 1178
– nanoribbon (GNR) 1177
– photonic application 1180
– synthesis 1174
graphite sliding-boat system 312
graphitised SWNT 1165
grating structure 914
Grätzel cell 1266, 1473
grid method (GM) 576
GRINSCH laser 1473
group II oxide
– MOCVD 328
group III nitride 283
– crystal structure 745
– mechanical property 748
group III–V 314
group III–V alloy
– deformation potential 737
group III–V alloy system
– thermal conductivity 732
group III–V binaries 735, 736, 739, 740
– crystal density 727
– elastic properties 729
– lattice parameter 727
group III–V compound 737
– diffusion 143
group III–V material
– MOCVD 324
group III–V nitride semiconductor 830, 835
– material 845
– material parameter 832
– recent development 848
– transient electron transport 834
group III–V quaternaries 728, 735, 736, 740
– lattice-matching condition 728
group III–V semiconducting compound alloy
– electronic transport device 726
– laser diode 726
– light-emitting diode 726
– optoelectronic device 726
– photodetector 726
group III–V ternary 733, 740
– band-gap energy 733
– optical mode 730

group III–V ternary and quaternary alloy
– optical mode 730
group II–VI
– binary compound 144
– compound epilayer 372
– materials 283
– semiconductor, MOCVD 321
group II–VI compound 737
– bulk single crystal 366
– wide-bandgap 365
group II–VI compound semiconductor 376
– bulk single crystal 368
– high quality film 368
group II–VI wide-bandgap compound
– Bridgman 379
– gradient freezing 379
group index 53
group-III acceptor 489
group-III nitride semiconductor 315
group-IV semiconductor 141
group-V donor 489
growth
– characteristics 880
– from Te solutions 349
– gallium arsenide 280
– GaP 282
– InP 282
– layer-by-layer 330
– method 880
– model, μc-Si:H 578
– on silicon 360
– oxide 286
– regime 311
– silicon 278
– silicon carbide 279
– system, MBE 331
– TlInGaAsN quinary alloys 547
growth method
– Bridgman 273
– high-temperature solution 276
– hydrothermal 277
– low-temperature solution 276
growth technique, compound semiconductor 281
Grüneisen
– coefficient 1388
– constant 1383
– law 436
– parameter 437
GST alloy 1152
Gunn
– diode 1473

H

Hagen–Rubens relationship 55
half-metal 103
halides 288
Hall
– carrier concentration 463
– coefficient 490, 758
– coefficient n-Si 503
– constant 196
– factor 36, 490
– measurement 462
– mobility 36, 490, 741, 759
– mobility III–V binary 740
– resistivity 43
– scattering factor 464
Hall effect 19, 35, 1473
– ambipolar conduction 35
– carrier concentration 463
– coefficient 36
– materials characterization 462
– physical principle 462
Hall–Petch relationship 1316
Haloid Corporation 969
Hamiltonian, vibrational 1001
hand-held display 952
hard
– axis 93, 1216, 1473
– ceramic 1189
– ferrites 97
– magnetic bias film 1473
Havriliak–Negami fractional power laws 225
Havriliak–Negami function 225
head field 1206
– slope parameter (Q) 1473
head skew 1207
head-media separation (HMS) 1187
headphone 97
heat
– affected zone (HAZ) 1316
– assisted magnetic recording (HAMR) 1220
– exchanger method 288
– flow 432
heat capacity 1473
– composition dependence 429
– glasses 430
– low-temperature experiments 427
– normalized 430
– temperature dependence 430
heated dot magnetic recording (HDMR) 1220
heat-flow differential scanning calorimetry (DSC) 430
heavy-hole 925
– band 924, 926
– effective mass 488
– spectrum 510
Heisenberg exchange 102
helical anisotropy 96
hemispherical-grain (HSG) 636
HEMT device 336
hermetic packaging 717
herringbone structure 1289
Hervé-Vandamme relationship 51
heteroepitaxial layer
– critical thickness 524
– metastable layer 525
– pseudomorphic layer 524
heteroepitaxy 310, 319, 356, 371, 1473
heterofulleride 1170
heterogeneous system 1091
heterointerface 1333
heterojunction 898, 1332, 1333, 1473
– band alignment 526
– bipolar transistor (HBT) 334, 524, 726, 1473
– field effect transistor (HFET) 746, 801
– laser 1473
heteropassivation 877
heterostructure 330, 1473
– design 357
– detector 351
– lattice-matched 185
– metamorphic 185
– waveguide 1017
Heusler alloy 1219
hexagonal
– As 1473
– close-packed (hcp) 745, 1196
– columnar phase 939
HF MPC 162
Hg partial pressure 344
Hg vapor pressure 350
Hg-based alternative 343
HgCdMnTe 362
HgCdSe 543
HgCdTe 543, 873
– 2-D arrays 881

– diodes 875, 878
– dual-band array 888
– high-performance infrared systems 892
– infrared detector 869
– photoconductive 871
– photovoltaic array 892
– photovoltaic detector 892
– solution 350
– thermal diffusion currents 876
HgCdZnTe 362
HgMnTe 361
HgZnTe 361
high
– electron mobility transistor (HEMT) 8, 336, 726, 1177, 1474
– electron-momentum parameter 792
– field transport 766
– frequency (HF) 464
– Hg vapor pressure 353
– impedance surface (HIS) 1367
– melting point material 271
– nitrogen pressure (HNP) 790
– refractive index glasses 80
– speed electronic 8
– temperature/pressure 277
– vacuum (HV) 647
high-angle annular dark field (HAADF) 408
high-brightness diodes 326
higher operating temperature (HOT) 317, 355, 867
– device 886
higher silane-related species (HSRS) 583
highest occupied molecular orbital (HOMO) 1264, 1329, 1332, 1474
high-field conduction 690, 699
high-frequency (HF) capacitance method 464
high-frequency electron device 829
high-index contrast 1473
high-k dielectrics 6, 633
highly doped n-type wafer 127
highly doped Si room-temperature resistivity 498
high-order Laue zone (HOLZ) 396
high-pressure Bridgman (HPB) 861
high-pressure depletion (HPD) 584
high-pressure diamond 279
high-purity n-Si 510
high-q detector 929
high-resistivity crystal 504
high-resolution
– displays 952
– electron microscopy (HREM) 397
– mask 1018
– transmission electron microscopy (HRTEM) 624, 793
– x-ray diffraction (HRXRD) 140, 746
high-T_c superconductor 1228
high-temperature
– oxide superconductor 1410
– solution growth 1474
– solution growth (flux) 276
– superconductivity 1226
high-temperature superconductor
– irreversibility line 1239
– physical property 1228
– upper critical field 1239
hillock 333, 356
hole
– array 1020
– drift velocity 508
– effective mass 735
– spectrum 488
hole transport 970
– trap-free 976
hole transport material
– DEH 973
– DPB 973
– TAPC 973
– TPD 973
– TTA 973
holographic patterning 1026
holography 89
homeotropic alignment 1474
homoepitaxial GaN 407
homoepitaxy 310, 319, 1474
homojunction 898, 1474
– device 877
HOMO–LUMO gap 1335
hopper coating 981
hopping
– band 1117
– charge transport 976
– conduction 696, 1474
– conductivity 203, 210
– distance, time-dependence 207
– energy relaxation 208, 213
– length 209
– rate 202, 206
– transition 201, 210
– transport 205, 212, 972
horizontal directional solidification crystallization (HDC) 288
host material, glass 1068
hot carrier 624
– diffusion 513
– phenomena 507, 1474
hot isostatic pressing (HIP) 603
HOT MCT device by LPE 317
hot-wall epitaxy (HWE) 365, 371
– II–VI compound 372
hot-wire chemical vapor deposition (CVD) 1474
humidity sensor
– thick-film 720
Huygens metasurface 1370
Huygens's principle 1474
HWE growth chamber 373
hybrid circuit 708, 1474
– substrate 710
hybrid graphene-gold nanoparticles (AuNp) 1285
hybrid thin film 1296
hybridization 1474
hydride vapor phase epitaxy (HVPE) 744, 763
hydrogen 632
– dilution ratio 575
– plasma 578
– sensor, titanium dioxide nanotube 1284
hydrogenated 573
– amorphous carbon transport properties 194
– amorphous germanium (a-Ge:H) 61
– amorphous silicon (a-Si:H) 61
hydrophobic substrate 666
hydrostatic pressure 928
hydrostatic stress 120
hydrothermal 277
– growth 284, 1474
– technique 377
hyperbolic dispersion 1363
hyperbolic metamaterial 1362
– enhancement of spontaneous emission 1365
– negative refraction 1363
– subwavelength imaging 1364
hypergeometric function, confluent 226
hyperlens 1364, 1365
hysteresis 633, 1474
– loop 89

I

IBM 986
IBM copier 969, 985
IC technology 617
ideality factor 177, 1474
identical absorption spectra 1336
IFIGS continua 176
IFIGS electric dipole
– intrinsic 183
IGZO 158
III nitrides 11
III–V compound diffusion 143
III–V material, band gap 1038
III–V MOCVD 324
III–V semiconductor 1, 928
III-V-Bi alloy
– growth 551
– optical properties 552
– semiconductor 550
III-V-Bi quaternary alloy 552
III-V-Bismides 911
II–VI MOCVD 327
II–VI narrow bandgap semiconductor 867
illmenite
– $LiNbO_3$ 596
– $LiTaO_3$ 596
image
– contrast analysis 397
– -force effect 177
– formation, photoreceptor 970
– lag 1474
– lag and ghosting 1128, 1144
imaginary susceptibility component 225
imaging 393
immunogamma globulin (IgG) 1296
IMP growth 362
impact avalanche transit time (IMPATT) 513
impact ionization 38, 513, 878, 929, 1474
IMPATT diode 514
improper (or extrinsic) ferroelectric 1474
impurity 1474
– absorption 65
– distribution 275
– ionization energy 489
– scattering 492
– solubility 302
– species 660
in situ monitoring 353
index of refraction, second-order 1083
index-guided structure 907
in-diffusion 144
indirect-gap material 1001
indium
– gallium zinc oxide (IGZO) 1117
– nitride (InN) 765, 781, 829
– tin oxide (ITO) 2, 262, 855, 1027, 1102, 1168, 1174, 1264, 1295
– zinc oxide (IZO) 1117
induced dipole moment density 223
induced uniaxial anisotropy 1187
inelastic scattering 1474
infrared (IR) 343, 491, 560, 611, 899, 987, 1288
– laser absorption spectroscopy (IRLAS) 574
– photon detector 868
– reflection absorption spectroscopy (IRRAS) 576, 1474
infrared detector 600, 869, 1051, 1055
– cooler 870
InGaAs 336
InGaAsN 917
InGaAsNSb 917
InGaN 7, 11
initial permeability 1188
injection
– circuits 876
– current 1138
– efficiency 876
InN 844
– electron transport 842
– material parameter 833
– mechanical property 749
– optical phonon frequency 750
inorganic crystal structure database (ICSD) 1402
inorganic glass 254
inorganic–organic hybrid material 669
in-plane
– coherence length 1239
– mass 905
– switching (IPS) 953
input/outputs (I/O) 1312
in-situ
– analysis 329
– doping 360
– monitoring 1475
insulating film 690, 697
insulation layer, thick film 713
insulator 62
– electrical characterization 454
– resistivity 455
– work function 690
insulator–metal transition 489
integrated
– detector – cooler assembly (IDCA) 870, 1475
– device 1082
– electronic absorption 1084
– magnetic concentrator (IMC) 108
– optics, glass 1068
– optics/photonics 1475
– photonics 1060
integrated circuit (IC) 1, 5, 293, 615, 714, 1288, 1475
– gate dielectrics 619
– manufacturing 616
– performance 617
– scaling 617
– wire bonding 716
intelligent catalyst 1410
interaction
– cross section 656
– material 386
– particle-material 387
– radiation 386
interatomic potential energy 436
interband
– optical transition 52, 1042
– transition region 739
– tunneling 877
intercation coulomb repulsion 1397
interconnect 639, 640, 1475
– capacitance 617
– conductivity 1325
– dielectric tetraethoxysilane (TEOS) 640
– dielectrics 619, 639
– electrical 1317
– environment 1317
– flip chip 1319
– Pb-free 1321
– polymer 1325
– second-level 1325
– wire material 617
interconversion 105
interdiffused multilayer process (IMP) 328, 353, 354, 1475
interdiffusion 144, 146
– compositional 137
– MCT 354

interdigitated electrode (IDE) 611, 1283
interdigitated structure 1335
interface 622
– band-bending 178
– capacitance 241
– dangling bond 628
– electrical characterization 454
– resistivity 459
– roughness 1063
– state density (D_{it}) 633
– trap 630
– trap capacitance 465
– trap charge 467
– trap density 1475
interface-induced gap states (IFIGS) 175, 1475
interfacial phase-change memory (iPCM) 1159
interfacial transport 255
intergrain critical current density 1475
intermediate diffuser 141
intermetallic compound (IC) 1226, 1319, 1475
intermetallic reactions and phases 1475
intermetallics 97
intermolecular force 963
internal
– energy 427
– gain 1076
– photoemission yield spectroscopy (IPEYS) 176, 179
– quantum efficiency (IQE) 800, 909
International Technology Roadmap for Semiconductors (ITRS) 616
interpolation scheme 726
interpolation, linear 737
interrupted field time-of-flight (IFTOF) 170
interstitial 1475
– diffusion mechanism 135
– recombination 137
– site 249
– trapping 127
interstitialcy mechanism 135
intersubband (ISB) 1052
– absorption 1045, 1055
– detector 1051
– emission 1045
– laser 1049
– transition 1045, 1475
inter-valley scattering 492
intraband transition 920
intragrain j_c 1475
intrinsic
– chirality 1371
– ionic disorder 253
– mobility 470, 496
– pinning 1244
– point defect 1475
– point defect aggregation 120, 123
– resistivity 689
– silicon 141
inverse spin-Hall effect (ISHE) 104
inversion 1475
– layer 465
– symmetry 1475
inverted
– metamorphic structure (IMM) 1104
– OLED (iOLEDS) 1265
– opal 1025
– staggered TFT 1116
ion
– beam milling 880
– bombardment 584
– conductor, disordered oxygen 253
– implantation 880, 1019, 1475
– species primary 420
– yield, positive secondary 421
– yield, relative 421
ionic
– conduction 193
– conductivity 253
– conductivity interface 255
– conductor 248
– conductor application 248
– crystal 51
– disorder 251
– radii 1406
ionization 1475
– energy 1475
ionized-impurity scattering 507
ion–lattice interactions 1001
ion-selective electrode (ISE) 1292, 1299
IPEYS technique (internal photoemission yield spectroscopy) 1475
iron pnictide 1249
– superconductor 1248
iron-based superconducting film 1250
iron-based superconductor 1248
– critical current density 1250
– doping 1249
– phase diagram 1249
irreversibility field
– MgB_2 1246
irreversibility line 1240, 1476
Irvin curve 1476
isobutyl iodide (IBI) 357
isoconcentration diffusion 136
isolation dielectrics 619
– LOCOS isolation oxide 635
isothermal capacitance transient spectroscopy (ICTS) 559
isotope effect 1231
isotope labeling 580
isotropic liquid 935
itinerant magnetism 85, 101
ITO 1265
I–V technique (current–voltage technique) 1474

J

Johnson–Mehl–Avrami (JMA) equation 443
Josephson junction 1242
– critical current 1231
Josephson–Fraunhofer interference 1232
Joule heating 870
Joule–Thompson cooler 870, 871
Judd–Ofelt (JO) analysis 66
jump process, diffusion coefficient 248
junction
– field-effect transistor (JFET) 478
– forming 880
– geometry 882
junction-formation method 352

K

Kauzmann paradox 439
Kauzmann temperature 441
Kelvin contact resistance (KCR) 454
Kerr effect 1027
Kevlar 935
Kikuchi lines 1476
kinematic viscosity 964
kinetic theory of gases 648
Kirkendall effect 137

Kirkwood factor 230
Kissinger plot 444
Knoop hardness 729
Knudsen cell 1476
Knudsen effusion 332
Kodak Ektaprint 100 969
Kohlrausch–Williams–Watts (KWW) 228, 441
Kramers–Kronig relation (KKR) 49, 72, 227, 491, 1083, 1085, 1476
Kretschmann's configuration 1293
Kyropoulos 273
– growth method 380
– method 288
– technique 1476

L

Landau diamagnetism 101
Landauer formula 1167
Landé factor 88
langasite material 287
Langevin function 228
Langmuir–Blodgett (LB) 646
– film deposition 665, 1476
– method 1287
– technique 662
Laplace transform 167
Laporte's rule 1000, 1476
large area electronic 1111
large-area growth 349
large-diameter crystal 273
laser
– ablation 606
– crystallization 1476
– detector 349
– diode (LD) 9, 365, 543, 726, 739, 743, 1047
– diode array 919
– spectral purity 913
– threshold 909
– trimming, thick film 715
laser light scattering (LLS) 112
laser light scattering tomography defect (LSTD) 115
laser-induced fluorescence (LIF) 574, 1476
LAST material 1387
latent image formation 970
lateral conductivity 685
lateral epitaxial overgrowth (LEO) 744, 747
lattice
– group III–V binary 727
– matched 921
– melting 251
– mismatch 356, 524, 1244
– mobility 504
– parameter 727
– parameter, iron-based superconductor 1251
– vibrations 22
lattice constant
– AlN 747
– GaN 747
– InN 747
lattice-matching 544, 1038, 1476
– condition 727, 728
law of mass action (LMA) 135, 1476
layer composition 982
layer thickness 351
layer-by-layer assembly 1268
layer-by-layer growth 330, 331
LD application, III-V-Bi 553
LDA (local density approximation) 103
lead
– iodide 1134
– oxide 1134, 1135
– phthalocyanine film 696
– zirconate titanate (PZT) 620, 718
– zirconium titanate (PZT) 639
lead–antimony–silver–telluride (LAST) 1387
lead-free piezoelectric 595
leakage current 616, 621, 875, 878, 911
LEC Czochralski 282
LEC technique 281
left-handed medium 1355
LF MPC 163
Li battery 258, 261
lifetime 348, 1073, 1476
ligands 66
light
– absorption 47
– amplification 1087
– emitting device 9
– hole 925
– point defect (LPD) 117
– source, glass-based 1071
light-emitting diode (LED) 1, 7, 9, 283, 310, 365, 726, 739, 743, 892, 898, 969, 997, 1037, 1047, 1075, 1169
– inorganic 997
– white 1006
light-hole
– band 924
– effective mass 488
light-induced defect creation 1476
light-induced phenomena 557
– amorphous semiconductors 567
lightly doped drain (LDD) 458
lineage 1476
linear
– interpolation 737
– interpolation scheme 726
– relaxation response 224
– response 222
– susceptibility 227
linearized muffin-tin orbitals (LMTO) 102
linewidth enhancement factor 926
liquid crystal (LC) 933, 1028, 1476
– alignment 940
– bulk nematic 942
– calamitic 935
– chemical structure 956
– chiral 937
– chiral nematic 955
– cholesteric phase 937
– columnar 939
– device 947
– dielectric anisotropy 960
– dipole moment 960
– director 940
– discotic 939
– elastic constants 961
– electrical conductivity 953, 958
– helical structure 937, 960
– long range order 940
– lyotropic 935
– material 957
– material, fluorinated mesogens 962
– molecular properties 958
– molecular shape 935
– molecular structure 935
– negative dielectric anisotropy 954
– nematic 935, 939
– optical properties 947, 958
– optical retardation 948
– order 936
– permittivity 960
– positive dielectric anisotropy 940, 948

– rod-like molecules 940
– shutter 956
– smectic 935
– thermotropic 934
– torsional elasticity 964
– twist grain-boundary phase 938
– viscosity 945, 964
– viscosity coefficients 964
liquid crystal display (LCD) 2, 14, 582, 933, 953, 1257
– addressing 949
– angle of view 949, 954
– bistable device 954
– deformed helix mode 964
– ECB-mode 948
– ferroelectric smectic display 955
– in-plane switching mode (IPS) 954
– materials 956
– multicomponent mixtures 957
– multiplexing 950
– optical properties 950
– passive matrix addressing 949
– relaxation time 950
– super-twisted nematic (STN) 951, 952
– thin-film-transistor twisted-nematic 952
– time response 950
– transmittance 950
– twisted nematic (TN) 951
– twisted vertically aligned (TVAN) 954
– vertically aligned nematic (VAN) 954
– viscous properties 950
– zenithal bistable device (ZBD) 955
liquid encapsulated Czochralski (LEC) 272
liquid phase electroepitaxy (LPEE) 312
liquid phase epitaxy (LPE) 8, 270, 349, 365, 371, 879, 898, 1038, 1476
– second-generation MCT 349
liquid silicon 297
– impurity solubility 302
liquid thin film 1270
liquid, key parameter 440
liquid-crystal mixtures 962
– two phase coexistence 957
liquid-crystal phase
– disc-like molecule 939
liquid-encapsulated Czochralski technique 1476
lithium ion battery 2, 258, 261
lithographic method 1018
local chemical state of surface 416
local density approximation (LDA) 102, 190, 748
local density approximation to density functional theory (LDA-DFT) 190
local field 1090
local oxidation of silicon (LOCOS) 620, 635
local spin density approximation (LSDA) 88
localized magnetism 85, 86
localized state 152, 195, 206, 1113
– potential barrier 211
localized vibrational mode (LVM) absorption 1476
lock-in detection 158
LOCOS isolation oxide 635
logic device dimensions 621
logic gates 1168
log-normal distribution 1200
log-pile structure 1024
London equations 1229
London penetration depth 1229
London theory 1232
lone pair electrons 1154, 1155
long wave (LW) 1476
longitudinal
– acoustical (LA) 491
– bias film 1473
– gauge factor 1119
– magnetic recording (LMR) 1186, 1477
– recording media 1199
– resistivity 43
– TFT 1121
longitudinal optical (LO) 54, 491, 731, 750, 758, 772
– modes 731
– phonon emission 1047
– phonons 54
longitudinal-mode peak 549
long-range disorder 1477
long-range orientational order 940
long-wavelength
– infrared (LWIR) 351, 543, 868
– LO phonon 737
– phonon 729
loophole device 880
Lorentz force 1220
Lorentz oscillators 51
loss coefficient 54
low
– field transport 758
– frequency (LF) 464
– phonon energy 1071
– pressure chemical vapor deposition (LPCVD) 523, 535, 664
low-dimensional
– structure 315, 1477
– transport phenomena 1167
low-energy
– electron beam irradiation (LEEBI) 787
– ion scattering (LEIS) 388
lower critical field
– cuprate superconductor 1240
– MgB_2 1247
lower explosive limit (LEL) 1282
lowest unoccupied molecular orbital (LUMO) 1264, 1329, 1332, 1477
low-field electron drift mobility
– GaN 837
low-frequency dispersion 240
low-frequency noise (LFN) 473, 1477
– characteristics 473
– measurement 476
– spectroscopy 473
low-impedance device 476
low-k application 640
low-k dielectric 6
low-loss dielectric material 1014
low-pass filter 1219
low-reluctance 1205
low-T_c superconductor 1228
low-temperature
– co-fired ceramic (LTCC) 1314, 1477
– solution growth 276, 1477
LPE
– apparatus 311
– background impurities 311
– characteristics 310
– GaAs 314
– growth 371
– growth mode 311
– history 310
– material point defects 311
– of MCT 312
– optoelectronic device 314
– second-generation MCT 349
LPE growth method
– dipping system 312

– graphite sliding-boat 312
– tipping furnace 311
LPEE of CdTe 316
L-pit 112, 127, 1477
LSDA (local spin density approximation) 103
LTCC 1322
lucky drift (LD) model 39
lucky electron 38
luminescence 1477
– band-to-band 1001
– centres 999
– deep-level 1001
– electronic configuration 999
– excitation/emission cycle 1002
– optically stimulated 1004
– rare-earth 999
luminescent ions 1001
luminescent material 997, 1006
Luttinger liquid (LL) 1167

M

macrocylic compound 1286
macrocylic molecule 1287
macrodefects 356
macromolecule, cylindrical graphitic 1163
macroscopic defect 333
magnetic
– anisotropy energy 94
– annealing 1477
– cap layer (MCL) 1214
– field applied continuous CZ (MCCZ) 305
– field applied CZ (MCZ) 304
– film 100
– material, class 96
– material, thick film 604
– metamaterial 1353
– mirror 1367
– noise 1220
– permeability 48
– random-access memory (MRAM) 1185, 1215, 1477
– recording 1186
– resonance imaging (MRI) 1354
– sector instrument 418
– semiconductor 893
– spacing 1191, 1477
– splitting 103
– squareness 1211
– surface charge 87
– susceptibility 93
– thin film 1187
– transition 1477
– tunnel junction 1215
– tunneling junction (MTJ) 1186, 1477
– viscosity 91
magnetic-semiconductor memory device 1185
magnetization 88
– curve 1210
– easy 1188
– hard 1188
magnetocrystalline anisotropy 87
magneto-optic material 309
magneto-optical effect 734
magneto-optics 89
magnetoresistance (MR) 490, 1477
magnetoresistive sensor 720
magnetostatic
– coupling field 1196
– dipolar coupling 100
– energy 87
magnetostriction 1188, 1191
– constant 1187
magnetron sputtering 661
majority carrier 161, 1477
many-body effect 102
Marconi company 956
mask processing 1019
masking effect 1333
mass flow controllers (MFCs) 334
master curve, polyallylbenzene 227
master oscillator power amplifier (MOPA) 919
master-curve technique 236
material
– charge transport 193
– composition 199
– disordered 193
– electronic 406
– for x-ray imaging 318
– optoelectronic 406
– organic thin film 698
– porous 1024
– sensor-active 1281
– thermal conductivity 431
– thick film 711
material LPE grown
– arsenic-based material 314
– atomically flat surface 318
– doping 315
– group III–V 314
– group II–VI 316
– group IV 313
– group-III nitride 315
– group-III–V antimonide 314
– mercury cadmium telluride 316
– mixed alloy layer 315
– narrowgap group II–IV compound 317
– phosphorus-based material 314
– silicon 313
– silicon carbide 313
– silicon/germanium 313
– widegap compound 316
material MBE grown
– GaAs/AlGaAs 334
– GaAs-based electronic device 337
– group-III nitride 336
– group-III–V nitride 337
– InGaAs 336
– precursor 337
– quality 336
– quantum dot 336
– sapphire substrate 336
– Stranski–Krastanov mode 336
material MOCVD grown
– $Al_{1-x}Ga_xAs$ 324
– antimonides 325
– arsenides 324
– GaN 325
– group II oxides 328
– group II–VI semiconductors 327
– HgCdTe 327
– nitrides 325
– phosphides 324
– selenides 328
– sulfides 328
– ZnSe 328
material parameter 832, 833
– quadratic dependence 726
– ternary 726
material properties, group III–V nitride semiconductor 845
Matthiessen's rule 23, 31, 686, 739, 758
Mauguin condition 951, 952
maximum differential permeability 90, 98
Maxwell–Boltzmann distribution 29, 649
MBE 374
– AlAs 330
– capital costs 336
– commercial systems 333
– GaAs 330

– group III–V compound 330
– growth chamber 332
– growth rate 330, 332
– history 329
– reaction mechanisms 330
– $Si_{1-x}Ge_x$ film 530
– technology, MCT 358
MBE growth
– defects 333
– substrate-preparation 332
MBE-grown MCT
– alternatives 361
– applications 360
– device-quality 359
– surface morphology 359
MCT
– advanced structure 359
– alloy composition 358
– array 357
– carrier lifetime 351
– composition 350
– dislocation density 357
– epitaxial growth technique 344
– epitaxial layer 351
– hybrid array 360
– impurity segregation 348
– infrared detector 360
– layer 359
– material characteristic 351
– MBE technology 358, 359
– MOVPE 353
– pseudobinary alloy composition 354
– sliding boat growth 350
– structure 359
– substrate material 356
MCT growth 285
– Bridgman 345
– Hg-rich melt 349
– monitoring 357
– solid state recrystallization (SSR) 345
– Te-rich 350
– traveling heater method) 347
MCZ silicon 304
mean free path 432, 685
– bulk electron 685
mean phonon velocity 432
mean-field approach 229
mechanical
– exfoliation 1285
– properties, thin film 646
– sensor, piezoresistive 718
– sensor, thick film 718
– slider 1186
media
– coercivity 1203
– flux 1199, 1478
– noise 1200, 1478
medium wave (MW) 1478
medium-energy electron diffraction (MEED) 402
medium-wavelength infrared (MWIR) 351, 868
meeting front (MF) 793
MEH-PPV 158
Meissner effect 1226, 1229
membrane
– conductivity 1304
– sensing 1286
– solid state ionic 248
– spin coating 1286
memory
– capacitor 619
– device tunneling oxide 619
– technology 636
memristors 262
mercuric iodide 1132
mercury cadmium telluride (MCT) 269, 310, 316, 343, 543, 873
mercury-sensitized photo-CVD 1478
mesa heterojunction (MHJ) 884
Meservey–Tedrow 103
mesogen 937, 1478
– chemical constitution 960
metafilm 1365
metal 25
– alloy sheet resistivity 711
– bulk nonlinearity 1090
– electrical conductivity 436
– electrical properties 25
– gate electrode 623, 634
– nanoparticle 1336
– oxide semiconductor field effect transistor (MOSFET) 5
– residual resistivity 1225
– resistivity 24, 455
– sheet resistivity 711
– thermal expansion coefficient 437
metal/oxide/semiconductor (MOS) 1478
metal-free phthalocyanine 1294
metal-induced gap state (MIGS) 183
metal–insulator transition 1235
metal–insulator–metal (MIM) 636
metal-insulator-metal-insulator-metal (MI-M-I-M) 1178
metal–insulator–semiconductor (MIS) 376, 695, 1273
– heavily doped polycrystalline Si 623
– structure 623
metallic film 686, 690
– electrical conductivity 686
metallic glasses 436
metallic nanocomposites 1478
metallic superconductor 1233
– isotope effect 1231
metalloporphyrin
– glass 1290
– Langmuir–Blodgett film 1290
– polymer 1290
metallurgical-grade silicon (MG-Si) 295
metal-molecule-metal junction 1273
metal-organic (MO) 546
metal-organic chemical vapor deposition (MOCVD) 8, 309, 365, 371, 407, 605, 665, 744, 857, 1021, 1038, 1105, 1478
metal-organic deposition (MOD) 605
metal-organic molecular beam epitaxy (MOMBE) 334, 365, 371, 374, 744
metal-organic vapor phase epitaxy (MOVPE) 8, 283, 319, 344, 365, 371, 546, 879, 1038, 1104, 1478
– reactor cells 322
metal–oxide–semiconductor (MOS) 464, 514, 523, 617, 620, 685, 1186, 1272
metal–oxide–semiconductor field-effect transistor (MOSFET) 5, 454, 524, 630, 1168, 1272
metal–semiconductor
– contact, J–V characteristic 459
– field-effect-transistor (MESFET) 8, 463, 1478
metal-semiconductor (MS) 464
metal-ultrathin insulator–semiconductor (MUTIS) 188
metamaterial 1339, 1346, 1351
metamolecule 1352
– geometrical resonance 1353, 1366
– resonant behavior 1352

metamorphic growth 921
metastability 1064, 1478
metasurface 1365
Meyer–Neldel rule 146
MgB_2
– critical temperature 1245
– crystal structure 1246
micro screen 708
micro-ampere 1330
microcrystalline silicon, hydrogenated 573
microdiffraction 399
micro-electromechanical (MEM) 262
micro-electromechanical system (MEMS) 262, 605, 1028
– thin film 605
microelectronic application
– thermally grown SiO_2 627
microelectronic circuit 707
– Si-based 615
microelectronic device
– dielectric material 615
– scaling 631
microelectronic materials 627
microelectronics
– capacitor 615
– interconnect 615
– transistor 615
microhardness 729
micromachining 1284
– technique 1020
microphotonics 1478
microprocessor
– performance 618
– scaling 617
– unit (MPU) 1267
microstrip design 1322
microstructure glass (MSG) 1299
– sensor 1299
microvia 1322
microwave
– circuit 1314
– device 314
– plasma chemical vapor deposition (MPCVD) technique 1478
mid-gap state 1087, 1091
migration energy 248, 254
Miller's rule 1084, 1478
milliampere 1332
milling 602
millions of instructions per second (MIPS) 1275
miniaturization 248
miniband 920, 1040, 1042, 1479
minimal thermal conductivity (MTC) 1386
minority carrier 161, 790, 928, 1479
– lifetime 343, 359
– silicon-based electronic device 497
mirror 680
MIS diode 630
misalignment angle 1242
misfit dislocation 1479
mismatch 12
mixed conducting 247, 260
mixed crystal
– charge carrier 200
– compositional fluctuation 200
mixed state 1479
mixture rule 33
mobile carrier
– concentration 250
– sublattice 251
mobile monitoring 1283
mobility 19, 248, 633, 740, 877, 1260, 1392, 1479
– degradation 622
– degradation coefficient 470
– dipole disorder model 975
– disorder model 975
– edge 195, 206, 210, 558, 1114, 1479
– effective 1116, 1117
– field dependence 974
– field-effect 1115
– gap 195, 205, 1113, 1114, 1137, 1479
– of electrons 208
– polaron model 975
– Poole-Frenkel field dependence 975
– temperature dependence 974
MOCVD 324
– basic reaction kinetics 319
– characteristics 319
– precursor 321
– technique 373
Moddera–Mathon 103
mode
– AlAs-like 731
– GaAs-like 731
model dielectric function (MDF) 739
mode-spacing 906
modulated photoconductivity (MPC) 157, 161
modulation transfer function (MTF) 1143, 1479
modulation-doped
– FET (MODFET) 336
– GaAs/AlGaAs 335
modulator 897
– electroabsorption 923
– electro-optic effect 926
– InGaAsP-based 925
– linewidth enhancement 926
– optical 898
– refractive index 926
– wave-guide structure 925
– Y-junction 926
molecular
– beam epitaxy (MBE) 8, 180, 283, 309, 344, 365, 371, 398, 523, 647, 744, 856, 879, 899, 1038, 1479
– chirality 937
– density 727
– dipole relaxation time 232
– dispersion 974
– dynamics (MD) 560
– junction 1273
– transistor 1274
– wire 939
molecular electronics 1479
molecular reorientation 1479
molecularly
– doped polymer (MDP) 194, 211, 972, 1479
– imprinted polymer (MIP) 1301
molten semiconductors, diffusion 146
momentum-flipping collisions 432
monolayer (ML) 330, 371, 795, 1479
monolithic solder 1321
monomer methacrylic acid (MAA) 1302
monopolar conduction 36
monopole inductive write element 1206
monosilane (SiH_4) 295
– production 296
– pyrolysis 296
Monsma–Parkin 103
Monte Carlo (MC) 560
Monte Carlo (MC) simulation 762, 829
– electron transport 837
Moore's law 6, 616, 1266, 1479

morphological characteristics 645, 677
morphotropic phase boundary (MPB) 595
Moseley's law 1479
MOSFET
(metal–oxide–semiconductor field-effect transistor) 454
Moss rule 1085
Mott's law 204
MOVPE/MBE combined growth 337
MQW
– band edge absorption 1043
– laser 336
MRAM architecture 1218
MTJ memory cell 1217
multi-beam optical stress sensor (MOSS) 262
multi-component material sputtering rate 416
multi-crystalline silicon (mc-Si) 1101, 1479
multiferroic 105
– magnetoelectric 598
multijunction cell 1103
multilayer
– ceramic (MLC) 602
– cuprate superconductor 1244
– device 981
– heterojunction (MLHJ) 359
multilayer film 569
– quantum size effect 563
multi-mode 913
multimode behavior 729
multioscillator model 738
multiple
– quantum disk (MQD) 800
– quantum well (MQW) 68, 137, 283, 374, 795, 797, 869, 1040, 1479
– reflection approximation 515
multiplets 999
multiplexer 876
multistrain well 925
multi-stripe 918
multisubstrate dipping method 313
multiwall nanotube (MWNT) 1163
– ballistic conductance 1167
– electrical transport 1166
MUTIS Schottky contact 187

N

n^+-p diodes 878
N_2O complex 125
N_2V complex 124
N_2V_2 complex 124
NAND-FLASH 1215
nano volume diffusion 146
nanochannel alumina (NCA) 1020
nanochannel glass (NCG) 1020
nanocrystalline diamond 159
nanoelectronic device 699, 1168
nanofibrous membrane 1301
nanogap electrode 1273
nanomaterial 1284
nanometer 1333
nanoparticle 679
– system 1089
nanoporous array, high-aspect 1020
nanostructure 647, 680
– GaN 794
– oxide 1170
– semiconductor 1167
nanotechnology 557
nanotube
– carrier mobility 1164
– electrical conductivity 1164
– electronic structure 1166
– electrooptics 1169
– logic gates 1168
– semiconducting 1168
– solar cell 1169
– thermal conductivity 1164
nanowire (NW) 799, 1016, 1027, 1028, 1053
nanowire light-emitting diode (NW LED) 799
narrow band gap 1479
– II-VI semiconductor 343
narrow gap compounds 317
narrow gap III–V compounds 892
NASICON 259
native defect 137, 1479
native interstitials 143
near-equilibrium grown process 310
nearest-neighbor hopping (NNH) 202
near-infrared (near-IR) 1013
– slices 346
near-lattice-matched heteroepitaxy 310
Néel temperature 1235
Néel–Arrhenius formula 91
negative
– -bias temperature instability (NBTI) 630
– differential mobility (NDM) 509
– differential resistance (NDR) 767, 1480
– FSS 1366
– -index medium 1355
– refraction 1356, 1363
– temperature coefficient (NTC) 719
– temperature coefficient of resistance (NTCR) 259, 719
negligible absorption 77
nematic
– birefringence 959
– dielectric anisotropy 959
– phase 935, 1480
– viscosity coefficient 945
nephelauxetic effect 66
Nernst effect 1236
Nernst equation 258
net electric dipole moment 222
neutral-impurity scattering 496
neutrality condition 489
neutron diffraction 251
neutron transmutation doping (NTD) 298
n-FET drive current 617
Ni 635
NiSi 635
nitride
– alloy 782
– crystal structure 745
– electrical property 757
– lattice parameter 746
– light emitter 1048
– nanorod 799
– optical property 769
– semiconductor 829
– thermal property 752
nitriding 664
nitrogen dioxide (NO_2)
– sensor 1291
nitrogen doped crystal 125
N–N dimer 123
NO complex 125
noise equivalent photons (NEPh) 891
noise equivalent temperature difference (NETD) 875, 1480
nonabsorbing film 684
non-alloyed Ohmic contact 187
noncrystalline material 696

nondegenerate material 697
nonepitaxial film 670
nonlinear
– absorption 1091
– directional coupler 1480
– Fabry–Pérot–interferometer 1480
– figure of merit 1480
– inductor 91
– KKR 1086
– Kramers–Kronig–transformation 1480
– Mach–Zehnder modulator 1480
– material figures of merit 1480
– optical medium 1480
– optical switching 1480
– periodic structure 1480
– phase shift 1480
– polarization 1087, 1480
nonmagnetic (NM) metal 23
non-oxide compounds 659
nonradiative
– decay 1073
– process 998
– recombination 559, 910
nonreversing heat flow (NHF) 447
non-stoichiometric oxide 1234
non-stoichiometry 1480
nonthermal energy exploration telescope (NeXT) 863
nonvolatile information, thin film 605
nonvolatile memory device, dielectrics 637
Nordheim coefficients 27
Nordheim's rule 26
normal hydrogen electrode (NHE) 1411
normal–insulator–superconductor (N-I-S)
– current–voltage characteristic 1231
– junction 1231
n-Si
– extinction coefficient 517
– refractive index 517
n-type conductivity 1480
n-type HgCdTe 872
n-type silicon
– electron mobility 500
– neutron transmutation doping 298
– phosphine (PH_3) 298
n-type TCO
– material design 1396
n-type-channel metal-oxide-semiconductor (NMOS) 617
nuclear magnetic resonance (NMR) 560, 987
nucleation
– activation energy 446
– field 1210
– model 579
– temperature 117, 1480
numerical aperture 1480

O

oblique incidence reflectivity 72
octahedral ferroelectric 597
off-chip interconnect 1322
Ohm mobility 498
Ohmic contact 976
Ohm's law 105, 1391
oil-free pumping 331
Onsager model 1480
Onsager relations 947
opaline lattice 1024
open circuit potential (OCV) 1415
open-circuit voltage 1335
optic 393
optical
– absorption 152, 567, 684
– absorption coefficient 154, 1480
– absorption edge 567
– activity 1358, 1371
– amplification gain 922
– amplifier 922
– axis 936
– bistability 1480
– communication 929
– constant model 518
– constant, Drude approximation 518
– constants 1481
– device, figures-of-merit 1086
– dielectric constant 182, 737
– emission spectroscopy (OES) 574, 1481
– emission spectroscopy (OES), glow discharge 416
– feedback 906
– gain 902
– gap 60, 562, 563, 569, 1118, 1481
– glass 50, 80
– integrated circuit 1087, 1088
– interference fringes 157
– limiter 1481
– lithography 1169
– magnetic mirror 1368
– material 79, 81
– memory device 557
– mode 730, 737
– nonlinearity 1085
– phonon scattering 28
– poling 1088
– pumping 998
– rectification 1345
– reflectance 70
– scattering 161
– sensor 1292, 1299
– system integration 1018
– texture 936, 1481
– thickness 682
– transition matrix 59
– transmission 157
– transmittance 69
optical fiber 2
– attenuation 1087
– material 1087
– sensor 1299
optical phonon scattering rate
– GaN 838
optical property 47, 557, 647, 680, 1481
– aluminum nitride 777
– a-Si:H 563
– chalcogenide glass 564
– SiGe 530
– thin film 680
optically
– detected magnetic resonance (ODMR) 559, 789, 1481
– pumped surface-emitting laser (OPSEL) 919
– stimulated luminescence (OSL) 1004
optoelectronic device 455, 726
– detector 927
– semiconductor 927
optoelectronic material 309
orbital angular momentum quenching 100
order parameter 940, 964, 1481
order–disorder
– ferroelectric 230
– transition 230
order–disorder model
– ferroelectric behaviour 592

organic
– glasses 194, 211
– laminant 1314
– light-emitting diode (OLED) 991, 997, 1111, 1264
– perovskite 1135
– photovoltaic device (OPV) 991
– solvent sensitivity 1295
– substrate module circuit 1323
– tandem cell 1336
– thin film material 698
– thin layer, dielectric constant 1293
organic photoconductor (OPC) 1481
– architecture 980
– function 980
– material 980
organic photoreceptor 967, 988, 1481
– charge-transport layer (CTL) 988
organic semiconductor 168, 170, 1001, 1330
– carrier concentration 1117
– film 1329
– molecule 1332
– zero-field mobility 212
organic solar cell 1330
– polymer system 1331
– small molecular system 1331
organometallic vapor phase epitaxy (OMVPE) 319, 886
organo-silicate glass (OSG) 640
orientational order 961
oriented crystal 347
oscillation
– atomic 223
– frequency, transition energy 224
– molecular 223
oscillator
– damping 222
– equation of motion 222
– strength 1481
OSF formation 128
OSF ring width 129
out-diffusion 622
out-of-plane coherence length 1239
oval defect 333
overcoat
– layer 989
– polyurethane 989
– silsesquioxane 989
overlap zone-melting (OZM) 861
Ovshinsky 1149
oxidation kinetics 628
oxidation state 626
oxidation-induced stacking fault (OSF) 113, 1481
oxide
– capacitance 466
– crystals 271
– electronic structure 1396
– film 659
– high-temperature superconductor 1391
– media 1212
– nanostructure 1170
– particle 122
– thermal 627
– trap density 475
oxide glass 51
– empirical rules 430
oxide/fluoride growth 318
oxide/nitride (ON) dielectric stacks 637
oxide/nitride/oxide (ONO) dielectric stacks 637
oxygen
– ion conductivity 254
– precipitates 128, 1481
– sensor 259, 1283
– separating membrane 1418
– vacancy concentration 249
ozone monitoring 1282

P

P3HT 156
P_b center 628
packaging 717
packaging system, conjugate 1312
paraelectric 1481
– phase 591, 597
paramagnetism 96, 97
parameter
– a 746
– S 792
parasitic
– capacitance 618, 623
– resistance 623
parity selection rules 65
partial response maximum likelihood (PRML) 1200
– recording channel 1481
particulate matter (PM) 1282
parts per million (ppm) 719
parylene 640
passivation 630, 1481
passivation layer, thick film 713
passive electronic component, thick film technology 714
passive matrix addressing 1481
Pauli
– exclusion principle 1260
– paramagnetism 101
– principle 1231
Pauling
– unit 181
Payne–Lacey model 1063
$Pb(Sc_{0.5}Ta_{0.5})O_3$ (PST) 613
Pb-free soft error 1321
Pb-free solder 1321
PC fabrication radiation patterning 1018
PCBM 158
PCM alloy 1150
PDMS stamp 1270
peeling tape 1344
Peierls distortion 1151, 1166
Peltier cooling 871
Peltier effect 1379
penetration depth (λ_L) 1239, 1481
– cuprate superconductor 1240
– iron-based superconductor 1252
– MgB_2 1247
percolation 1334
– distance 208
– mechanism 135
– parameter 214
perfect
– diamagnetism (Meissner effect) 1481
– lens 1357
– metamaterial absorber 1368
– silicon 111
perimeter-bonded die 1314
permalloy 1189, 1481
permeation rate 261
permittivity 620, 1481
– of free space 220
– relative 48, 220
perovskite ($CaTiO_3$) 250, 593, 639, 1405
– chemical element 1407
– ferroelectricity 1409
– photocatalytic activity 1411
– superconductivity 1409
– superstructure 1407
perovskite oxide 1405
– catalytic activity 1410
– property 1409

– proton conductivity 1416
– solid oxide fuel cell 1413
– tolerance factor 1406
perpendicular magnetic recording (PMR) 1186, 1205, 1482
perpendicular recording 1205
– development 1213
– media 1208
– system 1205
perturbation Hamiltonian, time-dependent 1000
p-FET drive current 617
phase
– equilibria 344
– lag 162
– matching 1082
– sensitive detection 1005
– transformation 441
phase-change
– alloy ternary phase diagram 1159
– material 1158
– material, neuromorphic computing 1160
– memory (PCM) 1149
– process mechanism 1156
– random access memory (PRAM) 1149
phase-separation 1334
phonon 47, 1342, 1482
– concentration 426, 427
– energy 1074
– optical 634
phonon scattering 634
– non-polar 28
– polar 28
– temperature dependence 28
phonon-glass-electron crystal (PGEC) 1380, 1385
phonon–phonon
– anharmonic interaction 432
– scattering 432
phosphate 288, 289
phosphor
– luminescence efficiency 1009
– optical display 1009
– persistent 1007
– strip 1007
– x-ray storage 1008
phosphorescence 997, 1482
phosphorus-based material 314
photoactive layer 1337
photoactive region 1332
photo-assisted amorphisation 1156
photo-catalysis 1482
photochromic 1090, 1482
photoconduction threshold 978
photoconductive (PC)
– array 874
– detector 871
– gain 871
– switch 1344
photoconductivity (PC) 151, 152, 567, 984, 1482
– layer 982
– material 349
– photoreceptor 984
– semiconductor 208
photoconductor 967
– array, detectivity 875
– loop 990
– organic (OPC) 967
– tribocharged 968
photocurrent 875, 1084, 1482
– decay 167
– generation 1333
photodarkening 1090, 1482
photodetector 726, 739, 898, 927, 1050
– avalanche 928
– photoconductive 927
– requirement 927
photodiode array
– manufacturing 879
– performance 875
photodiode junctions 880
photodischarge 979
– rate 979
photoemission spectroscopy (PES) 180
– technique 1482
photoexcited carrier 359
photoexcited transport molecule 972
photogeneration
– coating 977
– fluorescence quenching 979
– quantum efficiency 978, 979
– sensitized 978
photoinduced
– absorption (PA) 559
– anisotropy 1090
– charge generation 970
– current transient spectroscopy (PICTS) 165
– degradation 582, 1482
– discharge (PIDC) 979
– discharge characteristic (PIDC) 1482
– effects 1482
– electron transfer 1332
photoisomerization 1482
photolithography 1018
– serial patterning 1019
photoluminescence (PL) 140, 152, 371, 547, 562, 770, 861, 997, 1118, 1482
– amorphous semiconductors 562
– detector 1005
– long term 1005
– measuring system 1005
– short term 1005
– TlInGaAs 547
photon 1482
– density 906
photon-confinement 908
photonic
– band gap (PBG) 1013, 1014
– device 1013
– glasses, optical nonlinearity 1081, 1482
– lantern 1077
photonic crystal (PC) 1081, 1346, 1351
– alumina 1021
– architecture 1015
– cell structure 1015
– design limitation 1017
– device 1013
– electrochemical etching 1020
– fabrication method 1021
– fabrication methodology 1017
– lithographic fabrication 1022
– lithography fabrication 1018
– magnetic 1029
– material, optical property 1016
– materials and criteria 1016
– nanometer-scale 1017
– optical functionality 1024
– optical response 1027
– patterning 1019
– PBG structure 1017
– physical architecture 1016
– physical structure 1028
– self-ordered porous 1021
– semiconductor-based 1027, 1028
– structure 1014, 1017, 1021
– structure 3-D PBG 1022
– tunable 1027
– tuning 1028
photonic crystal fabrication
– charged particle 1018
– ion-beam technique 1021

– mask processing 1019
– self-assembly methods 1024
photonics 1340, 1347
photon-trapping structure (PTS) 878
photoreceptor 979, 983, 984, 990
– amorphous semiconductor 970
– architecture 980
– back-coat polymer 989
– characterization 978
– charge generation 971
– charge transport 971, 981
– charge transport layer 988
– charge-generation layer 983
– chemical resistance 991
– conductive layer 982, 990
– dark conductivity 970
– dark decay 971, 979
– dark injection 983
– device architecture 967
– dip coating 990
– discharge 979
– drift mobility 972
– drum 990
– dual layer 971
– dye-polymer aggregate 984
– electrical characteristic 967
– electrical uniformity 991
– electrical-only cycling 980
– electron transport 981
– electrophotographic 980
– fabrication 990
– hopper coating 982
– inorganic material 970
– Langevan recombination 980
– layer 981
– mechanical strength 991
– organic 967
– overcoat layer 989
– photoconductivity 984
– photodischarge rate 991
– photosensitivity 991
– PVK–TNF charge transfer complex 984
– quantum efficiency 971
– spatial uniformity 991
– surface charge 970
– surface charge injection 971
– thickness 980
– web coating 990
– xerographic 983
photoreceptor exposure
– LED array 983
– scanned laser 983
photoreflectance (PR) 770
photorefractive beam fanning 1482
photosensitivity 979, 1482
photosensor
– screen printed 719
– thick film 719
photo-stimulated luminescence (PSL) 1004, 1008
photothermal deflection spectroscopy (PDS) 559
photothermal ionization spectroscopy (PTIS) 1482
phototransistor 928
photovoltaic (PV) 316, 1097, 1257
– array 873
– drift current 105
– HgCdTe array 892
– module 1482
– solar cell 1483
photovoltaic device 720, 1482
– current–voltage characteristics 874
– ideal 874
phthalocyanine (Pc) 1291, 1296, 1330
– film 693
– polymorph 987
– polymorphic 1289
– rare-earth metal 1290
– thin film 1295
physical vapor deposition (PVD) 372, 647, 856, 858
physical vapor transport (PVT) 280, 365, 792
pi bond 1259
piezoelectric
– actuator 608, 1030
– ceramic 595
– coefficient 595
– device 609
– effect 1030
– elastic properties 609
– material 608
– polarization 1483
– sensor 608
– transducer 270
piezoelectricity 1483
piezoresistance 1118, 1119, 1483
piezoresistive sensor 718
pigment generation material
– bisazo 984
– perylene 984
– phthalocyanine 984
– squarylium 984
– trisazo 984
pigment material 977
pinned film 1483
pinned layer 1193
pinning center 1243
pinning efficiency 1244
Piper–Polich 277, 284, 376
pitch helix 937
pixel size 969
pixie dust 1203, 1483
planar
– chirality 1373
– diode 875
– lightwave circuit (PLC) 1013, 1060, 1483
– metamaterial 1365
Planck's law 901
planetary reactor 324
plasma 1342
– display panel (PDP) 582, 1009
– frequency 55, 519, 1394
plasma-assisted molecular beam epitaxy (PAMBE) 744
plasma-enhanced chemical vapor deposition (PECVD) 573, 665, 1069, 1112
– technique 1483
plasma-enhanced milling 880
plasmon 1090, 1483
plastic
– crystal 236
– deformation 1483
– electronics 1483
– encapsulation 1324
– optical fiber (POF) 916
plating 1190
– impurity 1316
platinum resistance thermometer (PRT) 719
platinum thick film 711
p-n diode structure
– *I–V* characteristic 470
p-n junction 137
– reverse-biased 908
p-n-control 1337
pocket calculator 582
point defect 628, 1483
– electrically active 477
– surface 138
point group 936
point of zero charge (PZC) 667
point source 650
Poisson statistics 1211

Poisson's equation 1483
Poisson's ratio (ν) 749, 1483
– semiconductor 730
polar crystal 589
polar dielectrics 590
polarisation 220
– microscopic origin 222
polarity 399, 1483
polarizability 48, 619, 634
polarization 69, 945, 1082
– insensitive modulator 924
– insensitivity 916, 924
– ratio 1217, 1483
polaron 1342
– model 975
pole tip 1205
pole write head 1212
poles of a write head 1483
poling 604, 1088
pollutant gas detection 1282
pollution monitoring 1282
poly(methylphenylsilylene) (PMPS) 977
– glass transition temperature 975
poly(N-vinyl carbazole) PVK 984
poly(vinyl butyral) 986
polyacetylene 1260
polyamorphous silicon 163
polyaniline (PAni) 1281
– conductivity 1298
polycrystalline
– alloy 436
– CIGS 170
– conductivity σ 31
– dielectric 601
– film 31, 32, 674
– growth 319
– interface state 473
– material 30
– piezoelectric 601
– pyroelectric 601
– silicon 1118
– silicon emitter 475
– solid 1483
– thin film 672
– vapour-deposited 673
polyethylene 1259
– crystallinity 434
polygonization 1483
polyimide 640
polymer 1082, 1088
– blend composition 434
– conducting 1297, 1298
– conducting, physical property 1281
– conjugated 211, 974
– effective thermal conductivity 434
– electrolyte membrane (PEM) 1301
– glass-transformation kinetics 447
– heat capacity 428
– molecularly doped 972
– substrates thick film 710
– tandem cell 1336
– thick film 708, 713
– thick film conductor 713
– thick films 1484
– transistor 474
polymeric
– composite thermal conductivity 434
– ferroelectric polyvinylidene fluoride 599
– ion-selective electrode 1299
– membrane 1299
– sensor 1300
polymorphous silicon 163
polypropylene structure 429
polysilicon
– emitter bipolar junction transistor (PE BJT) 454
– granular 296
– production 296
– randomly stacked 296
– seeding 297
polytype control (SiC) 279
polyvinylcarbazole (PVK) 211
polyvinylidene fluoride (PVDF) 599, 718
Poole–Frenkel
– coefficient 1137
– effect 695
– emission 694
population inversion 902
porous array
– domain size 1021
– pore shape 1021
porous material 1024
porphyrin, synthetic 1290
positive temperature coefficient (PTC) 719
– of resistance (PTCR) 591
– thermistor 719
positron annihilation 1484
– spectroscopy (PAS) 790
positron emission tomography (PET) 1010
postgrowth heat treatment 344
potassium dihydrogen phosphate (KDP) 592
power
– consumption 616
– dissipation 474
– miniaturization 248
– source 248
– stabilizer 1484
– supply voltage 619
power-law dependence 154
precipitates 138
precision doping 299
precursor 321, 665, 1484
– concentration 321
– diethyl telluride (DETe) 353
pressure sensors 718
primary
– bombarding particles 423
– ion species 420
– photocurrent 153
principal angle 71
principal component analysis (PCA) 1299
printed capacitor 714
printed circuit board (PCB) 708, 1060
printer, liquid-toner-based 989
process integration 1484
producible alternative to CdTe epitaxy (PACE-I) 357
profiling techniques 139
proper ferroelectric 1484
protective overcoat layer 981
proton conductivity 255
proton implantation 1088
pseudogap (cuprate superconductor) 1015, 1235–1237
pseudomorphic layer 1484
p-type
– conductivity 1484
– HgCdTe 877
– MOS (PMOS) 617
– silicon 298
– TCO, electronic structure 1400
– TCO, material design 1398
pulse width 1204, 1484
pulsed dye laser 166
pulsed laser deposition (PLD) 744, 1089
purified Si 488
push-rod arrangement 312

PV detector technologies 352
PVK 977
Pyrex glass 1088
pyrochlore 250
pyroelectric
– coefficient 595
– device, radiation detector 612
– effect 611, 1484
pyrolytic boron nitride (pBN) 1484
PZT ceramic 610

Q

QCL 1343
q-DC behaviour 241
– characteristic frequency 241
– equation of motion 242
– isolated response 241
– sub-infinite percolation 242
– suceptibility function 241
– transport path scaling 242
QLEDs 1009
quadratic Stark effect 1044
quantum
– ballistic transport 42
– cascade laser (QCL) 899, 920, 1037, 1049, 1484
– conductance 1167
– confined stark effect 925
– confinement 899, 923, 1038
– efficiency (QE) 800, 869, 874, 977, 1484
– Hall effect (QHE) 462
– information processing 1055
– interference, Josephson junctions 1232
– size, amorphous semiconducting multilayers 569
– wire 42, 557, 1039, 1053, 1485
quantum dot (QD) 325, 336, 374, 557, 796, 899, 1007, 1038, 1039, 1053, 1484
– array 87
– infrared photodetector (QDIP) 1055
– laser 1054
– layer 913
– photoluminescent 1007
quantum well (QW) 10, 68, 330, 374, 547, 557, 569, 794, 899, 903, 1037, 1039, 1484
– absorption 923, 1042
– avalanche photodiode 1050
– device 314
– emission 1044, 1047
– exciton 1043
– finite 1041
– GaAs/AlGaAs 1039
– infinite 1040
– infrared photodetector (QWIP) 530
– intersubband photodetector (QWIP) 1051, 1485
– laser 893, 1047
– light modulator 1037, 1052
– optoelectronic property 1037
– Schrödinger equation 1040
– solar cell 1037, 1050
– vertical transport 1046
quantum-confined
– Stark effect (QCSE) 923, 1044, 1052, 1484
– structure classification 1039
quantum-mechanical
– confinement 623
– tunneling 619
quantum-size effect 1484
quarter-wave stack 684
quartz 277, 287
– crystal microbalance (QCM) 1295
quasi particles 1485
quasi-Fermi level (QFL) 155, 902, 1485
quasi-static measurements 464
quaternary
– alloy 739
– binary parameters 726
– composition parameters 726
– Hall mobility 741
– optical modes 731
– unit cells 731
quench anneal (QA) 283, 345

R

radial distribution function (RDF) 560
radiation
– detector 612
– patterning 1018
– resistance 1485
radiative
– current 912
– decay 1073
– process 998
– recombination 559, 563, 910
Radio Corporation of America (RCA) 956
radio frequency (RF) 373, 576, 584, 605, 636, 661, 744, 1311
– circuit flip chip 1320
– circuit interconnect 1320
– circuits 1485
– device packaging material 1315
– digital wirebond 1317
– magnetron sputtering 606
– packaging substrate 1322
– packaging, thermal issues 1327
– wirebond material 1317
radioactive Pb^{210} 1321
radioluminescence 997
RAM memory 85
Raman
– amplifiers 1087
– crystallinity 580
– scattering 679
– scattering, stimulated 1087
– shift 679
– spectroscopy 678, 1179
random telegraph signal (RTS) 475, 879
randomly oriented particles 33
randomly stacked polysilicon, Siemens method 296
rapid slider LPE 315
rapid thermal annealing (RTA) 113, 137
rare earth (RE) 66
– doping 315
– ion 66
– spectroscopy 1005
RC delay 639
– time response 617
reaction kinetics, MOCVD 319
reactive
– evaporation 653
– ion beam etching (RIBE) 405
– sputtering 573, 659, 1485
reactor cell 322, 1485
– design 354
reactor planetary 324
read head 1485
read pulse 1194
read sensor 1187
readout integrated circuit (ROIC) 875, 1485
receptor technology 972
rechargeability 261

reciprocal lattice 1485
recombination 151, 1485
– bimolecular 154
– -enhanced diffusion 138
– monomolecular 154
– nonradiative 910
– process 208
– radiative 559, 563
recrystallization 347
rectangular columnar phase 939
reduced
– dimensionality 903
– Fermi energy 1382
– graphene oxide (RGO) 1174, 1285
– interface potential 178
re-entrant phase 940
refactive index 75
reflectance 48, 69, 682
– -transmittance (R–T) method 1485
reflection
– adsorption infrared spectroscopy (RAIRS) 1288
– coating 684
– difference spectroscopy (RDS) 358
– electron diffraction (RED) 401
– high-energy electron diffraction (RED) 401
– high-energy electron diffraction (RHEED) 330, 386, 546, 1485
reflectometer 1485
reflow 1485
– soldering 716
refractive index 47, 48, 51, 73, 81, 683
– average 51
– complex frequency-dependent 221
– intensity-dependent 1084, 1085
– oxide glass 51
– semiempirical 1084
– tuning 1027
– units (RIU) 1301
refractometry 74
refractory compound 247
refractory layer 989
relative
– dielectric permittivity 619
– humidity (RH) 1299
– permittivity 48, 220, 1331
– sensitivity 422
– susceptibility, frequency-dependent 223
relaxation
– defect diffusion 234
– ensemble 243
– ferroelectric 231
– frequency 233, 236
– interactions 243
– response 224, 226, 228, 231
relaxation rate 232
– activation entropy 235
– correlation length 236
relaxation time
– approximation 29, 491
– constraint hierarchy 238
– dipole density fluctuation 232
– distribution 237
– distribution shape 238
– extreme value statistics 238
– thermal activation 232
– viscosity 232
reliability of dielectrics 631
remaining challenges in nitrides 337
remanent magnetization 90, 1485
remanent polarization 1485
remote impurity scattering 41
residual attenuation 1087
residual potential 972
residual resistivity 23
– metal 1225
– ratio (RRR) 861
resistance
– contact 459
– metal-semiconductor 459
– modulation 1118
– × area (RA) 1218
– temperature device (RTD) 350
resistive transition
– broadening 1240
resistivity 25, 32, 462, 1198, 1485
– effective 23
– index 24
– longitudinal 43
– measurement 495
– mixture rule 34
– residual 23
– strain-induced 1119
– temperature coefficient 25
– Van der Pauw technique 457
resistor trimming 713, 715
resistor, thick film 712
resolution 1485
resonance response 223
resonant
– cavity light emitting diode (RCLED) 916, 1049
– loss amplification 1368
– nonlinear response 1485
– sensor 718
resonator, micromachined 1030
response function 227
– amorphous Gallium Arsenide 227
responsivity 1485
reststrahlen band 737
retina level processing 892
retrograde solidus 1485
reversing heat flow (RHF) 447
RHEED oscillation 331
Richardson constant 177, 1138
Richardson formula 1046
ring coating 982
ring write head 1205
Ritter–Weiser formula 157
RKKY coupling 99
Rochelle salt 590
rocksalt structure 745
rod-like molecules 945
root mean square (RMS) 1063
rotating substrate 330
rotational viscosity 964
– coefficient 946, 950
route formation 1334
Ruddelsden–Popper
– compound 1409
– structure 1408
Ruderman–Kittel–Kasuya–Yoshida (RKKY) 99
Ruthenium 1196
Rutherford backscattering spectrometry (RBS) 386, 546, 782
Rutherford scattering 1485

S

SAM film 1288
sample preparation 404
sapphire (Al_2O_3) 12, 288, 336
– substrate 357, 886
saturable absorber 64
saturated
– calomel electrode (SCE) 977
– vapor pressure (SVP) 321, 1486

saturation
– flux density 1189
– intensity 1486
– magnetization 88, 1486
– of absorption 1486
– parameter 64
– polarisation 591
scaling 1486
– law 1214
– of integrated circuits 615
– targets 616
scanning
– electron microscopy (SEM) 10, 359, 385, 560, 1285
– spreading resistance microscopy (SSRM) 458
– transmission electron beam 409
– transmission electron beam induced conductivity (STEBIC) 403
scanning probe 89
– microscopy 1486
scanning–tunnelling microscopy (STM) 646
scattering
– carrier–carrier 493
– carrier–lattice 492
– cross section 21
– mechanism 464
Schiff's base 956
Schott glass code 80
Schottky
– defect 135
– defect pair 249
– diode, Ni-a-Si:H 583
– emission 698
– formation energy 250
Schottky barrier 691, 790
– atomic interlayers 189
– H-induced modification 189
– lowering 1138
Schottky contact 175, 1486
– barrier height 183
– C/V characteristics 178
– graphene 190
– I/V characteristic 177
– ideal 177
– laterally homogeneous 178
– MoS_2 190
– patchy 178
– real 177
Schottky–Mott rule 175, 1486
Schroeder van Laar equation 958
schubweg 1127, 1140
scintillator 1010
SCLC mechanism 694
screen printing (SP) 707, 1486
– polymer thick films 713
s–d hybridization 102
secondary
– electron (SE) 1486
– ion mass spectrometry (SIMS) 138, 140, 351, 386, 413, 1486
– ion yields 421
– photocurrent 153
Seebeck coefficient 1380
– intrinsic semiconductor 1381
Seebeck effect 719, 1379
seed crystal 377
seeded chemical vapor transport (SCVT) 377
seeded physical vapor transport (SPVT) 377
segregation 137, 302, 311
– behavior 274
– coefficient 274
– coefficient (k) 1486
– effect 347
– of components 660
selection rule 1486
selective epitaxy 1486
selective LPEE 315
selective-area vapor-phase epitaxy (SA-VPE) 1053
selenide glass (SG) 1060
selenium 969
selenization 1486
self energy, hierarchical modes 239
self-assembled monolayer (SAM) 1288
self-assembly (SA) 1017, 1286, 1486
– inverse opal 1024
self-defocusing 1486
self-diffusion 136, 143
– dopants 141
self-focusing 1486
self-interstitials 141
– concentration 135
– semiconductor 135
self-organized process 726
self-scanned electronic readout 1126
self-similar correlation 236
Sellmeier equation 50
semiconduction
– CdTe film 693
– filmelectronic properties 1112
– layer 456
– nanotube 1168
– thin film 690
semiconductor 2, 22, 51, 59, 271, 275, 327, 486, 714, 829, 1016, 1165
– alloy 380, 726, 732
– amorphous 59
– amorphous organic 1117
– analysis 413
– band gap 477
– band picture 1085
– binary compound 135
– carrier scattering mechanisms 739
– characterization 413
– component 462
– conductivity 685
– crystal 1085
– crystalline 38, 133
– density of states 195
– detector 612, 1486
– device 453, 685, 898
– die 1315
– diffusion 133
– direct-band gap 900
– electrical characterization 454
– electron transport 830
– film 663, 1284
– free carrier 455
– gain 902
– gauges 1120
– glass 194
– group III–V 144
– group IV 144
– Hall effect 462
– impurity 62
– indirect-band gap 486, 900
– industry 293
– infrared detector 869
– interband transition region 739
– laser 897
– laser, current confinement 908
– laser, radiative recombination 910
– loss 902
– material 455
– microelectronics 615
– mobility 462
– modulator 923
– nanocrystals 1486
– nondegenerate 22
– n-type piezo-electric 378

– optical amplifier (SOA) 899, 922, 1072
– optical properties 1085
– permittivity 460
– properties 569
– pseudo-binary (CdTe–ZnTe) 380
– quantum well 923, 1037
– relative permittivity 22
– resistivity 454, 455
– silicon 294
– solid solution 1384
– specific heat 756
– spontaneous emission 900
– spontaneous ordering 726
– tetrahedral 904
– thermal conductivity 433
– thermal properties 732
– traveling heater method 380
– triode 3
– wafer 456
– work function 690
– wurtzite 730
– zincblende 730
semiconductor heterostructure 175, 180
– band offset 184
semiconductor layer
– analysis 416
– H diffusion 140
– hydrogen 140
– sheet resistivity 455
semiconductor–dielectric interface
– electrical characterization 454
semiconductor-grade silicon (SG-Si) 295
semiconductor–insulator interfaces
– charge pumping 472
semi-insulating 1486
– GaAs 272
– planar buried heterostructure (SIPBH) 908
semimetal 1176
Sendust 1190
sensing
– filament 1298
– mechanism 1300
– membrane 1292
– membrane fabrication 1301
sensitivity 1284
– relative 422
sensor 1281, 1290, 1299, 1300
– amperometric 259
– microelectronics 707
– microstructure glass 1299
– miniaturization 248
– optical absorption 1292
– organic vapor 1303
– oxygen 259
– solid-state 1284
– thick film technology 717
separate confinement heterojunction (SCH) laser 1487
separate confinement heterostructure (SCH) 549, 912
shallow
– impurity 489
– impurity doping 492
– structures 421
– trap 1129
– trap level 692
– trench isolation (STI) 620
shallow-energy-level dopant 1487
shape anisotropy 95
shape forming 602
shear modulus 748, 1487
sheet resistance 455, 461, 1487
shell cloak 1361
shields 1487
shift current 105
shifted-modulation-doped (SMD) 789
Shockley theory 5
Shockley–Queisser limit 1100
Shockley–Read–Hall (S-R-H) 876, 877
short wave (SW) 1487
short-circuit path diffusion 138
short-circuit paths 138
short-range atomic structure 1487
short-wavelength infrared (SWIR) 351
shot noise 875
showerhead reactor 324
Si 278, 279
– conduction-band effective mass 488
– diffusion coefficient of electrons 513
– doping effect 494
– drift velocity 512
– electrical properties 495
– energy levels of impurities 489
– Hall-effect parameter 503
– interstitial 113
– optical absorption spectrum 520
– oxidation kinetics 628
– room-temperature resistivity 497
Si/Ge 279
Si/SiGe quantum structures 530
Si/SiO_2 system, point defects 628
Si:H conduction tail states 59
$Si_{1-x}Ge_x$
– alloy 141
– bandgap 526
– density of states 527
– mobility 527, 528
Si-based transistor 628
SiC 142
– growth on Si 314
SiGe
– growth 534
– heterostructures 142
– hydrogen passivation 534
– in-situ hydrogen bake 534
– optical properties 530
– pre-epitaxy cleaning 534
– quantum wells 530
sigma bond 1259
signage 1007
signal decay rate 1487
signal processing in the element (SPRITE) 1487
signal-to-noise ratio (SNR) 1142, 1201, 1211
SiH_3, enhanced surface diffusion 579
silane (SiH_4) 329, 573
silica glass 561, 625
– PbO 1088
silicon 4, 12, 278, 523, 629
– amorphous 59
– bandgap 293
– carbide (SiC) 2, 12, 142, 313
– crystal 294
– crystal growth, magnetic-field-applied CZ 304
– crystals growth new methods 304
– CZ crystal growth 298
– device electrical performance 1312
– device power dissipation 1312
– dislocation-free 294
– dopants 303
– doping 298
– element 294
– FZ method 297
– growth 305
– hydrochlorination 295
– lattice constant 524
– magnetic-field-applied continuous CZ 305
– materials preparation 294

– metallurgical-grade 295
– microelectronics 1169
– multiplexer 882
– nitride 695
– nitride films 665
– n-type 298
– -on-insulator (SOI) 463, 794, 1063
– optical constants 515
– optical properties 515
– oxynitride (SiON) 631, 1069
– physicochemical characteristics 301
– p-type 298
– semiconductor industry 293
– silica 294
– single-crystal growth 296, 298
– solar cell 1101
– surface capacitance 468
– technology 1284
– tetrachloride 295
silicon-based light emitter 530
silicon-germanium (SiGe) 141, 523
– epitaxial layers 534
– growth 534
– physical properties 524
– selective epitaxy 535
– virtual substrate 528
silicon-on-insulator (SOI) 463, 794, 1063
– substrate 794
silicon-oxynitride (SiON) 631, 1069
silk screen printing 708, 1320
Sillenite $Bi_{12}TiO_{20}$ 170
silver conductor, thick film 711
silver/palladium conductor, thick film 711
SIMS (secondary ion mass spectrometry) 413
– profile 355
– static SIMS (SSIMS) 418
simultaneous unipolar multispectral integrated technology (SUMIT) 888
single crystal 269
– growth 299
– silicon 297
– silicon, CZ growth 299
– silicone, optical constants 515
single electron tunneling 1231
single photon counting 1005
single quantum well (SQW) 374
single-color device 355
single-crystal diamond 170
single-electron density of states 1230
single-layer
– organic photoconductor 1487
– photoreceptor 981
single-mode 913
single-photon
– avalanche diode (SPAD) 1050
– emission computed tomography (SPECT) 1010
– source 1055
single-wall
– carbon nanotube (SWCN) 1263
– nanotube (SWNT) 1163, 1262
sintering 98, 603
Si–O bond stretching vibrations 494
Si–O bonding network 628
SiO_2
– physicochemical properties 628
– quartz 625
– reliability 631
– thermally grown 625
SiON 622, 632, 1069
– film 631
S-I-S junction 1231
$Si–SiO_2$ interface 472
size of crystal 280
skutterudite 1386
slab waveguide 1017
slider 1487
sliding 880
– boat 311, 1487
– boat method 312
– boat process 880
sliding-boat LPE 350
Slonczewski loose-spin 100
slope parameter 1487
slow cooling 276
slush 1487
– growth 345
small outline (SO) 715
small pixel technology 886
small-angle x-ray scattering (SAXS) 560
smart sensor 1299
smart window 262
smectic liquid crystal 963
smectic phase 936, 1487
smoothing layer 983
Snell's law 71, 1356
SNR 1209
sodium nitrite ($NaNO_2$), order–disorder ferroelectric 592
SOFC
– anode 1414
– cathode 1413
– electrolyte 1415
– interconnector 1417
– stack 1417
soft ferrites 97
soft underlayer (SUL) 1205, 1212, 1487
soft-magnetic film, hysteresis 1187
soft-magnetic material 1185
solar cell 557, 573, 720, 1050, 1097, 1487
– a-Si:H-based 582
– μc-Si:H-based 585
– Si 278
solar energy 1098
solder
– alloy 1320
– alternative 1324
– conduction 1325
– dipping 716
– electronic interconnect 1317
– flip chip 1320
– interconnect 1317
– interconnect reliability 1318
– intergranular failure 1319
– intermetallic compound 1319
– paste 1320
– Pb-free product 1320
– Pb-Sn 1320
– plating 1320
– RF circuit 1320
– wetting 1319
solder joint
– flip chip 1319
– microstructure 1318
– strain 1318
soldering 716
sol–gel technique 667, 1089
solid
– electrical behavior 256
– electrolyte 258, 260, 261
– ionic conductor 256, 262
– oxide electrolyte 249
– oxide fuel cell (SOFC) 252, 260, 262, 1409
– oxide fuel cell, perovskite oxide 1413
– polymers 429
– silicon, impurity solubility 302

– source molecular beam epitaxy (SSMBE) 546
– state ionic 247
– state ionic, future trend 262
solid/liquid/gaseous phase equilibria 1487
solid–liquid phase relation 349
solid-source MBE
– effusion cell 333
solid-state recrystallization (SSR) 283, 343, 345, 380, 1487
solidus 1487
solidus–liquidus separation 344
solid–vapor phase relation 349
solitonic propagation of pulses 1487
sonication-assisted liquid exfoliation 1285
sonoluminescence 997
space charge region 256, 1243
space transformer 1312
space-charge-limited current (SCLC) 511, 691
spacer material 99
spacer resistivity 1198
spacing layer 1194
spatial frequency 1487
spatial resolutions 415
Spear method 168
species conservation equation 119
specific heat 1487
– III–V binaries 731
spectral broadening 1087
spectroscopic ellipsometry (SE) 73, 530, 548, 579
spectroscopic response 223
specular optical activity 1372
specularity parameter 687
spherical aberration 1487
spike noise 1213
spike-timing-dependent plasticity 1160
spin
– accumulation 105
– coating 1286
– coating, film thickness 1288
– coherence 105
– current 1218
– density wave (SDW) 1249
– detection 105
– filtering 105
– injection 105
– magnetic energy 43
– manipulation 105
– oscillator 106
– photo-galvanic effect (SPGE) 104
– polarization 102, 103
– pumping 105
– split-off band 911
– valve 1488
spin-based logic circuit 106
spin–charge separation 1167
spin-dependent recombination 171
spinel 98
– structure 1405
spin-Gunn effect (SGE) 104
spin-Hall effect (SHE) 104, 105
spin-Hall magnetoresistance (SMR) 104, 105
spin–orbit coupling 94, 999
spin-Peltier effect (SPE) 104
spin-polarized current (SPC) 104, 1218
spin-Seebeck effect (SSE) 104
spin-spin scattering 1197
spin-transfer torque (STT) 104, 1218
spintronic 104
– effect 104
spin-valve read element 1193
spin-valve read head 1192
spin-wave current (SWC) 104, 1218
spin-wave guiding 106
splay deformation 942
splay elastic constant 943
split-ring resonator (SRR) 1355
spontaneous
– emission (SE) 900, 1365
– emission spectrum 903
– ordering 726
– polarization 589, 590, 1488
spreading resistance technique 454, 456, 458
SPRITE detector 346, 872
– carrier concentration 872
– conductivity modulation 872
sputter deposition 1199
sputter depth profile 1488
sputter-induced roughness 422
sputtering 418, 654, 1488
– asymmetric-AC 660
– bias 660
– chemical information 419
– deposition (SD) 856
– diode 658
– direct current (DC) 660
– getter 660
– glow discharge 659
– high-energy 657
– low-energy 657
– magnetron 661
– non-oxide film 659
– oxide film 659
– radio-frequency (RF) 661
– rate 658
– reactive 573, 659
– target 658
– triode 658
– yield 655, 657
squareness 89
squeegee 1488
SSR crystals 347
stability 1065
Staebler–Wronski effect 567
stainless steel substrate 710
staking fault (SF) 793
stark effect 926
static dielectric constant 501
static disorder 61
static random-access memory (SRAM) 636, 1186
static secondary ion mass spectrometry (SSIMS) 418
steady-state
– electron transport 835, 843
– photocarrier grating (SSPG) 160
– photoconductivity (SSPC) 153
– plasma 574
Stepanov 273
– technique 1488
step-coverage method (SCM) 576
stimulated emission 900
Stirling cryocooler, long term reliability 871
Stirling engine cryocooler 871
stirred melts 274
stoichiometry 344, 345, 1488
Stokes Shift 1488
Stoner two-band model 102
Stoner–Wohlfarth model 90
stop band 1015
storage capacitor, leakage of charge 637
storage time 1209
strain 1488
– compensated 905
– component 1121
– effect 255
– engineering 524
– gauge 1119

– relaxation 374
– tuning 1029
strained bond formation 580
strained quantum well 1041, 1048, 1488
strained-layer technique 923
strain-induced modulation 1122
Stranski–Krastanov (SK) 374
– growth 374
– growth technique 904
– mode 336
– technique 1053
stray magnetization 1212
stress 1488
strong absorption 77
strong anchoring 1488
strontium bismuth tantalate (SBT) 639
structural parameter
– crystal density 727
– lattice parameter 727
– lattice-matching condition 727
– molecular density 727
structural properties 348
structural relaxation 1065, 1488
STT-MRAM 1215
sub-100 nm (CMOS) technology 633
sublimation traveling heater method (STHM) 284
submicron machining 1018
submicron quantum dot 107
suboxide 627
substitutional impurities 1488
substrate 1488
– Al_2O_3 ceramic 1322
– basic types 651
– debiasing 881
– dielectric constant 1322
– embedding of components 1323
– flip chip interconnect 1322
– lattice-matched 879
– LTCC 1322
– material 1322, 1323
– material, underfill 1323
– non-lattice-matched 356
– organic 1322
– orientation 356
– Poisson's ratios 1120
– rotation 332
– temperature 321
subthreshold slope 471
subwavelength imaging 1364
super structure grating 915
super-bandgap 1084
superconducting oxide 1226
superconducting state resistance 1228
superconductivity 1168
– BCS theory 1230
– history 1226
– mechanism 1230
– MgB_2 1245
– organic materials 1226
superconductor
– electrodynamics 1229
– field-cooled 1228
– magnetic behavior 1228
– metallic 1226
– phase diagrams 1233
– type I 1488
– type II 1488
– zero field-cooled 1228
supercontinuum 1082, 1087
supercooled liquid 439, 1061
supercooling temperature 1066, 1488
superlattice (SL) 10, 1037, 1040, 1042, 1198, 1488
– avalanche photodiode (SL-APD) 1037, 1051, 1488
– optoelectronic property 1037
– structure 136
– vertical transport 1046
superlens 1357
superparamagnetism 1185, 1488
supersaturation 120, 312, 376, 671, 1488
supersaturation versus misfit 311
superstrate 1489
superstructure grating (SSG) 915
super-twisted nematic (STN) 951
supramolecular assembly 1289
supramolecule 1282
surface
– absorption 579
– analysis, dopant concentrations 417
– anisotropy 95
– band bending 467
– energy 100
– kinetics 670
– mobility 673
– morphology 317
– passivation 1489
– photovoltage (SPV) 171
– plasmon 1293
– reaction concept 575
– roughness 1063
– scattering 30
– source 650
– tension 1062
– topography 415
surface acoustic wave (SAW) 600, 718
– device 600
surface alignment 941
– director orientation 948
– energy 942
– homeotropic 941
– hybrid 955
– pre-tilt 944, 953
– strong anchoring 942
– uniform planar 941
– weak anchoring 942
surface chemical analysis 413
– ISO TC 201 413
surface mount components 716
surface mount devices (SMDs) 1489
– component 714
– thick film techniques 714
surface plasmon resonance (SPR) 1293
– imaging (SPRI) 1296
susceptibility 64, 241
– frequency dependence 237
– linear 227
– nonlinear 1082
– relative, frequency-dependent 223
– self-similar scaling 237
susceptibility increment 230
– dipole contribution 242
susceptor 1489
Swanepoel's method 77
s-wave symmetry 1235
switching current 1219
switching time 617, 1192
switch-off transient 165
switch-on photocurrent 164
switch-on transient 164
SWNT (single wall nanotube)
– electrical transport 1166
– intrinsic superconductivity 1168
– ropes 1166
synthetic
– antiferromagnet (SAF) 1192, 1195, 1489

– crystal 270
– ferrimagnetic media (SFM) 1202, 1489
system-on-a-chip (SOC) 293

T

T nonlinear figure of merit 1489
TA Instruments 449
tab automated bonding (TAB) 715
tail states 1489
tandem cell 1336
tape-casting 603
target material 658
Tauc gap 1085, 1489
Tauc–Lorentz parameterization 79
Teal–Little (TL) 495
– method 297
Te-based compounds 284
technology node 1489
– cycle time 616
TEM-cathodoluminescence (TEM-CL) 403
temperature coefficient of resistivity (TCR) 25, 686, 711, 1489
temperature coefficient of the refractive index (TCRI) 51
temperature profiles 275
temperature-modulated differential scanning calorimetry (MDSC) 446
template, self-assembled 1021
templating 1017
tensile strain 904, 1489
terahertz (THz) 1347
– component 1346
– detector 1345
– frequency 1339
– gap 1340
– materials 1339
– parametric oscillator 1344
– radiation 1339
– range 1340
– source 1343
terahertz-frequency photons 1340
ternaries
– III–V 730
ternary 282
– alloy mobility 740
– parameter 726
– system 345
ternary and quaternary alloys 1489
tetraethoxysilane (TEOS) 640
tetragonal ferroelectric phase 597
tetrahydrofuran (THF) 605
TFT
– fabrication 1112
– inverted staggered 1116
– longitudinal 1121
thallium bromide 1135
thallium-based compound, layering scheme 1234
theoretical maximum gain 1072
thermal
– analysis 425
– budget 1489
– coefficient of expansion (TCE) 710
– contrast ratio 875
– decomposition 664, 1174
– emission 158, 1046
– energy 1332
– expansion coefficient (TEC) 425, 436, 752, 1490
– expansion coefficient (TEC), group III–V 732
– expansion principle 436
– generation 927
– generation current 1137
– generation rate 877
– imaging 871, 887
– nonlinearities 1490
– poling 1088
– resistivity 1388
– resistivity, group III–V binaries 732
– runaway power load 918
– stability 899, 1209
thermal conductivity 425, 1489
– ceramic 431
– crystalline insulators 432
– crystalline polymers 434
– crystalline semiconductor 433
– glass 431
– glasses 433
– lattice 433
– metal 431, 435
– metal alloy 431
– nanotube 1164
– noncrystalline insulator 433
– polymer 431
– polymeric composite 434
– substrate 710
thermally
– activated delayed fluorescence (TADF) 1265
– driven switching 1202
– grown SiO_2 697
– stimulated current (TSC) spectroscopy 1490
– stimulated luminescence (TSL) 1003
thermionic emission (TE) 177
– current 460
thermistor 719, 1490
– positive temperature coefficient (PTC) 719
thermochromism 937
thermocompression 716
– bonding 716
thermodynamic melting temperature 438
thermoelectric energy conversion 1380
thermoelectric material, semiconductor 1381
thermoelectricity 1379
thermoluminescence (TL) 1003
thermomechanical fatigue 1318, 1490
thermo-photovoltaic (TPV) 325, 1490
thermoplastic polymer
– adhesive 1324
– glass-transition temperature 1324
– thick film 713
thermosetting
– material 1325
– polymer, thick film 713
thermosonic bonding 716
thermotropic 1490
thick CdF_2 layer 318
thick film 604, 687, 707, 710, 711, 1284, 1490
– chemical sensor 720
– component, resistor 715
– component, tolerance 715
– conductor, characteristics 711
– conductor, sheet resistivity 711
– copper 711
– definition 707
– dielectric paste 713
– fabrication 707
– photoconductor 719
– piezoelectrics 718
– platinum 711
– platinum conductor 718
– polymer 713
– resistor material 712
– resistor paste 712
– solar cell 720

– substrate 710
– technology 708, 714
– technology, printing 708
thick film hybrid circuit
– active components 714
– manufacturing 707
– packaging 717
thin dielectric film 690
thin film 47, 151, 645, 685
– amorphous 674
– antireflection coating 680
– deposition 158, 647
– deposition method 646
– deposition variable 669
– dielectric 654
– electrical properties 646, 685
– Fresnel coefficient 681
– fundamental properties 680
– metallic 674
– mirror 680
– morphological characteristic 676
– nanostructure 680
– optical coefficients (constants) 681
– optical properties 646, 680
– optical thickness 682
– optics 75, 680
– organic 1289
– phthalocyanine 1295
– polycrystalline 674
– properties 646, 669
– reflectance 682
– resistivity 31
– roughness 161
– semiconductor 1111
– sputtering 654
– stack transmission properties 683
– structure 669, 670
– technology 645
– transducer 1118
– transistor (TFT) 2, 557, 573, 582, 952, 1111, 1125, 1126
thin metallic films 685, 688
thin organic film 1286
– island density 1286
third-harmonic generation 1087, 1088
third-order nonlinearities 1089
THM MCT 286
THM technique 285
Thomas–Fermi screening 1243
threading dislocation (TD) 793, 1490
threading dislocation density (TDD) 537
three-dimensional atom probe microscopy (3DAP) 10
three-dimensional lattice 426
three-valley Monte Carlo simulation 830
three-zone model, microstructure 677
threshold current, characteristic temperature 913
threshold electric field 943
threshold voltage (V_T) 633
threshold voltage (V_{th}) 470, 1116, 1122, 1490
– modulation 1122
Tien model 1063
tilted anisotropy 1211
tilted perpendicular recording 1211
time division multiplexing 1490
time-delay and integration (TDI) 873
time-dependent dielectric breakdown (TDDB) 631
time-domain charge measurement (TDCM) 1490
time-of-flight (TOF) 165, 168, 564, 972, 973
– electron mobility 207
– SIMS 418
– SIMS (ToFSIMS) 418
time-resolved microwave conductivity (TRMC) 171
TiO_2 168, 172
tipping 880, 1490
– boat 311
– technique 349
Tl (thallium)-based III-V ternary alloy semiconductor 544
Tl-III-As ternary alloy growth 546
Tl-III-P ternary alloy growth 546
Tl-III-Sb ternary alloy growth 546
Tl-III-V alloy growth 546
Tl-III-V alloy semiconductor 544
TlInGaAs 544
– band-gap energy 548
– photoluminescence 547
– refractive index 548
TlInGaAsN 544, 545
– electroluminescence 549
– photoluminescence 549
TlInGaP 544
TOF mobility 168
toner 969
Tool–Narayanaswamy–Moynihan (TNM) 441
topological insulator 1159
total heat flow 447
track density 1210
track width 1206
trailing-edge shield 1207
transconductance 470, 630
transfer reaction 664
transferred electron effect 1490
transformation optics 1361
transient
– enhanced diffusion (TED) 137
– photoconductivity (TPC) 166
– photocurrent spectroscopy 166
transient electron transport 834, 847
– characteristics 845
– drift velocity 843
transimpedance pre-amplifier 166
transistor 454, 714, 1490
– channel scaling 620
– design 3
– design technology, switching time 617
– electrical characteristics 617
– FinFET 621
– gate dielectric, industry scaling 622
– gate length 619
– non-planar structure 620
– performance 622
– planar structure 620
– scaling 618
– trigate 621
transit time 1490
transition
– dipole 223
– insulator–metal 489
– length 1203
– metal 100
– probabilities, relative 1000
– region 1187
– region entropy 233
– shape 1210
transition temperature 936
– liquid crystal–isotropic 940
– nematic-isotropic 936
transition-metal ions 1000
transmission coefficient 1217
transmission electron diffraction (TED) 393

transmission electron microscopy (TEM) 10, 115, 385, 560, 579, 676, 1076, 1178, 1200
transmission line measurement (TLM) 454
transmission line model (TLM) 460
– test structure 461
transmittance 69
transparent
– conducting oxide (TCO) 328, 855, 1104, 1391, 1490
– connection layer 1337
– film 683
– layer 683
– oxide semiconductor (TOS) 1392
transport
– agent 375
– chamber 358
– currents 1243
– energy 205, 213
– phenomena, low-dimensional 1167
transport coefficient 202, 211
– field dependence 211
transport material
– oxidation potential 988
– reduction potential 989
transport-limited growth 324
transverse
– acoustical (TA) 491
– bias 1195
– electric (TE) 905, 1303
– magnetic (TM) 899, 1293, 1303
– magnetic bias film 1490
transverse optical (TO) 53, 491, 532, 561, 731, 750
– modes 731
trap 624
– concentration 693
– distribution 1117
– fixed charge 624
– level 692
trap-assisted tunneling (TAT) 878
TRAPATT diode 514
trap-free space-charge-limited current density (TFSCLC) 976
trap-free transport 976
trap-limited conduction (TLC) 1118
trapped carrier density 1114
trapped plasma avalanche-triggered transit (TRAPATT) 513
trapping centers 1490
traveling heater method (THM) 275, 284, 343, 347, 380, 861, 1491
triangular lattice 1015
triboluminescence 997
trichlorosilane ($SiHCl_3$) 295
trimming 1491
triode sputtering 659
triphenylenes 940
triple phase boundary (TPB) 1413
tri-*p*-tolylamine (TTA) 986
trisazo compound 986
TROK model 167
true specific heat capacity 429
tuneable lasers 915
tungstate 289
tungsten bronze
– $PbNb_2O_6$ (lead metaniobate) 597
tuning photonic crystal response 1029
tunnel conductivity 1235
tunnel magnetoresistance (TMR) 1205, 1218
tunneling 927, 1046
– conduction 1491
– current 616, 624
– magnetoresistance (TMR) 1491
– magnetoresistive head 1218
– single electron 1231
– transition of charge carrier 211
– transition probability 201
turbo-disc reactor 324
twin crystal 1491
twinning 357
twist deformation 942
twist vector 1373
twisted nematic (TN) 951
– display, chiral dopant 952
twisted vertically aligned (TVAN) 954
two-beam experiment 1086
two-beam system 160
two-color detector 360
two-dimensional (2-D)
– electron gas (2DEG) 40, 105, 335, 785
– hole gas (2DHG) 789
– projection 344
– structure 1039
two-electron transition (TET) 774, 775
two-photon absorption 1491
two-step absorption 1083, 1087
type I band alignment 1040
type I superconductor 1233
type II band alignment 1040, 1051
type II superconductor 1232
– mixed state 1233
– pinning 1243
– vortex–vortex interaction 1243
type of film 669
Tzero technology 449

U

ultrafast nonlinear response 1491
ultrafine printing 1270
ultrahigh vacuum (UHV) 329, 649, 1491
ultrahigh vacuum-electron cyclotron resonance-radio-frequency magnetron sputtering (UHV-ECR-RMS) 766
ultrahigh-vacuum chemical vapour deposition (UHV-CVD) 523, 535
ultrahigh-vacuum environment 358
ultra-large-scale integration (ULSI) 293, 622
ultrasonic 716
– bonding 716
– micro-spectroscopy (UMS) 749
ultrathin films 1244
ultraviolet light absorption spectroscopy (UVLAS) 574
Umklapp process 432
unbalanced magnetron technique 662
under bump metallurgy (UBM) 1319
underfill 1323, 1491
uniaxial anisotropy 92, 1491
uniform planar alignment 1491
uniformly doped (UD) 789
unipolar avalanche photodiode (UAPD) 1052, 1491
unit cell 669
upper critical field
– cuprate superconductor 1240
– iron-based superconductor 1251
– MgB_2 1247
upper critical field (B_{c2}) 1491
upper shield 1214
Urbach law 685
Urbach tail 61
UV lithography 1018
UV poling 1088

V

V/G parameter 116
V/I boundary 128, 1491
vacancy (V) 113, 1491
– aggregate 115
– concentration 135, 137
– flux 136
– formation energy 114
vacuum
– deposition 649
– distillation (VD) 861
– evaporation (VE) 1089
– ultraviolet (VUV) 72
valence band (VB) 56, 899, 1083, 1491
– amorphous semiconductor 558
– deformation potentials 737
valence change glass 1009
valence electron 60
valence-alternation pairs (VAP) 559
valence-band offset
– $Al_{1-x}Ga_xAs/GaAs$ 185
– core-level XPS 180
– GaN and Ge heterostructure 187
– SiO_2 and Si heterostructure 186
van der Pauw technique 457
van Hove singularities 58
vanadium oxide (VOx) 869
vapor
– growth 278, 281, 284–286, 1491
– phase epitaxy (VPE) 5, 270, 310, 365, 879, 882, 899, 1491
– pressure constant, precursor 322
– sensing 1295
– transport deposition (VTD) 858
vapor-liquid-solid (VLS) 325, 799
– nanowire growth 1053
vapor-pressure-controlled Czochralski (VCZ) 281
vapor-solid (VS) 799
variable-range hopping (VRH) 203, 696, 1117, 1168
Varshini 771, 783
VCSEL technology 337
Vegard's law 727, 728, 1491
vehicle 1491
velocity–field characteristic 835
– AlN 836
– GaN 836
– group III–V nitride semiconductor 836
– InN 836
Verdet constant 69
Verneuil 271
– technique 1491
vertical
– external cavity surface emitting laser (VECSEL) 918
– gradient freezing (VGF) 270, 275, 283, 380, 884, 1491
– superlattice 1334
– transport 1491
vertical cavity 796
– surface emitting laser (VCSEL) 796, 899, 916, 1037, 1048, 1491
vertical external cavity surface emitting laser (VECSEL) 918
vertical unseeded vapor growth (VUVG) 284
vertically aligned nematic (VAN) 954
very high mobility 335
very-large-scale integration circuit (VLSI) 801
V-grove nanowire 1053
VI recombination 118, 1492
via-hole device 880
vibration dynamics 239
vibrational mean energy 426
videotape 97
virtual gap states (ViGS) 181, 1492
viscosity 1061
viscosity coefficient
– Miesowicz experiments 945
– nematodynamics 947
viscous torque 946
visible (VIS) 72
visual graphics array (VGA) 949
V-MRAM 108
void 112, 1492
– density 121
– morphology 123
volatile organic compound (VOC) 1282
Volmer–Weber (VW) 374
– growth 374
voltage noise measurement 475, 476
– low-impedance device 476
voltage scaling 618
voltage-controlled magnetism 107
vortex
– property 107
– transistor 108
VPE growth
– group II–VI wide bandgap 372
– source materials 372

W

W nonlinear figure of merit 1492
wafer scale packaging (WSP) 13
Wannier–Mott excitons 62
Wannier-type exciton 1331
waste in electrical and electronic equipment (WEEE) 1321
water-soluble materials 276
wave function, Bloch-type 486
wave soldering 716
wave-function overlap 924
wavelength 1340
– dispersive x-ray (WDX) 385
– division multiplexing (WDM) 543, 1030, 1068, 1492
wavenumber 1340
weak absorption 77
– tail (WAT) 61
weak link 1492
weak-link behavior, polycrystalline cuprate superconductor 1242
web photoreceptors 1492
Wentzel–Kramers–Brillouin (WKB) 697
white-light LED 1006
wide bandgap
– compounds 365, 374
– epitaxial growth 371
– group II–VI compound semiconductors 367
– group II–VI compounds 371
– material 743
– quantum dot 374
– quantum well 374
wide-band x-ray 863
– imager (WXI) 863
widegap compounds 316
width at one-half amplitude PW_{50} 1194
Wiedemann–Franz–Lorenz law 435
wire bonding 716, 1492
– digital applications 1315
– materials 1315
WO_3 168
work function 635, 690, 1492
write
– efficiency 1492
– gap 1492
– head 1189, 1205, 1492
write-head material 1191
wrong bond 1492
wurtzite 589, 743, 745
– GaN 835

– group III nitride alloy 782
– group III–V binaries 737, 738
– group III–V semiconductors 737
– structure 1492
WZ-GaN 747

X

xerogel 640
xerographic discharge 978, 1492
xerographic gain 971, 1492
xerography 968, 969, 1492
XeroX Copier 969
x-ray 1492
– absorption near-edge structure (XANES) 1153
– absorption spectroscopy (XAS) 748
– detector 314
– diffraction (XRD) 372, 385, 1151, 1178, 1286, 1492
– fluorescence (XRF) 390
– imaging 1008
– lithography 1024
– photoconductor property 1127, 1128
– sensitivity 1140, 1141, 1492
– storage phosphor (XRSP) 1008
– topography 1492
x-ray diffraction (XRD) 390
x-ray photoelectron spectroscopy (XPS) 413, 626, 627, 1492
– primary excitation 415

Y

YAG (yttrium aluminium garnet) 715
yellow luminescence (YL) 777
yield strength 1492
yoke 1191
Young's modulus 748, 1173, 1493
– semiconductors 730
yttrium iron garnet (YIG) 105
Y-type construction 1287

Z

ZBLAN glass 429
Zeeman energy 87
zeolite 1090
zero band-gap semiconductor 1176
zero temperature coefficient of resistance (ZTCR) 259
zero-dimensional (0-D) structure 1053
zero-phonon transition 1002
zinc compound 671
zincblende structure 1493
zirconia auto exhaust 259
zirconia based system 260
Zn-based compound 316
ZnO 590
– single crystals 377
ZnSe
– epitaxial layer 371
– single crystal 379
zone refining 4, 1493

Recently Published Springer Handbooks

Springer Handbook of Electronic and Photonic Materials (2nd) (2017)
ed. by Kasap, Capper, 1450 p., 978-3-319-48931-5

Springer Handbook of Global Navigation Satellite Systems (2017)
ed. by Teunissen, Montenbruck, 1328 p., 978-3-319-42926-7

Springer Handbook of Model-Based Science (2017)
ed. by Magnani, Bertolotti, 1157 p., 978-3-319-30525-7

Springer Handbook of Odor (2017)
ed. by Buettner, 1151 p., 978-3-319-26930-6

Springer Handbook of Electrochemical Energy (2017)
ed. by Breitkopf, Swider-Lyons, 1016 p., 978-3-662-46656-8

Springer Handbook of Robotics (2nd) (2016)
ed. by Siciliano, Khatib, 2227p., 978-3-319-32550-7

Springer Handbook of Ocean Engineering (2016)
ed. by Dhanak, Xiros, 1345 p., 978-3-319-32550-7

Springer Handbook of Computational Intelligence (2015)
ed. by Kacprzyk, Pedrycz, 1633 p., 978-3-662-43505-2

Springer Handbook of Marine Biotechnology (2015)
ed. by Kim, 1512 p., 978-3-642-53970-1

Springer Handbook of Acoustics (2nd) (2015)
ed. by Rossing, 1286 p., 978-1-4939-0754-0

Springer Handbook of Spacetime (2014)
ed. by Ashtekar, Petkov, 887 p., 978-3-642-41991-1

Springer Handbook of Bio-/Neuro-Informatics (2014)
ed. by Kasabov, 1230 p., 978-3-642-30573-3

Springer Handbook of Nanomaterials (2013)
ed. by Vajtai, 1222 p., 978-3-642-20594-1

Springer Handbook of Lasers and Optics (2nd) (2012)
ed. by Träger, 1694 p., 978-3-642-19408-5

Springer Handbook of Geographic Information (2012)
ed. by Kresse, Danko, 1120 p., 978-3-540-72678-4

Springer Handbook of Medical Technology (2011)
ed. by Kramme, Hoffmann, Pozos, 1500 p., 978-3-540-74657-7

Springer Handbook of Metrology and Testing (2nd) (2011)
ed. by Czichos, Saito, Smith, 1229 p., 978-3-642-16640-2

Springer Handbook of Crystal Growth (2010)
ed. by Dhanaraj, Byrappa, Prasad, Dudley, 1816 p., 978-3-540-74182-4

Springer Handbook of Nanotechnology (3rd) (2010)
ed. by Bhushan, 1961 p., 978-3-642-02524-2

Springer Handbook of Automation (2009)
ed. by Nof, 1812 p., 978-3-540-78830-0

Springer Handbook of Mechanical Engineering (2009)
ed. by Grote, Antonsson, 1576 p., 978-3-540-49131-6

Springer Handbook of Experimental Solid Mechanics (2008)
ed. by Sharpe, 1096 p., 978-0-387-26883-5

Springer Handbook of Speech Processing (2007)
ed. by Benesty, Sondhi, Huang, 1176 p., 978-3-540-49125-5

Springer Handbook of Experimental Fluid Mechanics (2007)
ed. by Tropea, Yarin, Foss, 1557 p., 978-3-540-25141-5

Springer Handbook of Engineering Statistics (2006)
ed. by Pham, 1120 p., 978-1-85233-806-0

Springer Handbook of Atomic, Molecular, and Optical Physics (2nd) (2005)
ed. by Drake, 1506 p., 978-0-387-20802-2